A Additional Figures and Tables

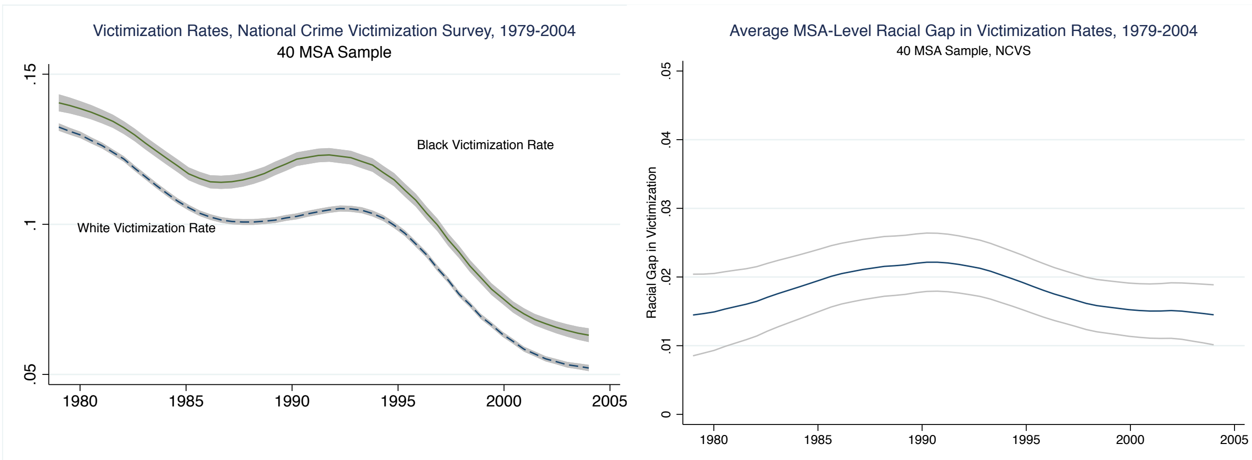


Figure A.1

B United States v. City of Chicago (1973)

The potential causal mechanisms discussed in the article text are illustrated by the history of the employment discrimination litigation filed against the Chicago Police Department (CPD) in 1973, culminating in a 1976 judicial order to end discriminatory employment practices and to hire and promote more Black police officers (United States v. City of Chicago, 411 F. Supp. 218 (N.D. Ill. 1976)).[14]

In the late 1960s approximately 17% of CPD officers were Black, while approximately 32% of the city's residents were Black [Pihos, 2015]. In a series of surveys of police officers in Chicago, Boston, and Washington D.C. conducted in 1966, Black and Reiss [1967] found that 38% of white officers were "highly prejudiced, extremely anti-Negro" while 34% of white officers were "prejudiced, anti-Negro" [Black and Reiss, 1967]. In surveys of residents of these cities, Black and Reiss [1967] found that white civilians were more satisfied with the policing in their city than Black civilians.

In 1968 a group of Black CPD officers formed the Afro-American Patrolmen's League (AAPL) in response to what they perceived to be the failure of the largely white Chicago Police Department to protect residents of largely Black neighborhoods. The AAPL's leaders decried the response of the CPD to calls for service from Black neighborhoods, including not responding to calls, responding too slowly, refusing to take reports or to investigate crime complaints, and/or using force against those reporting crime victimization.[15] The AAPL sought to achieve greater safety for Black civilians both by changing the behavior of white CPD officers ("They hope to teach their white counterparts

[14] The Chicago litigation is also featured in McCrary [2007]. We draw heavily upon the narrative history of these events in Pihos [2015].

[15] Renault Robinson, "A Backwards View of Problems," Chicago Defender, June 26, 1971, 30; Renault Robinson, "Beware the 'Big Trick,"' Chicago Defender, July 17, 1971, 4; Renault Robinson, "What's Wrong with City's Police Boss," Chicago Defender, April 10, 1972, 8.

that respect for the black community is essential in enforcing the law"),[16] and by advocating for the hiring and promotion of more Black police officers.

By 1973, the AAPL's leaders had not seen the CPD implement the changes they had been seeking. Renault Robinson and other Black AAPL leaders filed a federal employment discrimination claim against the CPD in April 1973; in May 1973 a second employment discrimination suit was filed against the CPD by Black and Hispanic CPD applicants.[17] A key claim of these complaints was that evidence of the CPD's discriminatory practices could be found not only in its hiring and promotion procedures and outcomes, but also in the CPD's lack of responsiveness to crime in Black neighborhoods. Renault Robinson editorialized in 1973 in the *Chicago Defender*, the city's Black newspaper, "The next time your precinct captain has the nerve to knock at your door and ask for your vote for King Daley remember what you are voting for: a police department that does nothing to stop serious crime in the black community – a police department that refuses to hire and promote black police officers – a police department that disrespects black people."[18] In August 1973, the Department of Justice (DOJ) filed its own employment discrimination suit against the CPD, the first time the DOJ had used its newly available powers under the 1972 amendments to Title VII to bring an employment discrimination case against a police department. The three lawsuits were consolidated the following year.

