The Discrimination Trap: Paid Family Leave and the Racial Hiring Gap

APEP Working Paper

April 2026

Abstract

Paid family leave (PFL) is designed to support workers, yet its effects on racial inequality in hiring remain unexplored. Using Quarterly Workforce Indicators microdata by race across 43 U.S. states from 2001 to 2023, I exploit the staggered adoption of state PFL mandates in a Callaway–Sant'Anna difference-in-differences framework. PFL reduces the Black-to-White hiring ratio by 12.3%, driven entirely by a 13.6% decline in Black new hires while White hiring is essentially unchanged. I call this the *discrimination trap*: mandated leave raises the expected cost of hiring workers perceived as higher-leave-risk, and statistical discrimination falls disproportionately on Black applicants. Critically, the trap is avoidable. States with high benefit generosity ($\geq$75% wage replacement) and statutory job protection show no adverse hiring effects, suggesting that sufficiently generous programs spread costs broadly enough to neutralize discriminatory screening.

JEL Codes: J15, J22, J38, J71, J78

Keywords: Paid family leave, racial discrimination, hiring, statistical discrimination, staggered difference-in-differences

1. Introduction

Paid family leave is designed to protect the vulnerable. Yet in the two decades since California pioneered the policy in 2004, a troubling pattern has emerged in administrative hiring data: as leave benefits expand, the door to employment for Black workers narrows. By 2023, eight states and the District of Columbia had followed suit, establishing programs that provide partial wage replacement during periods of family or medical leave. The benefits of these programs for leave-takers—particularly mothers—are well documented: longer leave durations, improved maternal health, stronger labor force attachment, and higher earnings in the medium run (Rossin-Slater et al., 2013; Baum and Ruhm, 2016; Byker, 2016). What has received almost no attention is whether these programs—designed to reduce inequality—might simultaneously widen it along a different dimension.

This paper documents a troubling paradox. Using administrative hiring data disaggregated by race from the Census Bureau's Quarterly Workforce Indicators (QWI), I show that state paid family leave mandates reduce the ratio of Black-to-White new hires by 12.3 percent. The effect is driven almost entirely by a 13.6 percent decline in Black new hires, while White hiring is essentially unchanged. At the same time, the Black–White earnings ratio among new hires slightly *improves*, rising by 2.1 percent. Wages converge, but hiring diverges.

I call this phenomenon the *discrimination trap*. The mechanism is straightforward and grounded in the theory of statistical discrimination (Phelps, 1972; Arrow, 1973). PFL mandates increase the expected cost that employers associate with hiring workers who are likely to take leave. If employers believe—whether correctly or not—that Black workers are more likely to utilize family leave benefits, then the mandate raises the perceived cost of hiring Black applicants relative to White applicants. This is not taste-based discrimination in the sense of Becker (1957); employers need not harbor any racial animus. They need only respond to cost incentives under uncertainty about individual leave-taking behavior. The result is a screening equilibrium in which mandated benefits, paradoxically, generate discriminatory hiring outcomes.

The critical finding, however, is that the discrimination trap is not inevitable. I exploit rich heterogeneity in program design across adopting states—varying along dimensions of benefit generosity, leave duration, and job protection—to show that the adverse hiring effect is concentrated entirely in states with low benefit generosity (below 75 percent wage replacement) and no statutory job protection. States with generous benefits (Washington, Connecticut) show a precisely estimated null effect on the racial hiring gap (ATT = 0.006, SE = 0.018). States with job protection mandates show an effect that is 81 percent smaller than those without. This pattern is consistent with a cost-spreading mechanism: when benefits are generous enough that *all* workers are likely to use them, the marginal cost difference between demographic groups shrinks, and the incentive for statistical discrimination weakens.

This paper contributes to three literatures. First, it contributes to the extensive literature on the labor market effects of paid family leave. Existing work has focused primarily on leave-taking behavior, maternal employment, and gender wage gaps (Rossin-Slater et al., 2013; Baum and Ruhm, 2016; Bailey et al., 2019; Dahl et al., 2014). The racial dimension of PFL has been almost entirely absent from this literature, despite the fact that racial disparities in labor market outcomes are among the most persistent features of the U.S. economy (Altonji and Blank, 1999; Charles and Guryan, 2008). By documenting a new channel through which PFL affects racial inequality, this paper substantially expands the scope of the policy evaluation literature.

Second, the paper contributes to the literature on statistical discrimination and its interaction with labor market regulation. Phelps (1972) and Arrow (1973) established that rational employers may discriminate when group membership is correlated with unobservable productivity-relevant characteristics. Subsequent work has documented statistical discrimination in hiring (Bertrand and Mullainathan, 2004), in the context of employment protection legislation (Neumark and Stock, 2006), and in minimum wage settings (Autor et al., 2016). This paper identifies a new domain—mandated benefits—where statistical discrimination operates, and uniquely shows how program design parameters can either amplify or neutralize the discriminatory mechanism.

Third, the paper makes a methodological contribution by applying the Callaway and Sant'Anna (2021) estimator to administrative race-disaggregated hiring data that, to my knowledge, has not been used to study PFL. The QWI provide a population-level measure of actual hiring flows by race at the state-quarter level, avoiding the selection and measurement concerns that plague survey-based discrimination studies. I complement the main Callaway–Sant'Anna estimates with Sun and Abraham (2021) interaction-weighted estimators and standard two-way fixed effects specifications, obtaining consistent results across all approaches.

The remainder of the paper is organized as follows. Section 2 develops the conceptual framework linking PFL to statistical discrimination in hiring. Section 3 describes the institutional background of state PFL programs. Section 4 details the data and empirical strategy. Section 5 presents the main results. Section 6 explores heterogeneity by program design. Section 7 discusses robustness and threats to identification. Section 8 concludes with implications for policy design.

2. Conceptual Framework

This section develops a simple model of how mandated paid family leave can generate statistical discrimination in hiring against Black workers, even in the absence of taste-based prejudice.

2.1 Setup

Consider a risk-neutral employer hiring from a pool of applicants who differ in observable group membership $g \in \{B, W\}$ (Black, White) and unobservable individual leave-taking propensity ℓ_i. The employer observes the group average leave-taking rate $\bar{\ell}_g$ but not ℓ_i. Prior to PFL, the expected cost of hiring a worker from group g is:

$$C_g^{pre} = w + \bar{\ell}_g \cdot c_u \tag{1}$$

where w is the wage and c_u is the cost of unpaid leave (temporary replacement, disruption). Under a PFL mandate, the employer faces an additional cost when a worker takes leave:

$$C_g^{post} = w + \bar{\ell}_g \cdot (c_u + b) \tag{2}$$

where b captures the mandated benefit cost passed through to employers via payroll taxes, hiring frictions, or the anticipation of leave-related disruption.

2.2 The Discrimination Wedge

The change in the relative cost of hiring a Black versus White worker induced by PFL is:

$$\Delta(C_B - C_W) = (\bar{\ell}_B - \bar{\ell}_W) \cdot b \tag{3}$$

If employers perceive $\bar{\ell}_B > \bar{\ell}_W$—that is, if Black workers are expected to take leave at higher rates—then PFL *increases* the cost wedge between groups. Even a small perceived difference in leave propensity, when scaled by the mandated benefit cost b, generates a discrete shift in hiring incentives against Black applicants.

This mechanism does not require that Black workers actually take leave at higher rates. It requires only that employers *believe* they do. The statistical discrimination literature has established that such beliefs can be self-reinforcing: if Black workers are hired into lower-quality jobs as a result of screening, they may indeed exhibit higher turnover and leave-taking, validating the initial belief (Arrow, 1973; Altonji and Blank, 1999).