As later recounted in the District Court's 1976 opinion, the 1973 plaintiffs produced testimony not only on the disparate impact of the CPD's hiring and promotion procedures, but also on the Department's failure to protect Black civilians from crime: "We could take judicial notice of the tensions that have existed between white officers and black citizens in predominantly black neighborhoods, but we need not do so for that was the thrust of a good portion of the testimony of the witnesses called by the Robinson plaintiffs. Indeed, one of the avowed purposes of the Afro-American Patrolmen's League is to relieve those tensions through, inter alia, the recruitment and employment of more black officers to serve and protect the black citizenry which is so desperately in need of that service and protection" (United States v. City of Chicago, 411 F. Supp. 218 (N.D. Ill. 1976)).

The city's response to the 1973 filings was immediate. Police Superintendent James Conlisk resigned within two months of the DOJ filing. His interim replacement, James Rochford, was directed by the Chicago Police Board, the CPD's civilian oversight body, to consider the adoption of a variety of policies related to the CPD's hiring and promotion of Black and Hispanic officers, and to the department's policing practices in Black neighborhoods.[19] Upon taking office, Rochford immediately requested letters of resignation from all seventy CPD command officers. At his first press conference, he promised "a safer Chicago, control of street gangs, elimination of police misconduct, a reduction in street crime, police respect for citizens and the public's respect for policemen."[20] Rochford began to meet regularly with Black-led activist organizations.[21] Mayor Richard Daley likewise met with activists and promised to expand and diversify the Police Board.[22] These actions

[16] "Black Police League Has Tough Job Ahead," Sheryl Fitzgerald, Chicago Defender, September 14, 1968.

[17] The AAPL plaintiffs had filed an earlier suit in 1970 alleging retaliation for their organizing efforts; their 1973 filing expanded their 1970 complaint to include a generalized employment discrimination claim against the CPD.

[18] Renault Robinson, "Black Watch," Chicago Defender, June 16, 1973.

[19] William Mullen and Pamela Zekman, "Probe Brutality, Police Told: Board Orders Rochford to Take Immediate Steps," Chicago Tribune, November 16, 1973.

[20] Philip Wattley, "Rochford Starts Shakeup: 70 Top Cops Asked to Quit," Chicago Tribune, February 16, 1974.

[21] Citizens Alert, "Report to the Illinois Law Enforcement Commission: October 1, 1974 to September 30, 1975," 1975, Citizens Alert.

[22] Alan Merridew and George Bliss, "Daley Rips Police Brutality: Pledges Action to Combat It," Chicago Tribune,

were all taken in 1973 and 1974, well before the 1976 court order imposing a hiring and promotion plan on the Department.

The history of the litigation effort to end employment discrimination in the Chicago Police Department illustrates how plaintiffs could use to their advantage evidence of racial disparities in a department's response to crime victimization. The responsiveness of the CPD to the plaintiffs' allegations in the year that the litigation was filed likewise illustrates the incentives that departments had to address racial disparities in their responses to crime victimization once they faced the prospect of judicial intervention.

C Construction of NCVS Sample

The NCVS data used here include respondents from "core counties" within the forty MSAs included in the sample. The 50,000 housing units included in the NCVS are selected using a stratified, multi-stage cluster design. The Primary Sampling Units (PSU's) composing the first stage of the sample are counties, groups of counties, or large metropolitan areas. Large PSU's were included in the sample automatically, and each large PSU was assigned to its own stratum. These PSU's are considered to be self-representing (SR) since all of them were selected. The remaining PSU's, called non-self-representing (NSR) because only a subset of them were selected, were combined into strata by grouping PSU's with similar geographic and demographic characteristics, as determined by the Census, to design the sample. The design consists of 84 SR PSU's and 153 NSR strata [United States Department U.S. Dept. of Justice, 2007, pg. 8].

Housing units within a PSU are selected into the sample in two stages. The stages were designed such that, prior to any weighting adjustments, each sample housing unit had the same probability of being selected. The first stage involved selecting a sample of Enumeration Districts (ED's), geographic areas established for each decennial Census encompassing a population of 750 to 1,500 persons, from designated PSU's. ED's were systematically selected proportionate to their 1980 or 1990 population size. In the second stage, each selected ED was divided into segments using clusters of about four housing units each, and a sample of segments was selected. From this sample was compiled a list of addresses recorded during the 1980 and 1990 Censuses [United States Department U.S. Dept. of Justice, 2007, pg. 8-9].