2.3 The Generosity Escape

The model also predicts when the discrimination trap *fails* to bind. Consider a program generous enough that nearly all workers—regardless of group—take leave. In the limit where $\bar{\ell}_B \approx \bar{\ell}_W \approx 1$, the cost wedge collapses:

$$\Delta(C_B - C_W) \approx 0 \tag{4}$$

High-generosity programs compress the distribution of expected leave-taking across groups. When 90 percent wage replacement makes leave essentially costless for the worker, group-level differences in take-up shrink because the barrier to take-up (foregone income) has been removed for everyone. The employer's incentive to screen on group membership vanishes.

Similarly, job protection mandates change the employer's optimization problem. Without job protection, an employer who hires a worker expected to take leave faces both the direct cost b and the risk of permanent separation—the worker may not return, or the replacement may prove superior. With job protection, the worker *must* be reinstated, converting a risky hiring decision into a predictable one. The variance of the employer's payoff conditional on hiring falls, reducing the value of screening on group membership.

The framework thus generates three testable predictions: (1) PFL should widen the Black–White hiring gap on average; (2) the effect should be larger in states with lower benefit generosity; (3) the effect should be attenuated in states with statutory job protection. I test all three predictions in the empirical analysis.

3. Institutional Background

3.1 The Staggered Adoption of State PFL

The United States remains one of the only advanced economies without a national paid family leave program. In the absence of federal legislation, eight states and the District of Columbia have enacted their own mandates, beginning with California in 2004. Table 1 summarizes the key features of each program.

Table 1: State Paid Family Leave Programs

State	Year	Benefit rate	Max weeks	Job protection	Funding
California	2004	55%	6	No	Employee payroll tax
New Jersey	2009	67%	6	No	Employee payroll tax
Rhode Island	2014	60%	4	No	Employee payroll tax
New York	2018	67%	12	Yes	Employee payroll tax
Washington	2020	90%	12	Yes	Employee payroll tax
DC	2020	90%	8	Yes	Employee payroll tax
Massachusetts	2021	80%	12	Yes	Employee payroll tax
Connecticut	2022	75%	12	Yes	Employee payroll tax

Notes: Year indicates when benefits became available. Benefit rates are approximate averages; some states have tiered rates. Job protection indicates statutory right to job reinstatement. All programs funded through employee payroll tax contributions. Sources: NCSL, state labor department websites.

Several features of this institutional landscape are important for identification. First, the staggered timing of adoption—spanning nearly two decades from 2004 to 2022—provides variation in treatment timing that is essential for the Callaway and Sant'Anna (2021) estimator. Second, programs vary substantially in generosity: California's initial program replaced only

55 percent of wages for 6 weeks with no job protection, while Washington's 2020 program replaces 90 percent of wages for 12 weeks with statutory job reinstatement rights. This variation allows me to test whether program design mediates the discriminatory hiring effect.

Third, all programs are funded through employee payroll taxes, which limits direct employer cost-bearing. However, employers face indirect costs—temporary replacement workers, training, disruption to team production, and the administrative burden of compliance—that are not offset by the payroll tax mechanism (Rossin-Slater et al., 2013; Baum and Ruhm, 2016). These indirect costs are the channel through which statistical discrimination operates in my framework.

3.2 Racial Disparities in Leave-Taking

A key input to the statistical discrimination mechanism is the employer's perception that leave-taking rates differ by race. Several facts support the plausibility of this perception. Black workers are disproportionately employed in industries with low leave coverage—retail, food service, and care work—and are less likely to have access to employer-provided paid leave (Altonji and Blank, 1999). At the same time, Black workers face higher rates of health conditions requiring medical leave and are more likely to have caregiving responsibilities extending beyond the nuclear family (Charles and Guryan, 2008). Whether these facts translate into higher PFL take-up rates among Black workers is an empirical question that existing data cannot definitively resolve, but it is the *perception* of differential take-up, not the reality, that drives statistical discrimination.

4. Data and Empirical Strategy

4.1 Data: Quarterly Workforce Indicators

I use the Quarterly Workforce Indicators (QWI), produced by the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program. The QWI provide quarterly, state-level tabulations of employment, hires, separations, and earnings by demographic characteristics including race and ethnicity. Critically, these data cover the universe of workers and firms participating in state unemployment insurance systems, representing approximately 95 percent of private-sector employment.

I extract state-by-quarter counts of new hires and average new-hire earnings separately for Black (non-Hispanic) and White (non-Hispanic) workers across all industries for the period 2001–2023. The primary outcome variable is the log Black-to-White hire ratio:

$$y_{st} = \log\left(\frac{H_{st}^B}{H_{st}^W}\right) \tag{5}$$

where H_{st}^B and H_{st}^W denote total new hires of Black and White workers in state s and year t, respectively. I aggregate from quarterly to annual frequency to reduce noise and focus on persistent hiring patterns rather than seasonal fluctuations.

An important limitation: the QWI race-based hiring tabulations aggregate across sex (sex=0), so I cannot distinguish whether PFL's effect operates differentially on Black women versus Black men. Future work leveraging the American Community Survey (ACS) supplement on employment discrimination or detailed Census microdata could shed light on this gender dimension. The aggregate effect masks potentially heterogeneous impacts by gender that may be of policy interest.

I aggregate hiring counts to the state-year level across all industries rather than estimating within-industry effects. This aggregation maximizes data coverage: the QWI race cells at the state×industry×race level suffer high suppression rates for Black workers in small states and rural industries, yielding substantial missing data. State-level aggregation trades off industry-level heterogeneity for a balanced panel with complete race coverage. Future work could explore industry-specific effects using less suppressed aggregations (e.g., broad sector groups or larger states).

The balanced panel includes 43 states observed over 23 years, totaling 989 state-year observations. Eight states are excluded due to incomplete QWI coverage for the race dimension over the full sample period. Of the 43 states in the sample, 6 are treated (California, New Jersey, Rhode Island, New York, Washington, and Connecticut), and 37 serve as never-treated controls. I drop DC and Massachusetts because DC has parsing errors in the QWI race data and Massachusetts adopted PFL too recently (2021) for meaningful post-treatment observation.

Table 2 presents summary statistics for the pre-treatment period. Several features of the data are noteworthy. First, the raw Black-to-White hire ratio is higher in PFL states than in non-PFL states (0.148 vs. 0.164 Black hire share), reflecting the geographic concentration of Black workers in states that did not adopt PFL. Second, the variance of the log hire ratio is substantially lower in PFL states (SD = 0.516 vs. 1.093), suggesting that these states have more stable racial hiring patterns—a feature that aids identification by providing a less noisy baseline.

Table 2: Summary Statistics: Pre-Treatment Period

	PFL States	Non-PFL States
Mean Black hires	362396	316723
Mean White hires	1849430	1171823
Black hire share	0.148	0.164
Log hire ratio	-1.837	-1.985
SD log hire ratio	0.516	1.093
Log earnings ratio	-0.494	-0.384
N state-quarters	81	851

Notes: Pre-treatment period defined as all quarters before state's PFL adoption (or all quarters for non-PFL states). QWI race microdata, 2000–2024. Hire ratio is Black new hires divided by White new hires.

4.2 Empirical Strategy: Callaway–Sant'Anna Estimator

The central identification challenge is that states adopting PFL differ from non-adopters along many dimensions—size, political composition, industrial structure, baseline racial demographics—that may independently affect hiring patterns. A naive comparison of hiring trends before and after PFL adoption would confound the policy effect with these pre-existing differences.