Each month during the sample period the U.S. Census Bureau selected respondents for the NCVS using a "rotating panel" sample design. Households were randomly selected from the selected ED's and all age-eligible individuals became part of the panel. Once in the sample, respondents were interviewed every six months for a total of seven interviews over a three-year period. The first and fifth interviews were face-to-face; the rest were by telephone. After the seventh interview the household left the panel and a new household was rotated in to the sample.

Beginning in 1992, the NCVS categorized crime as personal or "property" covering the personal crimes of rape and sexual attack, robbery, aggravated and simple assault and purse-snatching/pick-pocketing; and the property crimes of burglary, theft, motor vehicle theft, and vandalism. [United States Department U.S. Dept. of Justice, 2007, pg. 6]. Therefore, in 1992 and 1993, the sample was split in half. Fifty percent of sample households were assigned to receive the continuing questionnaire and the other half was revised to the new questionnaire. The half-samples were designed to be as comparable as possible in terms of crime statistics. The 50-50 split with an 18-month overlap was designed to permit comparative analyses between both methods, to provide

Dec. 6, 1973, 1.

a basis for measuring the impact of the new methods on crime rates, and to lay the foundation for statistical adjustments to connect the historical and the new time series of crime rates [United States Department U.S. Dept. of Justice, 2007, pg. 9].

D Parallel Trends

Table D.1 reports estimates of the probability of treatment and, conditional on treatment, the year of treatment, as a function of 1970 log population, 1970 percent Black, 1970 median age, 1970 median family income (in thousands), 1970 median years of school, 1970 percent urban, and an indicator for whether an MSA experienced a riot between 1961 and 1968 [Spilerman, 1970, McCrary, 2007]. Law enforcement agencies that would eventually be litigated for racially discriminatory employment practices and subjected to post-litigation affirmative action plans are located in counties with larger percentages of Black residents in 1970 ($p < .05$ for all specifications), fewer median years of schooling in 1970 ($p < .05$ for agency-level specification), and higher median family incomes in 1970 ($p < .05$ for all specifications), relative to agencies that would never be litigated. By contrast, no covariates are significant in the models that predict treatment timing. These estimates are consistent with those reported by McCrary [2007].

Table D.1: Predicting Treatment and Treatment Timing

	Treatment			Treatment Timing		
	Agency	County	MSA	Agency	County	MSA
Log Population	0.09	0.06	0.11	0.35	-0.05	-0.04
	(0.07)	(0.07)	(0.11)	(0.68)	(0.65)	(1.44)
Pct Black	0.02**	0.02***	0.02**	-0.10	-0.03	-0.11
	(0.01)	(0.01)	(0.01)	(0.11)	(0.09)	(0.25)
Median Age	0.01	-0.00	0.01	-0.23	-0.20	-0.17
	(0.02)	(0.02)	(0.03)	(0.18)	(0.18)	(0.28)
Median Yrs School	-0.18**	-0.16	-0.07	1.10	0.76	-0.11
	(0.08)	(0.08)	(0.17)	(1.17)	(0.92)	(3.04)
Median Fam Income	0.11**	0.14***	0.14**	-0.98	-0.41	-0.46
	(0.05)	(0.04)	(0.06)	(0.72)	(0.65)	(0.98)
Pct Urban	-0.00	-0.00	0.00	0.03	0.02	0.10
	(0.00)	(0.00)	(0.01)	(0.05)	(0.04)	(0.15)
Riot 1961-1968	0.13	0.03	0.15	-1.63	-1.62	-1.03
	(0.21)	(0.18)	(0.20)	(1.79)	(1.38)	(2.08)
N	149	81	37	108	60	26

** p<.05, *** p<.01. Median family income in 1,000s. All models estimated with OLS. Models predicting treatment timing are restricted to only those agencies that will eventually be subjected to treatment. Standard errors clustered on MSA for agency and county models.

Table D.2 reports estimates of the probability of treatment and, conditional on treatment, the year of treatment, using the sample described in Table 1. There are several correlations between respondent-level pretreatment covariates and whether an MSA will experience litigation over race-based employment discrimination in law enforcement. The joint p-values are < 0.001 both for the model without interactions and for the model interacting covariates with respondent race. These correlations disappear, however, for the models predicting treatment timing. The joint p-values for these models are 0.11 for the model without interactions and 0.33 for the model interacting covariates with respondent race.