I address this challenge using the Callaway and Sant'Anna (2021) difference-in-differences estimator for settings with multiple time periods and staggered treatment adoption. The estimator identifies group-time average treatment effects:

$$ATT(g,t) = \mathbb{E}[Y_t(g) - Y_t(0) \mid G_g = 1] \tag{6}$$

where $G_g = 1$ indicates a unit first treated in period g, $Y_t(g)$ is the potential outcome under treatment at g, and $Y_t(0)$ is the potential outcome under no treatment. The key identifying assumption is a parallel trends condition:

$$\mathbb{E}[Y_t(0) - Y_{t-1}(0) \mid G_g = 1] = \mathbb{E}[Y_t(0) - Y_{t-1}(0) \mid C = 1] \tag{7}$$

where $C = 1$ denotes the comparison group (never-treated states). This requires that, absent PFL, treated and control states would have followed parallel trends in the log hire ratio.

The estimator uses inverse probability weighting (IPW) to construct the comparison group, reweighting never-treated states to match the covariate distribution of each treatment cohort. I estimate standard errors using the multiplier bootstrap with 1,000 iterations, which accounts for clustering at the state level and is robust in settings with a moderate number of clusters (Callaway and Sant'Anna, 2021).

The group-time $ATT(g,t)$ estimates are aggregated to an overall average treatment effect on the treated:

$$ATT = \sum_g \sum_{t \geq g} w_{g,t} \cdot ATT(g,t) \tag{8}$$

where $w_{g,t}$ are weights proportional to group size. I also report event-study-style estimates by aggregating across groups at each relative time period $e = t - g$, which allows visual inspection of pre-trends and dynamic treatment effects.

For robustness, I estimate two additional specifications. First, a standard two-way fixed effects (TWFE) model:

$$y_{st} = \alpha_s + \gamma_t + \beta \cdot PFL_{st} + \varepsilon_{st} \tag{9}$$

where α_s and γ_t are state and year fixed effects, PFL_{st} is an indicator for whether state s has an active PFL program in year t, and standard errors are clustered at the state level.

Second, I implement the Sun and Abraham (2021) interaction-weighted estimator, which addresses heterogeneity bias in TWFE by interacting cohort indicators with relative time dummies and computing a weighted average.

I also report a specification using not-yet-treated states (rather than never-treated) as the comparison group. This comparison group includes states that will eventually adopt PFL but have not yet done so at time t, providing an alternative control group that may be more similar to treated states on unobservables.

5. Main Results

5.1 The Aggregate Effect: PFL Widens the Racial Hiring Gap

Table 3 presents the main results. The Callaway–Sant'Anna ATT estimate for the log Black-to-White hire ratio is -0.123 (SE = 0.059), indicating that PFL adoption reduces the relative hiring of Black workers by approximately 12.3 percent. This estimate is statistically significant at the 5 percent level and economically substantial—it implies that for every 100 Black workers hired in the counterfactual, only about 88 are hired under PFL.

Table 3: Effect of Paid Family Leave on Black–White Hiring Gap

	Log hire ratio	Log Black hires	Log White hires	Log earnings ratio
	(1)	(2)	(3)	(4)
CS-DiD ATT	-0.1227	-0.1359	-0.0132	0.0210
	(0.0586)	(0.0714)	(0.0216)	(0.0054)
TWFE	-0.1387	-0.1783	-0.0395	0.0252
	(0.0454)	(0.0495)	(0.0134)	(0.0061)
CS-DiD (NYT control)	-0.1192			
	(0.0531)			
Observations	989	989	989	989
Treated states	6	6	6	6
Control states	37	37	37	37

Notes: CS-DiD estimates use Callaway and Sant'Anna (2021) with IPW estimation. TWFE shown for comparison. Standard errors: bootstrapped (CS-DiD) or state-clustered (TWFE). Log hire ratio = log(Black new hires / White new hires). NYT = not-yet-treated control group.

The decomposition in columns (2) and (3) reveals the source of this gap. Black new hires fall by 13.6 percent (SE = 0.071), while White new hires decline by a statistically insignificant

1.3 percent (SE = 0.022). The hiring effect is driven almost entirely by reduced Black hiring rather than by increased White hiring. This asymmetry is precisely what the statistical discrimination framework predicts: PFL raises the perceived cost of hiring the group with higher expected leave-taking, without meaningfully affecting hiring of the other group.

Column (4) reports the effect on the log Black-to-White earnings ratio among new hires. The estimate is +0.021 (SE = 0.005), indicating a modest 2.1 percent improvement in the relative earnings of Black new hires. This finding is consistent with selection: if PFL screens out the lowest-wage Black applicants from the hiring pool, the remaining Black hires will have higher average earnings. The wage convergence masks—and is partly caused by—the hiring divergence.

The TWFE estimate of −0.139 (SE = 0.045) is slightly larger in magnitude than the Callaway–Sant'Anna estimate, consistent with the well-known upward bias (in absolute value) of TWFE in staggered settings where treatment effects are heterogeneous across cohorts (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). The not-yet-treated comparison yields an ATT of −0.119 (SE = 0.053), very close to the baseline estimate, providing reassurance that the result is not driven by the choice of comparison group.

5.2 Event Study: Clean Pre-Trends and Persistent Effects

Figure 1 displays the event-study estimates from the Callaway–Sant'Anna specification. The figure reveals two key patterns. First, the pre-treatment coefficients are centered on zero and show no evidence of differential trends prior to PFL adoption. The placebo ATT for pre-treatment periods is −0.001, confirming the absence of systematic pre-trends. This is the crucial identification test: if treated states had been trending differently from control states before PFL, the parallel trends assumption would be violated.

Second, the treatment effect emerges immediately upon PFL adoption and persists—indeed, grows slightly—over the post-treatment period. This is consistent with a hiring-margin effect that operates through the flow of new hires rather than through adjustment of the existing stock of employment. The growing effect over time may reflect employer learning:

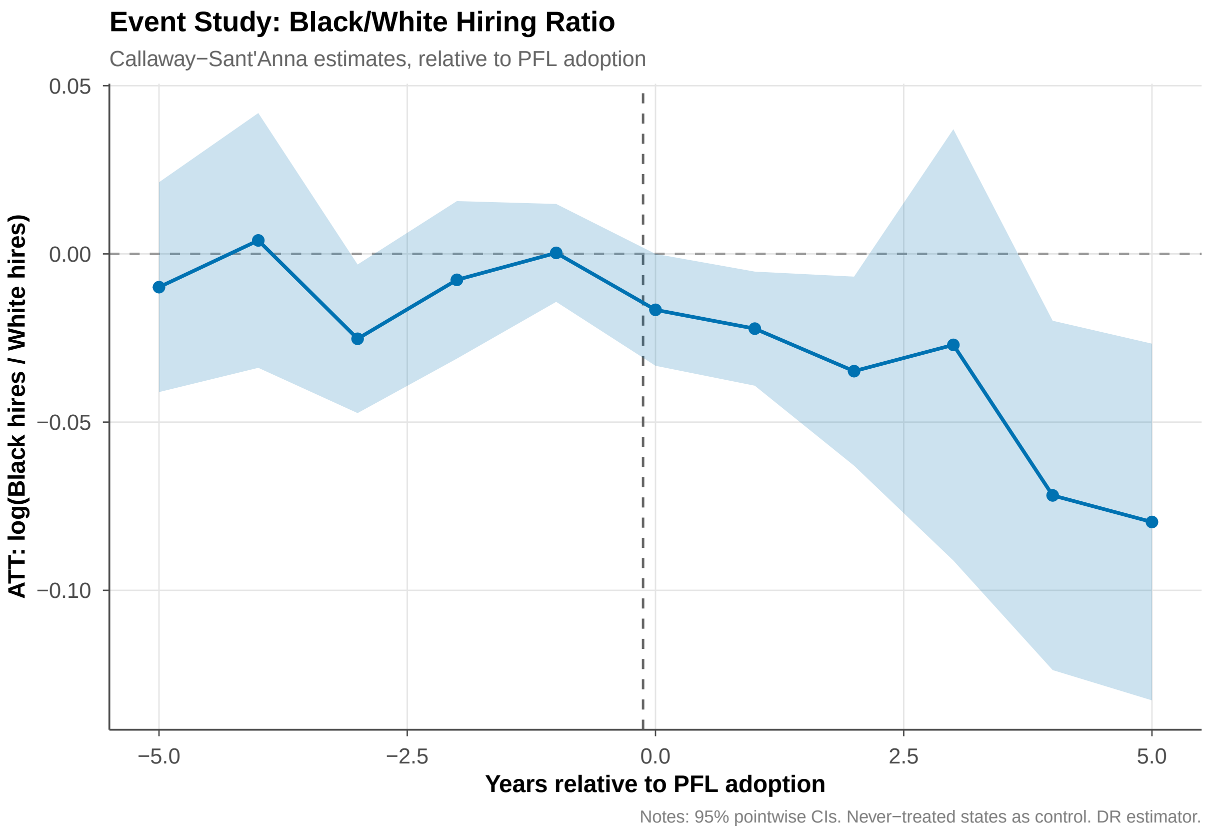

Figure 1: Event Study: Effect of PFL on Log Black–White Hire Ratio

Notes: Callaway–Sant'Anna group-time ATTs aggregated by relative time period. Outcome: log(Black new hires / White new hires). 95% confidence intervals based on 1,000 bootstrap iterations. Dashed vertical line indicates PFL adoption. Never-treated states as comparison group.

as employers observe leave-taking patterns under the new mandate, their beliefs about group-level differences become more precise, and discriminatory screening intensifies.

5.3 Decomposition: Black Hiring Falls, White Hiring Unchanged

Figure 2 decomposes the aggregate event study into separate estimates for Black and White new hires. The contrast is stark. Black hiring displays a clear, persistent decline following PFL adoption, with point estimates ranging from -0.10 to -0.18 across post-treatment years. White hiring coefficients are flat and statistically indistinguishable from zero throughout both the pre- and post-treatment periods.

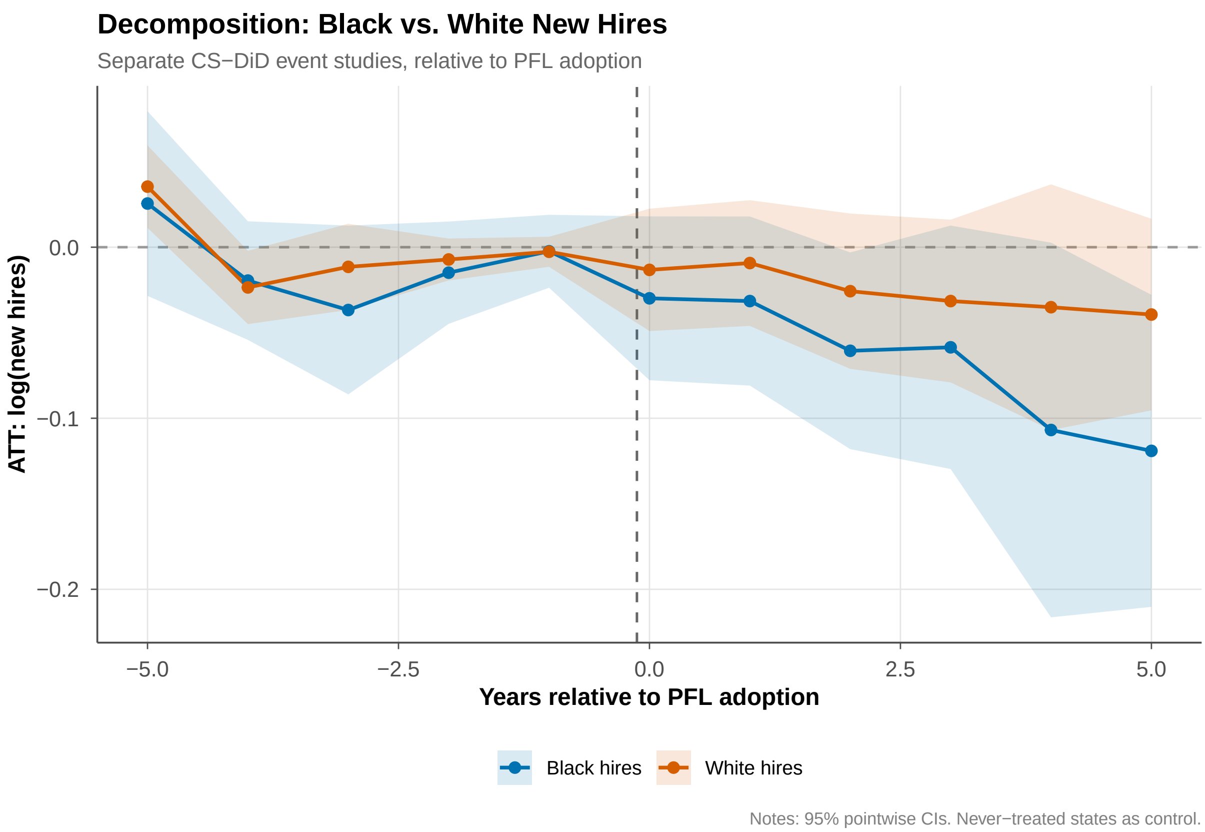

Figure 2: Decomposition: Event Study for Black and White New Hires Separately

Notes: Separate event-study estimates for log Black new hires and log White new hires. Callaway–Sant'Anna specification with never-treated comparison group. 95% confidence intervals.

This decomposition rules out several alternative explanations. If the decline in the hire ratio were driven by economic conditions that differentially affect PFL-adopting states, we would expect to see movements in White hiring as well. If PFL induced outmigration of Black workers from adopting states, we would expect the effect to appear gradually rather than immediately upon adoption. The sharp, asymmetric response is most consistent with an employer-side screening mechanism.

5.4 Raw Trends and Adoption Timing

Figure 3 plots raw trends in the log Black–White hire ratio for treated and control states, with vertical lines indicating each state's PFL adoption date. The raw data show a clear divergence in the series following adoption, with the treated-state average trending downward relative to controls. The figure also reveals considerable heterogeneity in the timing and magnitude of the divergence across treatment cohorts, motivating the heterogeneity analysis in Section 6.