Table D.2: Predicting Treatment and Timing of Litigation Leading to Affirmative Action NCVS Sample, Respondent Level

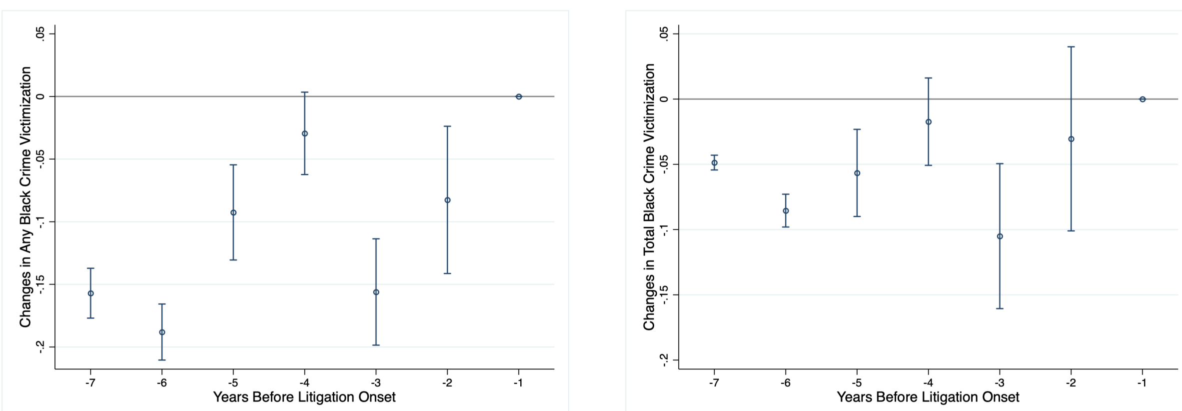
	DV: Treatment		DV: Treatment Timing	
Black	0.19	0.33	0.86	1.32
	(0.12)	(0.16)	(0.52)	(0.62)
Homeownership	0.12**	0.11**	-0.20	-0.20
	(0.05)	(0.05)	(0.23)	(0.24)
Single Family Home	0.03	0.06	0.60	0.74
	(0.05)	(0.05)	(0.32)	(0.34)
Age 18-29	0.03**	0.03***	-0.05	-0.02
	(0.01)	(0.01)	(0.06)	(0.07)
Household Income 30K +	0.00	-0.00	0.67	0.67
	(0.05)	(0.05)	(0.36)	(0.36)
Some College	-0.08**	-0.08**	-0.08	-0.08
	(0.03)	(0.03)	(0.08)	(0.08)
Married	-0.01	-0.01	-0.07	-0.05
	(0.01)	(0.01)	(0.06)	(0.05)
Female	-0.01	-0.01**	0.01	0.01
	(0.01)	(0.01)	(0.02)	(0.02)
Total Victimization	0.01	0.01	0.07	0.06
	(0.01)	(0.01)	(0.05)	(0.04)
Black X Homeownership		0.09		0.05
		(0.08)		(0.28)
Black X Single Family Home		-0.29***		-0.67
		(0.04)		(0.32)
Black X Age 18-29		-0.05		-0.10
		(0.02)		(0.14)
Black X Household Income 30K+		0.07**		-0.17
		(0.03)		(0.31)
Black X Some College		-0.01		0.09
		(0.03)		(0.11)
Black X Married		-0.03		-0.08
		(0.03)		(0.11)
Black X Female		0.03***		-0.00
		(0.01)		(0.04)
Black X Total Victimization		0.04***		0.06
		(0.01)		(0.12)
Constant	0.28	0.26	1984.26***	1984.14***
	(0.14)	(0.13)	(1.06)	(1.04)
N	85,157	85,157	32,699	32,699
Joint F-Statistic	9.90	122.14	2.02	1.59
P-Value for Joint F-Statistic	0.00	0.00	0.23	0.31

** p<.05, *** p<.01. Standard errors clustered on MSA.

One potential concern is that Black victimization rates were already trending downward prior to litigation onset in not-yet-litigated MSAs. If so, then post-litigation decreases in Black crime victimization may have simply represented the continuation of preexisting secular trends. Although Figures 2, 3, and 4, plotting raw victimization rates, do not reveal pre-litigation downward trends in Black crime victimization, we can more systematically explore the potential for pre-trends in Black crime victimization.

A common test for pre-trends is to implement a dynamic difference-in-difference estimator, with indicators for leads and lags of a treatment indicator, and then to examine the coefficients on the indicators for treatment leads (pre-treatment periods), which report changes in the outcome of interest relative to the last period prior to treatment. In our case, we are looking for significant positive coefficients on pre-litigation period indicators, which would indicate that Black crime victimization was higher in pre-periods before the last pre-period, suggesting that Black crime victimization was already decreasing prior to litigation onset.

Figure D.1 reports the coefficient estimates for treatment leads from dynamic TWFE DD specifications, using a sample restricted to those MSAs and years for which pre-litigation data are available between 1979 and 1985, and including year and MSA fixed effects and clustering standard errors on MSAs. All pre-treatment coefficients are negative; there is no evidence of a downward trend in Black crime victimization prior to litigation onset. Instead, there is suggestive evidence that Black crime victimization in not-yet-litigated MSAs was trending upward prior to litigation onset.



(a) Total Black Crime Victimization

(b) Any Black Crime Victimization

Figure D.1: Estimated Changes in Black Crime Victimization, 1979-1985
Dynamic TWFE DD; Pre-Litigation Years Only

Row 1 of Table D.3 reports the estimated slopes and standard errors for the lines of best fit through the pre-litigation coefficients for Black crime victimization reported in Figure D.1. Estimated pre-trends are indeed upward sloping, although not significant at $p < .05$. As discussed in more detail in Appendix G, Sun and Abraham [2020] report the potential for bias in dynamic TWFE DD specifications that do not account for treatment heterogeneity. Row 2 of Table D.3 reports the estimated slopes and standard errors for the lines of best fit through the pre-litigation coefficients for Black crime victimization estimated using the dynamic Sun and Abraham [2020] specification, again restricting the sample to those MSAs and years for which pre-litigation data are available between 1979 and 1985, including year and MSA fixed effects, and clustering standard errors on MSAs. Estimated pre-trends are again upward sloping although not significant.

Roth [2021a] suggests estimating linear violations of the parallel trends assumption that are possible within the range of estimation uncertainty for treatment leads and lags, but that might be missed by standard tests for pre-trends. We use the estimated coefficients on treatment leads and lags and variance-covariance matrices from the TWFE DD specifications reported above to estimate the linear violations of parallel trends that would be detected only 80% and 50% of the time by standard tests (i.e., a significant pre-treatment coefficient in the opposite direction as post-treatment coefficients). Rows 3 and 4 of Table D.3 report these estimated pre-trends. They are again positive, although flatter than the pre-trends estimated using the point estimates for pre-litigation change in Black crime victimization rates. There is again no evidence of pre-litigation downward trends in Black crime victimization.

Table D.3: Pre-Litigation Trends in Black Crime Victimization, 1979-1985

	Total Victimization		Any Victimization	
	β	SE	β	SE
TWFE DD	0.022*	(0.011)	0.007	(0.007)
Sun and Abraham [2020]	0.020	(0.019)	0.022	(0.016)
Roth [2021a,b] 80%	0.003		0.001	
Roth [2021a,b] 50%	0.002		0.001	

Rows 1 and 2 report estimated slopes and standard errors for the line of best fit through the pre-litigation coefficients on Black crime victimization. Dynamic TWFE DD and Sun and Abraham [2020] specifications include year and MSA fixed effects and cluster standard errors on MSAs. Rows 3 and 4 report the estimated slopes of linear violations of parallel trends that would be detected 80% and 50% of the time, respectively, by conventional pre-trends tests, using the variance-covariance matrices from the TWFE DD estimates reported in row 1 [Roth, 2021a,b].

E McCrary (2007)

McCrary [2007] reported that employment discrimination litigation had no discernible impacts on offenses known to law enforcement. However, the FBI's Uniform Crime Reporting (UCR) data used in McCrary [2007] do not report victim race. As reported in Table 3, there are no discernible effects of litigation on victimization experienced by non-Hispanic white respondents. The inability to differentiate victim race in the UCR data may obscure racially heterogeneous treatment effects of litigation.

Further, although we are unable to identify reporting effects given our data and design, any post-litigation increases in reporting would mask decreases in victimization when analyzing data on reported victimization. Miller and Segal [2018] suggested that litigation leading to gender-based affirmative action in law enforcement increased the rate at which female victims reported gender-based violence to law enforcement agencies, conditional on an incident of violence occurring, and attributed post-litigation decreases in the incidence of gender-based violence at least in part to this increased reporting. It is likewise possible that employment discrimination litigation, by increasing the responsiveness of police departments to Black crime victimization, increased the trust of Black crime victims in the likely police response, thereby increasing their willingness to report victimization to the police.