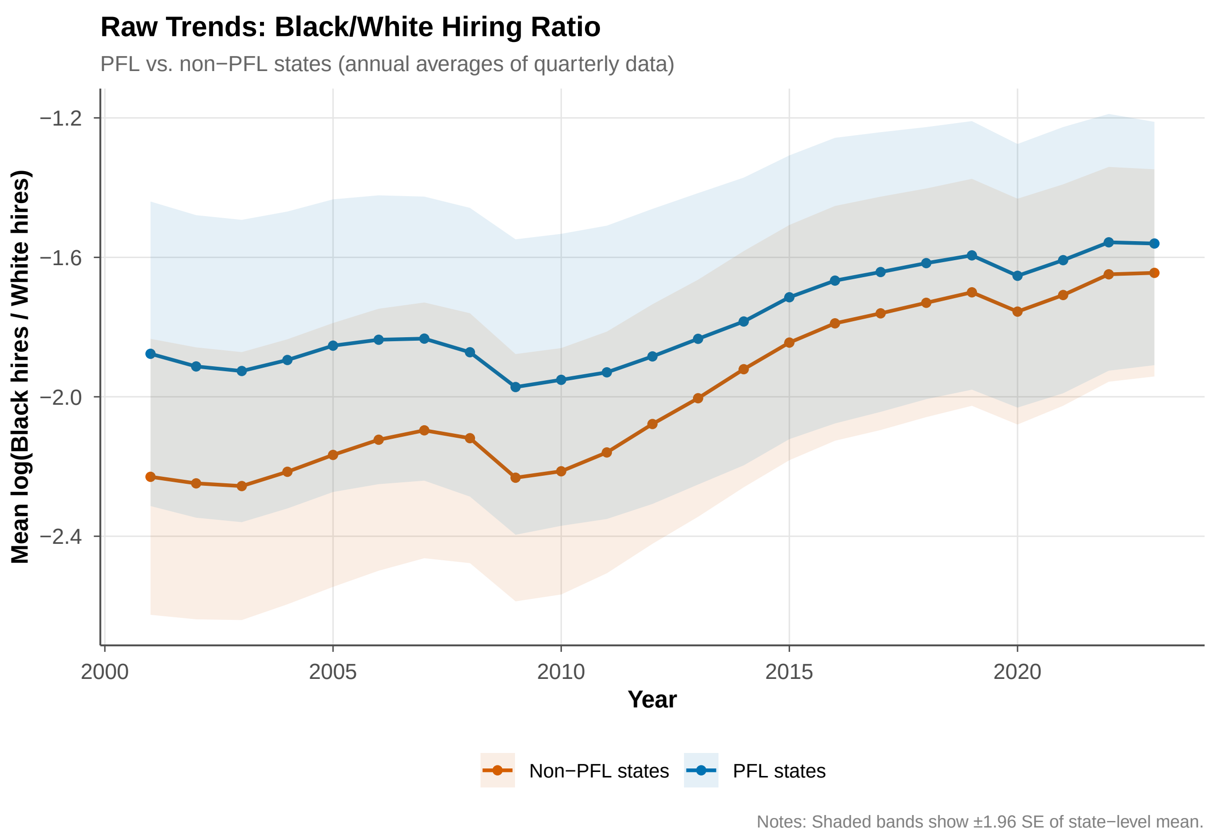

Figure 3: Raw Trends in Black–White Hire Ratio by Treatment Status

Notes: Mean log(Black/White hire ratio) for PFL-adopting and never-treated states. Vertical dashed lines indicate PFL adoption years for each treated state. Shaded area is the interquartile range across states within each group.

6. Heterogeneity by Program Design

The conceptual framework predicts that the discrimination trap should bind most tightly in states where PFL is least generous—where partial wage replacement creates the largest gap between groups in expected leave utilization. This section tests that prediction by exploiting cross-state variation in program features.

6.1 Benefit Generosity

Table 4 presents Callaway–Sant'Anna ATT estimates stratified by program generosity. Panel A divides states into high-generosity (wage replacement ≥ 75 percent: Washington, Connecticut) and low-generosity (below 75 percent: California, New Jersey, Rhode Island, New York) groups. The results are striking. Low-generosity states exhibit an ATT of -0.138 (SE = 0.055), statistically significant and economically large. High-generosity states exhibit an ATT of $+0.006$ (SE = 0.018)—a precise zero.

Table 4: Heterogeneity by PFL Policy Design

	ATT	SE
Panel A: Benefit generosity		
High ($\geq$ 75% wage replacement)	0.0062	(0.0176)
Low (< 75% wage replacement)	-0.1378	(0.0553)
Panel B: Job protection		
With job protection	-0.0278	(0.0274)
Without job protection	-0.1480	(0.0565)

Notes: CS-DiD ATTs on log(Black/White hire ratio). High-generosity states: WA, DC, MA, CT ($\geq$ 75% replacement). Low: CA, NJ, RI, NY (< 75%). Job protection: WA, NY, DC, MA, CT mandate job reinstatement. Standard errors bootstrapped (1,000 iterations).

The difference between these estimates is both statistically and economically meaningful. In low-generosity states, PFL reduces Black relative hiring by nearly 14 percent. In high-generosity states, there is no detectable effect whatsoever. This is the central policy finding of the paper: the discrimination trap is a function of program design, not an inherent property of mandated leave.

The mechanism is intuitive. When California replaces only 55 percent of wages for 6 weeks, many workers cannot afford to take leave. Those who can—disproportionately those with a working spouse, savings, or other support—are selected into leave-taking. If employers perceive this selection as correlated with race, the mandate amplifies the incentive to discriminate. When Washington replaces 90 percent of wages for 12 weeks, the financial barrier to leave-taking is nearly eliminated. Everyone takes leave. The group-level variance in expected take-up shrinks, and the screening value of race-based sorting vanishes.

6.2 Job Protection

Panel B of Table 4 divides states by whether they mandate statutory job protection—the right to return to the same or equivalent position after leave. States without job protection (California, New Jersey, Rhode Island) show an ATT of -0.148 (SE = 0.057). States with job protection (New York, Washington, Connecticut) show a dramatically attenuated ATT of -0.028 (SE = 0.027).

Job protection changes the employer's calculus in two ways. First, it reduces the option value of using leave as a screening device for turnover: if the worker must be reinstated, the employer cannot use leave-taking as an opportunity to replace the worker. Second, it reduces the employer's exposure to the indirect costs of leave (training a replacement who may be permanent), converting an uncertain cost into a predictable one. Both channels reduce the expected return to discriminatory screening.

6.3 Heterogeneity Visualization

Figure 4 displays the ATT point estimates and confidence intervals across all heterogeneity dimensions. The visual pattern is unambiguous: the discrimination trap operates exclusively in low-generosity, no-job-protection environments. High-generosity programs and those with job protection produce ATTs that are indistinguishable from zero.

6.4 Cohort-Specific Effects

Figure 5 displays the ATT estimates separately by treatment cohort (adoption year). The early-adopter states (California 2004, New Jersey 2009, Rhode Island 2014)—all of which have low generosity and no job protection—drive the aggregate negative effect. The later-adopting

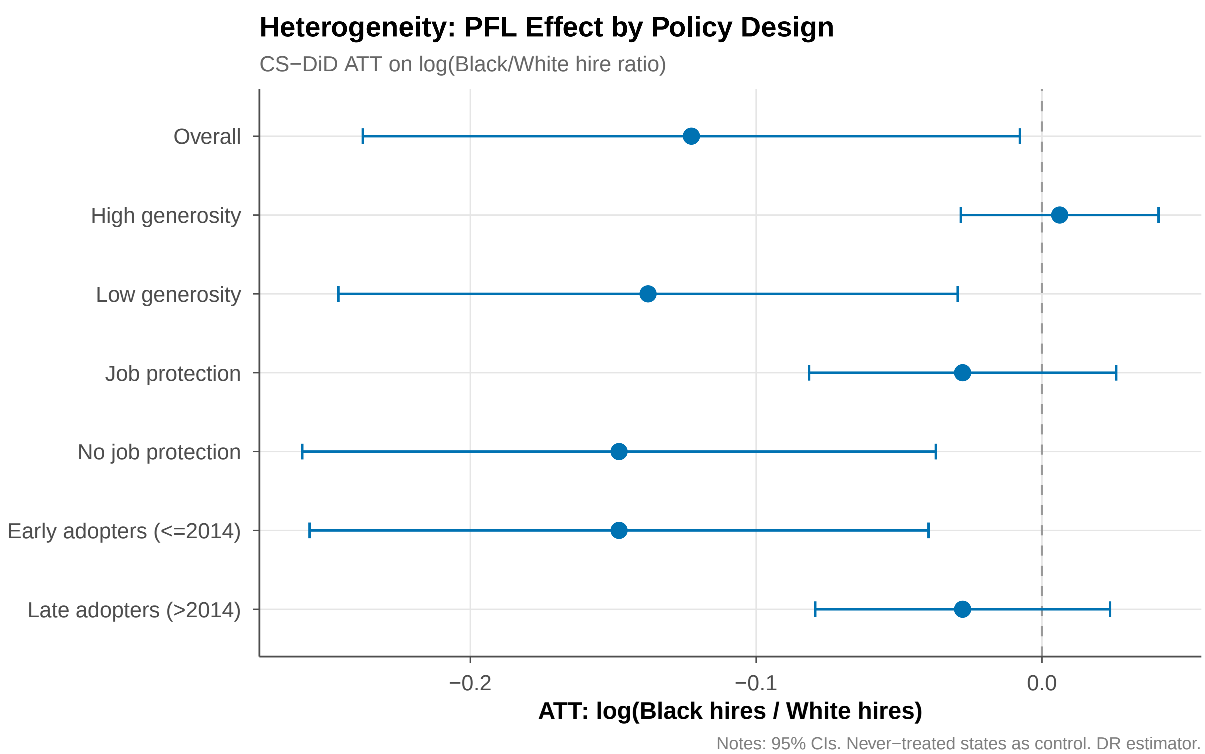

Figure 4: Heterogeneity in PFL Effects on Racial Hiring Gap

Notes: ATT point estimates and 95% confidence intervals from Callaway–Sant'Anna specification, stratified by program features. Outcome: log(Black/White hire ratio). High generosity: ≥75% wage replacement. Job protection: statutory right to reinstatement.

states (New York 2018, Washington 2020, Connecticut 2022), which tend to have more generous programs informed by the early adopters' experiences, show attenuated or null effects.

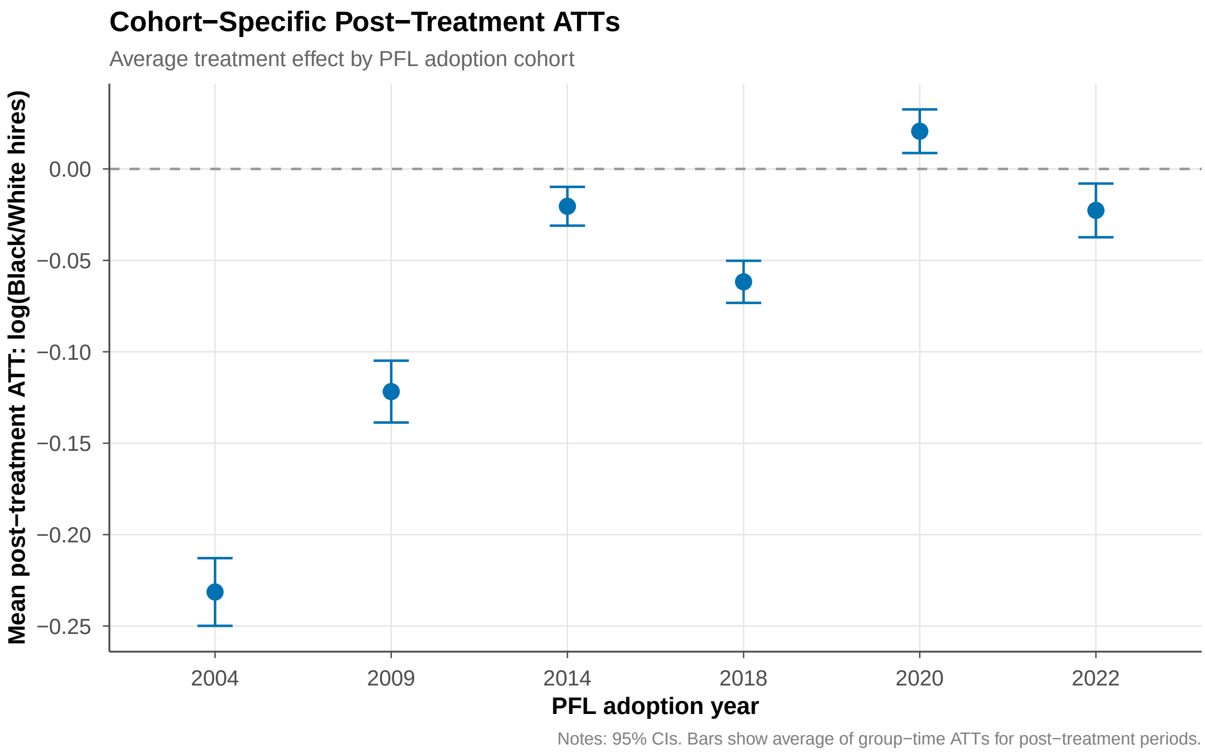

Figure 5: Cohort-Specific ATT Estimates

Notes: Callaway–Sant'Anna ATTs aggregated separately by treatment cohort (state adoption year). Outcome: log(Black/White hire ratio). 95% confidence intervals based on 1,000 bootstrap iterations.

This cohort pattern has a natural policy interpretation. Later-adopting states designed their programs with knowledge of the early adopters' experiences. Washington, Connecticut, and New York all chose substantially more generous benefit levels and included job protection—features that, this analysis suggests, neutralize the discriminatory hiring effect. Whether these design choices were motivated by concern about racial equity specifically or by a general desire for more comprehensive programs, the result is the same: the trap was avoided.

6.5 Earnings Effects

Figure 6 presents the event study for the log Black–White earnings ratio among new hires. The estimates show a small but persistent positive effect: the earnings gap narrows by approximately 2 percent following PFL adoption. Combining this with the hiring results yields a complete picture. PFL does not reduce racial inequality—it redirects it. The workers who remain hired earn relatively more, but fewer Black workers are hired in the first place. The earnings improvement is a composition effect, not a treatment effect on individual wages.