It is also possible that increased effort devoted by law enforcement agencies to Black crime victimization could increase the rate at which victimizations reported to the police are actually recorded as criminal offenses. Law enforcement agencies do not record all victimization reports as criminal offenses. In one study in which independent observers accompanied law enforcement officers responding to 911 calls, by the time that officers arrived on the scene, victims were no longer present for approximately 33% of calls. These calls were not recorded as criminal offenses. Of the calls coded by observers as involving criminal incidents in which a victim was present, officers failed to report the incident as a criminal offense in another 36% of cases [Black, 1970]. Calls with longer response times may be less likely to be recorded as criminal offenses [Asher, Jan 29, 2018]. Officers may be slower to respond to calls originating in less white neighborhoods, leading to fewer victimization reports from less white neighborhoods being recorded as criminal offenses, relative to victimization reports from more white neighborhoods.[23]

The discretion of law enforcement agencies to record complaints as criminal offenses (or not) may also generate positive bias in estimates of the effects of litigation onset on offenses known to law enforcement. If, after the onset of litigation leading to affirmative action, agencies decreased response times to calls involving Black victims, and/or were more likely to record incidents involving Black victims as criminal offenses, then a larger number of victimization reports may have been recorded as criminal offenses post-litigation [Levitt, 1998, Vollaard and Hamed, 2012].[24] Post-litigation increases in the proportion of victimizations recorded as "offenses known to law enforcement" would also generate upward bias in estimates of the effects of litigation leading to affirmative action on offenses known to law enforcement.

As reported in the first plot of Figure E.1 for litigated MSAs, using a dynamic TWFE DD specification to better enable comparisons to McCrary [2007], the effects of litigation on changes in victimization reported to law enforcement (pooling Black and white NCVS respondents) are indistinguishable from zero for several years after the onset of litigation, and are negative and significant but small in magnitude after that.

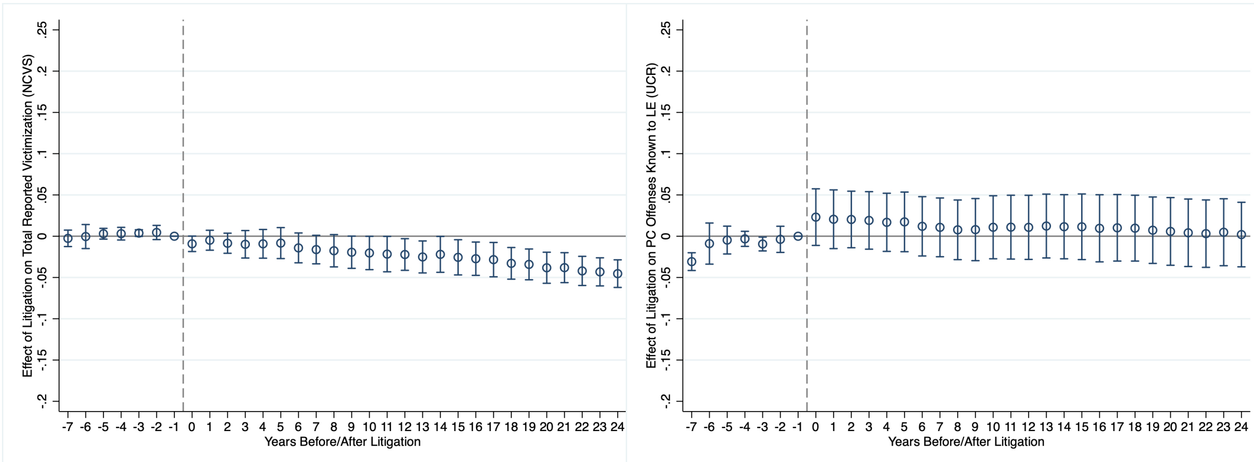
The second plot in Figure E.1 reports estimates of the effects of litigation on changes in offenses known to law enforcement, using the UCR data for agencies located in the core counties of the 26 treated MSAs. The FBI's UCR program is voluntary; not all agencies participate, or participate consistently, in the program. Offenses known to law enforcement in the UCR data include homicide, rape, robbery, aggravated assault, burglary, theft, and motor vehicle theft. We eliminate the homicide counts in order to better approximate the offenses that could have been reported to law enforcement by victims. We match the 108 agencies located in our sample of 26 treated MSAs to the agencies reporting monthly offense data in the UCR, as cleaned and aggregated to the yearly level by Kaplan [2019]. 87 of these agencies consistently report UCR data.[25] We aggregated total

[23] https://www.aclu-il.org/en/press-releases/newly-released-data-shows-city-continues-deny-equitable-police-services-south-and.

[24] In 1983, the FBI began asking law enforcement agencies to report complaints that were determined by agencies not to involve criminal offenses as "unfounded" complaints. Unfortunately, we do not have enough pretreatment years of data on these unfounded complaints to systematically analyze this potential source of measurement error.

[25] Agencies that report zero crimes in any year are treated as missing data in that year. We interpolated isolated missingness in the annual total offense series between 1979 and 2004.

offenses and population served from the agency level to the MSA level, and constructed per capita offenses known to law enforcement at the MSA/year level from these aggregated data.