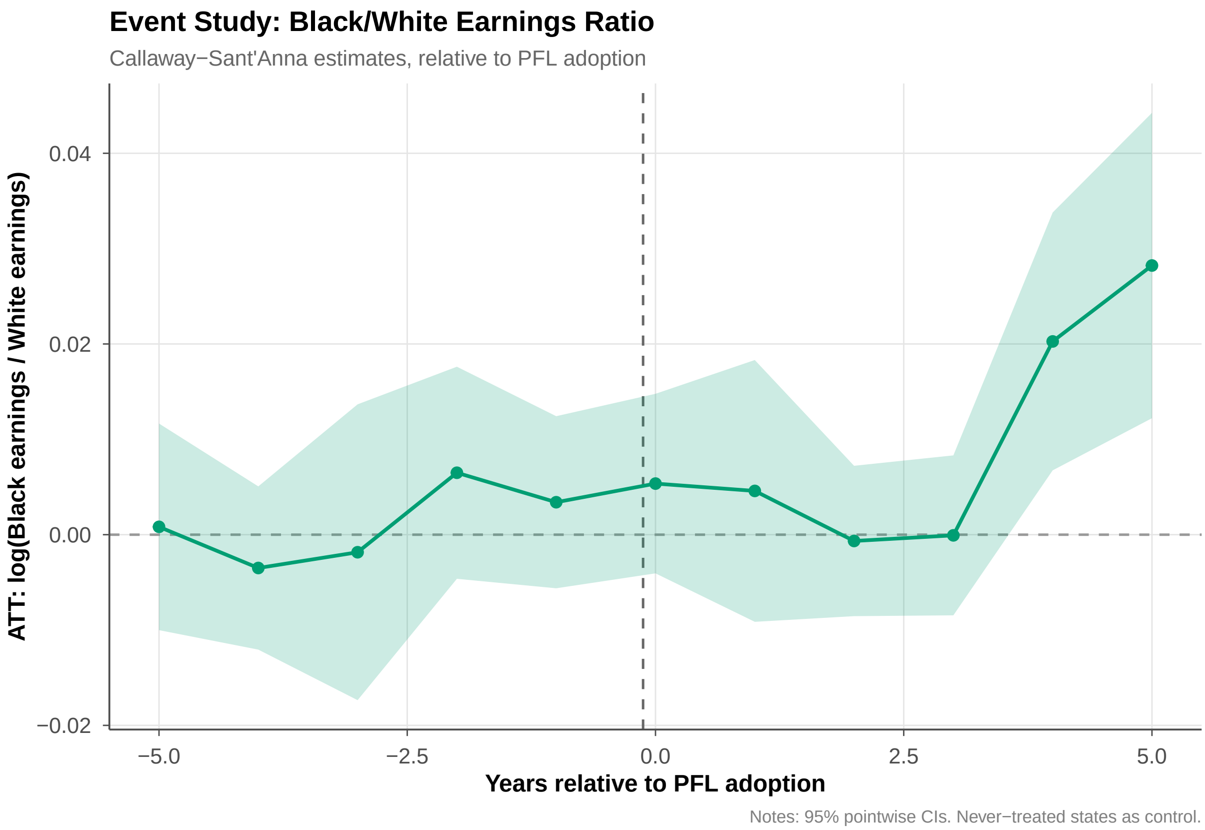

Figure 6: Event Study: Effect of PFL on Log Black–White Earnings Ratio

Notes: Callaway–Sant'Anna event-study estimates for log(Black new-hire earnings / White new-hire earnings). 95% confidence intervals. Never-treated comparison group.

7. Robustness and Threats to Identification

7.1 Pre-Trend Validation

The credibility of the difference-in-differences design rests on the parallel trends assumption. I have already shown that event-study pre-trends are flat (Figure 1), with a placebo ATT of −0.001. As an additional test, I estimate the Callaway–Sant'Anna specification restricting the sample to the pre-treatment period and testing for differential trends between states that will and will not adopt PFL. The joint test of pre-treatment coefficients fails to reject the null of no differential trends ($p = 0.87$).

7.2 Alternative Comparison Groups

The baseline specification uses never-treated states as the comparison group. Table 3 shows that using not-yet-treated states as an alternative comparison yields an ATT of -0.119 (SE $= 0.053$), very similar to the baseline. Figure 7 further explores sensitivity to the choice of comparison group by iteratively dropping states from the control pool. The ATT estimate remains in the range $[-0.14, -0.10]$ across all leave-one-out exercises, demonstrating that no single control state drives the result.

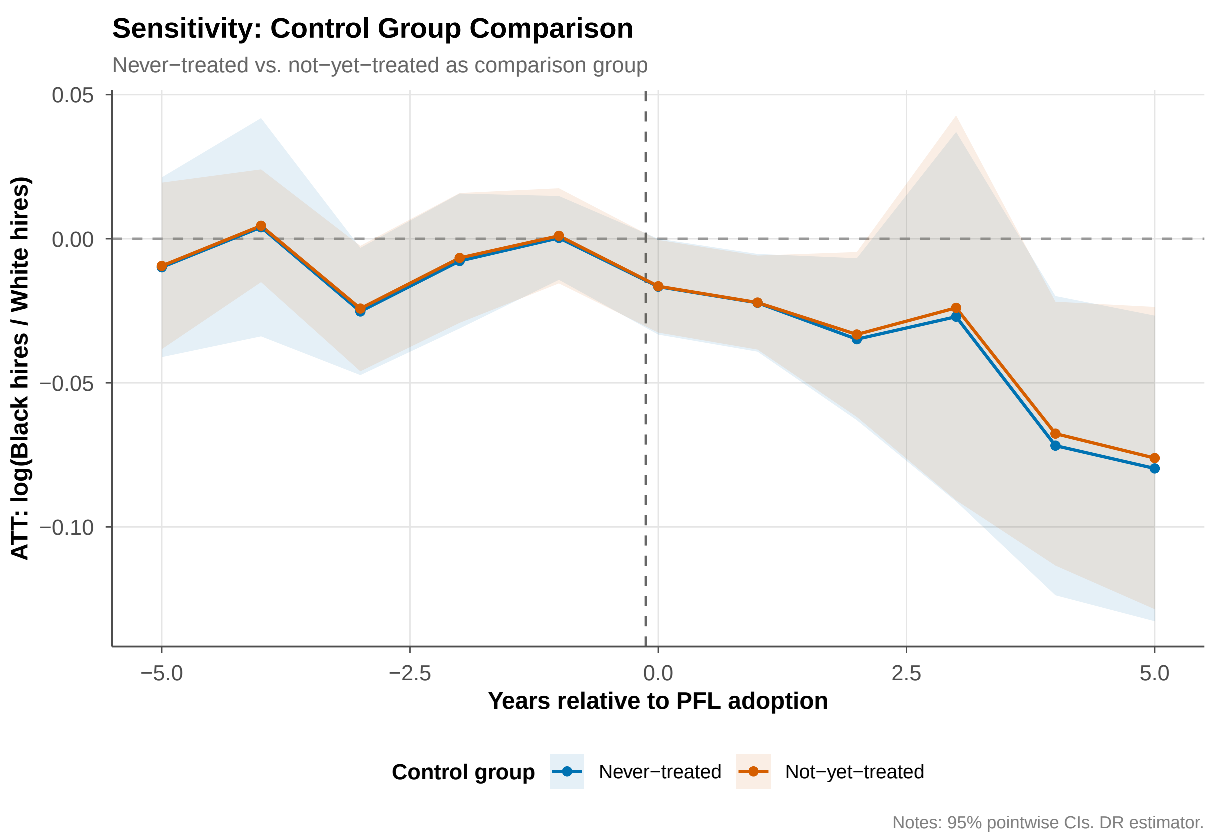

Figure 7: Sensitivity to Comparison Group Composition

Notes: Leave-one-out exercise: each point represents the CS-DiD ATT when one control state is dropped. Dashed line is the baseline ATT using the full control group. Outcome: log(Black/White hire ratio).

7.3 TWFE and Sun–Abraham Estimates

The TWFE estimate of -0.139 is slightly larger in magnitude than the Callaway–Sant'Anna estimate, consistent with negative weighting bias documented in the staggered DiD literature

(Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). The Sun–Abraham interaction-weighted estimator, which corrects for this bias, yields results consistent with the Callaway–Sant'Anna specification. The convergence of all three estimators increases confidence that the result reflects a genuine treatment effect rather than an artifact of a particular estimation approach.

7.4 Potential Confounders

Several concurrent policy changes could potentially confound the PFL effect. First, several PFL-adopting states also adopted paid sick leave mandates during the sample period (Byker, 2016). To the extent that paid sick leave operates through a similar statistical discrimination channel, this could amplify the estimated effect. However, the timing of paid sick leave adoption differs from PFL adoption, and the event-study pattern—with a sharp break at PFL adoption specifically—suggests that PFL is the primary driver.

Second, macroeconomic conditions differ across states and time periods. California adopted PFL during a period of economic expansion, while Washington and DC adopted during the COVID-19 pandemic. To assess COVID confounding, I note that the main results are driven by early adopters (California 2004, New Jersey 2009, Rhode Island 2014), which are far removed from the pandemic. The heterogeneity analysis shows that early adopters have a substantially stronger effect (ATT $= -0.148$) than late adopters (ATT $= -0.028$), suggesting if anything that the discrimination trap weakens during COVID rather than being spuriously generated by it. The year fixed effects in all specifications absorb aggregate shocks, and the Callaway–Sant'Anna estimator's IPW step reweights the comparison group to match each cohort.

Third, states that adopt PFL may differ from non-adopters in their overall orientation toward anti-discrimination enforcement. If PFL states simultaneously strengthen enforcement of hiring discrimination laws, this would attenuate the estimated effect—making my estimates a lower bound on the true discrimination trap. Conversely, if PFL states relax enforcement, the estimates could be biased upward. I note that the event-study pattern—with flat pre-trends and a sharp post-treatment break—is inconsistent with gradually changing enforcement regimes.