(a) NCVS Reported Victimization

(b) UCR Reported Offenses

Figure E.1: Estimated Effects of Litigation Leading to Affirmative Action on Changes in Reported Victimization (NCVS) and Offenses Known to Law Enforcement (UCR), 1979-2004. Dynamic TWFE DD estimators with MSA and year fixed effects; standard errors clustered on MSAS, treated sample only.

As reported in Plot 2 of Figure E.1, estimates of the effects of litigation leading to affirmative action on changes in offenses known to law enforcement are very similar to the estimates from the NCVS data, albeit noisier and slightly more positive on average. The somewhat less negative estimates of the effects of litigation on changes in offenses known to law enforcement, relative to changes in reported victimization, are consistent with the potential sources of positive bias described above. Consistent with the estimates reported by McCrary [2007], all of the post-litigation estimates in Plot 2 are indistinguishable from zero at the 95% confidence level.

F Causal Mechanisms

Table F.1: Estimated Effects of Litigation on Officers Per 100K Population, Arrest Rates, and Clearance Rates
Treated-Only Sample

	(1) TWFE DD 1979-2004	(2) TWFE DD 1979-1985	(3) Sun and Abraham [2020]
Officers/100K Population	20.31	-1.72	32.83
	(21.16)	(16.46)	(16.54)
Arrests/100K Population	4931.06	131.39	-489.16
	(4025.96)	(492.56)	(362.05)
Black Arrests/100K Black Population	-2014.42	-4372.61	-728.10
	(2385.27)	(2684.01)	(1879.28)
White Arrests/100K White Population	661.55	-203.81	-216.22
	(388.74)	(202.65)	(310.38)
Avg Clearance Rate	-0.02	0.04	0.04
	(0.01)	(0.02)	(0.03)
N	676	35	35
MSA FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

* p<.10, ** p<.05, *** p<.01. All models include year and MSA fixed effects and cluster standard errors on MSAs. Models (2) and (3) are estimated using data only from MSAs litigated after 1979, during the years 1979 - 1985.

G Technical Appendix

The following section summarizes the estimators used in Tables 3 - 6.

G.1 TWFE Estimator

The TWFE estimator of the average treatment effect of an absorbing binary treatment D_{irt} on an outcome Y_{irt} in a repeated cross-section of individuals i in regions r in periods t, where all regions are eventually treated, is defined as:

$$Y_{irt} = \lambda_r + \delta_t + \gamma D_{irt} + \epsilon_{irt}$$

where λ_r are region fixed effects and δ_t are time fixed effects. We estimate γ using OLS with standard errors clustered on regions (MSAs).

G.2 Borusyak et al. [2021] ATT Estimator

Borusyak et al. [2021] define an estimator of causal effects of an absorbing binary treatment D_{irt} on some outcome Y_{irt} in a repeated cross-section of individuals i in regions r in periods t, where all regions are eventually treated. The set of treated observations is defined by $\Omega_1 = \{irt \in \Omega : D_{irt} = 1)$ of size N_1 and the set of not-yet-treated observations by $\Omega_0 = \{irt \in \Omega : D_{irt} = 0)$ of size N_0. $Y_{irt}(0)$ is the potential outcome without treatment. The causal effect of interest is $\tau_{irt} = Y_{irt} - Y_{irt}(0)$. The overall average treatment effect on the treated (ATT) is estimated using the following steps:

1. Estimate a model for non-treated potential outcomes using the not-yet-treated observations only, including region λ_r and time δ_t fixed effects and using OLS:

$$Y_{irt} = \lambda_r + \delta_t + \epsilon_{irt}$$

2. Extrapolate the model from Step 1 to treated observations, excluding always-treated observations and observations in years for which there are no not-yet-treated observations, imputing non-treated potential outcomes for each treated observation, and obtain treatment effects for each treated observation: $\hat{\tau_{irt}} = Y_{irt} - \hat{Y_{irt}}(0)$;

3. Estimate the ATT as the average of $\hat{\tau_{irt}}$ for treated observations $= 1/N_1 \sum_{irt \in \Omega_1} \hat{\tau_{irt}}$.

We estimate the overall average treatment effect on the treated (ATT) using the package developed by Borusyak [2021], clustering standard errors on MSAs.

G.3 Callaway and Sant'Anna [2020] ATT Estimator

The Callaway and Sant'Anna [2020] ATT estimator is defined for T periods, where $t = 1, ..., T$, Y_{it} is the outcome of interest, and D_{it} is a binary variable equal to 1 if a unit is treated in period t and 0 otherwise. G_g is defined as a binary variable equal to 1 when a unit is first treated in period g; treatments are assumed to be irreversible. Timing group/time average treatment effects are defined as:

$$ATT(g,t) = E\left[Y_t - Y_{g-1}|G_g = 1\right] - E\left[Y_t - Y_{g-1}|D_t = 0\right] \tag{1}$$

These timing group/time ATTs are defined only for timing groups with pretreatment outcomes (Y_{g-1}), excluding always-treated timing groups, and only for posttreatment periods for which there exist not-yet-treated observations ($D_t = 0$), excluding the period in which the last timing group is treated and all periods thereafter. Timing group average treatment effects are defined as:

$$\theta(g) = \frac{1}{T - g + 1} \sum_{t=g}^{T} ATT(g,t) \tag{2}$$

This parameter reports treatment effect heterogeneity with respect to treatment adoption period. Callaway and Sant'Anna [2020] then construct θ^O, which is the overall average treatment effect on the treated (ATT), as:

$$\theta^O = \sum_{g \in G} \theta(g) P(G = g | G \leq T) \tag{3}$$

We estimate the overall ATT using the doubly robust estimator developed by Sant'Anna and Zhao [2020] for repeated cross-sections, with robust standard errors clustered on MSAs, using the package developed by Rios-Avila and Sant'Anna [2021].

G.4 de Chaisemartin and D'Haultfœuille [2021] ATT Estimator

de Chaisemartin and D'Haultfœuille [2021] use panels of groups, indexed by g, that are exposed to a binary absorbing treatment $D_{i,g,t}$, which represents the treatment of observation i in group g at period t. Their $DID_{t,l}$ estimator identifies the effect of treatment on an outcome $Y_{i,g,t}$ for units in groups g and period t switching from untreated in period $t-l-1$ to treated in period $t-l$, using units that remain untreated between period 1 and period t as controls. $F_{g,1}$ is the first period in which group g is treated; $N_{g,t}$ is the number of observations in group g at period t; $N^1_{t,l}$ is the number of units treated for the first time $l \geq 0$ periods ago at t; and N^{nt}_t is the number of observations in groups untreated from period 1 to t.

$$DID_{t,l} = \sum_{g:F_{g,1}=t-l} \frac{N_{g,t}}{N^1_{t,l}} (Y_{g,t} - Y_{g,t-l-1}) - \sum_{g:F_{g,1}>t} \frac{N_{g,t}}{N^{nt}_t} (Y_{g,t} - Y_{g,t-l-1})$$

$DID_{t,l}$ can only be estimated for units in groups that have pretreatment outcomes ($Y_{g,t-l-1}$), excluding units in always-treated groups, and only in periods with units that have been untreated from period 1 to t ($N^{nt}_t > 0$), excluding periods in which all units have been treated. de Chaisemartin and D'Haultfœuille [2021] then define the average treatment effect for units switching from untreated in period $t-l-1$ to treated in period $t-l$ as:

$$DID_l = \frac{1}{N^1_l} \sum_{t=l+2}^{T} N^1_{t,l} DID_{t,l}$$

Finally, de Chaisemartin and D'Haultfœuille [2021] define the overall average treatment effect for treated units as a weighted average of the DID_l estimators, giving to each estimator a weight proportional to the number of units DID_l applies to. We estimate the overall ATT using the package developed by de Chaisemartin et al. [2021], with bootstrapped standard errors clustered on MSAs.

G.5 Sun and Abraham [2020] ATT Estimator

Sun and Abraham [2020] define an interaction-weighted estimator for cohorts e containing units i observed over time periods t first receiving a binary absorbing treatment $D_{i,t}$ at time Ei. Sun and Abraham [2020] first estimate cohort- and period-specific average treatment effects $CATT_{e,l}$ for a sample with no never-treated units using a linear two-way fixed effects specification interacting relative period indicators l with cohort indicators e, excluding indicators from the latest-treated cohort C, and excluding units in always-treated cohorts:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{e \notin C} \sum_{l \neq -1} \delta_{e,l}(\mathbf{1}\{E_i = e\} \cdot D^l_{i,t}) + \epsilon_{i,t}$$

Unit and time fixed effects are represented by α_i and γ_t respectively. Each $\delta_{e,l}$ represents an estimate of the average treatment effect l periods from the initial treatment for the cohort of units first treated at time $E_i = e$. Sun and Abraham [2020] then estimate weights $Pr\{E_i = e | E_i \in [-l, T - l]\}$ by sample shares of each cohort in the relevant periods l. To construct the overall ATT, they then form a weighted average of the estimates of $CATT_{e,l}$ using these weight estimates. We estimate the overall ATT using the package developed by Sun [2021], with standard errors clustered on MSAs.