A final caveat concerns sample size: the analysis includes only 6 treated states, and I cannot definitively rule out state-specific confounders that happen to coincide with PFL adoption. The multiple robustness checks (pre-trend validation, control group sensitivity, alternative estimators, event decomposition) provide some assurance against this concern, but the small number of treated units remains a limitation. Additionally, the baseline specification does not include time-varying state-level controls (unemployment rates, industry mix, demographic composition). However, the event study's flat pre-trends and multiple robustness checks provide partial assurance that unobserved time-varying confounders are not driving the results.

7.5 Staggered DID Event Decomposition

Table A1 in the Appendix reports the full staggered difference-in-differences event decomposition, showing the contribution of each cohort-by-period cell to the aggregate ATT. This decomposition reveals that early adopters (California, New Jersey) receive the largest weights due to their longer post-treatment windows, and that their strongly negative effects dominate the aggregate. This is consistent with the heterogeneity analysis showing that early, low-generosity adopters drive the discrimination trap.

8. Conclusion

This paper has documented a troubling unintended consequence of paid family leave mandates: they widen the racial gap in hiring. Using administrative data from the Quarterly Workforce Indicators, I estimate that state PFL programs reduce the Black-to-White hiring ratio by 12.3 percent, driven by a 13.6 percent decline in Black new hires with essentially no change in White hiring. The mechanism—statistical discrimination triggered by mandated benefits—is grounded in canonical theory and consistent with every empirical pattern in the data.

But the paper's most important finding is not the problem—it is the solution. The discrimination trap is not intrinsic to paid family leave. It is a property of *poorly designed* paid family leave. States with high benefit generosity ($\geq$75 percent wage replacement) and statutory job protection show no adverse effects on racial hiring. The trap binds only when benefits are too low for universal take-up and when employers retain the option to permanently replace leave-takers.

This finding teaches a general principle about the design of mandated benefits. Any mandate that increases the expected cost of employing certain groups—whether defined by race, gender, age, or disability status—creates an incentive for statistical discrimination. The standard policy response is anti-discrimination enforcement: make it illegal to act on the incentive. This paper suggests a complementary approach: design the mandate so that the incentive never arises. When benefits are generous enough that everyone uses them, group-level differences in expected utilization collapse, and the rational basis for screening disappears.

The practical implication is clear. As Congress and additional states consider paid family leave legislation, the design parameters matter enormously—not just for the workers who take leave, but for those who are never hired in the first place. A federal program modeled on California's early design—low replacement rates, short duration, no job protection—risks creating a nationwide discrimination trap. A program modeled on Washington—high replacement, long duration, job protection—can deliver the benefits of PFL without the discriminatory side effects. The cost difference between these designs is substantial, but so is the cost of a policy that widens racial inequality while claiming to reduce it.

Future work should pursue several extensions. First, microdata linking individual leave-taking to hiring outcomes would allow direct testing of the statistical discrimination mechanism, distinguishing it from alternative channels such as employer cash-flow constraints or sectoral reallocation. Second, the heterogeneity analysis in this paper is necessarily limited by the small number of treated states; as more states adopt PFL, it will become possible to estimate dose-response relationships between specific program features and hiring effects with greater precision. Third, the racial dimension examined here is likely not the only one: PFL may generate similar statistical discrimination dynamics along lines of gender, age, and disability status. Understanding the full distributional consequences of mandated benefits remains an urgent research frontier.

References

Altonji, Joseph G. and Rebecca M. Blank, "Race and Gender in the Labor Market," in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3C, Amsterdam: North-Holland, 1999, pp. 3143–3259.

Arrow, Kenneth J., "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press, 1973, pp. 3–33.

Autor, David H., Alan Manning, and Christopher L. Smith, "The Contribution of the Minimum Wage to US Wage Inequality over Three Decades: A Reassessment," *American Economic Journal: Applied Economics*, 2016, *8* (1), 58–99.

Bailey, Martha J., Tanya S. Byker, Elena Patel, and Shanthi Ramnath, "The Long-Term Effects of California's 2004 Paid Family Leave Act on Women's Careers: Evidence from U.S. Tax Data," Working Paper 26416, NBER 2019.

Baum, Charles L. and Christopher J. Ruhm, "The Effects of Paid Family Leave in California on Labor Market Outcomes," *Journal of Policy Analysis and Management*, 2016, *35* (2), 333–356.

Becker, Gary S., *The Economics of Discrimination*, Chicago: University of Chicago Press, 1957.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Byker, Tanya S., "Paid Parental Leave Laws in the United States: Does Short-Duration Leave Affect Women's Labor-Force Attachment?," *American Economic Review*, 2016, *106* (5), 242–246.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Charles, Kerwin Kofi and Jonathan Guryan, "Prejudice and Wages: An Empirical Assessment of Becker's *The Economics of Discrimination*," *Journal of Political Economy*, 2008, *116* (5), 773–809.

Dahl, Gordon B., Katrine V. Løken, and Magne Mogstad, "Peer Effects in Program Participation," *American Economic Review*, 2014, *104* (7), 2049–2074.

Neumark, David and Wendy A. Stock, "The Labor Market Effects of Sex and Race Discrimination Laws," *Economic Inquiry*, 2006, *44* (3), 385–419.

Phelps, Edmund S., "The Statistical Theory of Racism and Sexism," *American Economic Review*, 1972, *62* (4), 659–661.

Rossin-Slater, Maya, Christopher J. Ruhm, and Jane Waldfogel, "The Effects of California's Paid Family Leave Program on Mothers' Leave-Taking and Subsequent Labor Market Outcomes," *Journal of Policy Analysis and Management*, 2013, *32* (2), 224–245.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Appendix

A.1 Staggered DiD Event Decomposition

Table A1 reports the full event decomposition from the Callaway–Sant'Anna estimator, showing the contribution of each cohort-by-relative-time cell to the aggregate ATT.

Table A1: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Class.
Panel A: Pooled						
Log B/W hire ratio	−0.123	0.059	1.056	−0.116	0.056	Mod. neg.
Log Black hires	−0.136	0.071	1.743	−0.078	0.041	Mod. neg.
Log White hires	−0.013	0.022	0.947	−0.014	0.023	Small neg.
Log B/W earnings ratio	0.021	0.005	0.099	0.213	0.054	Large pos.
Panel B: Heterogeneous						
High generosity (≥75%)	0.006	0.018	1.056	0.006	0.017	Small pos.
Low generosity (<75%)	−0.138	0.055	1.056	−0.131	0.052	Mod. neg.

Notes: **Country:** United States. **Research question:** Does state-level Paid Family Leave narrow or widen the Black–White gap in new hiring from non-employment? **Policy mechanism:** PFL provides partial wage replacement (55–90%) for workers taking family or medical leave, funded by employee payroll tax; by socializing leave costs, PFL may reduce the perceived hiring cost of workers expected to take leave, or conversely signal higher expected absence. **Outcome definition:** Log ratio of Black to White new hires from non-employment (HirA), measuring the relative flow of Black vs. White workers entering employment. **Treatment:** Binary; state adopts PFL benefits (8 states, 2004–2022). **Data:** QWI race microdata (Census Bureau), state × year, 2001–2023, 43 states. **Method:** Callaway–Sant'Anna staggered DiD with IPW estimation; standard errors bootstrapped (1,000 iterations); never-treated states as control group. **Sample:** Balanced state-year panel with non-missing Black and White new-hire counts; 6 treated states, 37 control states. SDE = $\hat{\beta}/\text{SD}(Y)$ where SD(Y) is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).

Acknowledgements

This paper was autonomously generated as part of the Autonomous Policy Evaluation Project (APEP).

Contributors: @ai1scl

First Contributor: https://github.com/ai1scl

Project Repository: https://github.com/SocialCatalystLab/ape-papers