# Wage Caps and Police Performance: Evidence from Arbitration Reform

Pim Pinitjitsamut*

October 2026

**Abstract**

This paper examines whether fiscal constraints on public-sector compensation reduce police performance, using New Jersey's two percent cap on police arbitration awards (2011–2017). I assemble a new thirty-year dataset of police contracts that pinpoints each agency's contract expiration date, which determines when the cap became binding. The cap bound. Across 460 agencies it reduced negotiated salary schedule increases by 1.99 percentage points against a pre-cap mean of 3.6 percent, and it lowered the entry wage written into contracts by about 4.5 percent. Staggered difference-in-differences estimates of the effect on crime clearance rates are negative but small and imprecise, declines of 1.9 to 2.6 percentage points that are not statistically distinguishable from zero. The muted response is not explained by officers being made whole elsewhere in the contract. The compressed value does not reappear in the uncapped components, such as uniform allowances, tuition stipends, and leave, and at least four fifths of it is unaccounted for on the margins contracts record.

**JEL Classification:** J45, J52, H72, K31
**Keywords:** police performance, arbitration, public sector unions, compensation, wage caps

*Postdoctoral Fellow, National Bureau of Economic Research. Email: pp804@rutgers.edu. I would like to express my sincere gratitude to Anne Morrison Piehl for her invaluable guidance and support. I also appreciate the insightful comments and feedback from Amanda Agan, Hector Blanco, Joseph Fry, and Emily Owens. Additionally, I thank the participants of the Applied Microeconomics Seminars at Rutgers and the Applied Micro Breakfast Seminar, where I had the opportunity to present this research. All errors are my own.

# 1 Introduction

Police unions negotiate comprehensive contracts that extend far beyond traditional wages and hours to include disciplinary procedures, complaint investigation protocols, and use of force guidelines (Rushin, 2016; Huq and McAdams, 2016). In occupations like policing where discretion is high and performance is difficult to observe, these collective agreements directly shape both working conditions and standards of accountability. Police unions are a pervasive feature of municipal law enforcement, shaping the employment terms of officers across thousands of agencies nationwide. Many states mandate binding interest arbitration for police contract disputes (Anderson and Krause, 1987; Feuille and Delaney, 1986; Ashenfelter and Hyslop, 2001).

Police arbitration provides an unusually informative setting for studying how compensation determination affects worker behavior. Unlike private sector workers who can strike to pressure employers during negotiations, police officers in most jurisdictions are prohibited from striking because work stoppages would create public safety crises. This prohibition necessitates alternative dispute resolution mechanisms when bargaining reaches impasse (Anderson and Krause, 1987). In effect, binding interest arbitration has become the police substitute for the right to strike. When the parties cannot reach a voluntary agreement, a neutral arbitrator imposes the contract terms (Anderson and Krause, 1987; Ashenfelter and Hyslop, 2001). Under final offer arbitration, the arbitrator must select one party's complete final offer, which creates incentives for parties to submit reasonable proposals (Farber, 1980).

Mas (2006) shows that these arbitration outcomes affect police performance. Studying New Jersey contracts from 1978 to 1996, he finds that when employers win in final offer arbitration, so that raises fall short of what unions proposed, clearances per capita decline by roughly 5 to 8 percent.

This pattern suggests that officers retain considerable latitude over how intensively they pursue cases, follow up on leads, and engage with investigative work. When they feel un-

fairly compensated, they withdraw effort on these discretionary margins, and clearance rates decline as a consequence. This behavioral response aligns with broader evidence on reciprocity in employment relationships, where workers reduce effort when they perceive unfair treatment (Akerlof and Yellen, 1990; Fehr et al., 2009; Card et al., 2012; Gneezy and List, 2006). Despite growing attention to police union contracts, we know relatively little about how officers respond when arbitration institutions themselves change. This raises the question of whether predictable, uniform constraints produce the same behavioral responses as individual disappointments. When everyone knows raises will be limited to two percent, officers can adjust their expectations in advance rather than feeling blindsided when outcomes fall short (Koszegi and Rabin, 2006; Abeler et al., 2011; Crawford and Meng, 2011).

This distinction matters for policy. If wage compression always reduces police effort, fiscal reforms carry hidden costs through diminished public safety. But if predictable constraints operate differently than unexpected disappointments, then how constraints are designed and communicated becomes critical. Transparency and uniformity may prevent the perceived unfairness that triggers effort withdrawals.

In this paper, I examine how a statutory cap on police arbitration awards affected officer behavior in New Jersey. Between 2011 and 2017, New Jersey implemented the Interest Arbitration Reform Act (Chapter 105), which limited salary increases awarded through arbitration to two percent per year. The reform compressed wage growth by over 50 percent for agencies proceeding to arbitration, reducing average awards from 3.9 percent annually before 2011 to 1.85 percent during enforcement.[1] Using staggered difference-in-differences and drawing on thirty years of arbitration records matched to crime clearance data from 463 agencies, I find small negative effects on police performance that are not statistically significant. Under the preferred specification, clearance rates decline by 1.9 to 2.6 percentage points, or about 9 to 13 percent of pre-cap baselines. The confidence intervals are wide,

[1]These award figures, and the union-offer figures reported below, are nominal. Because consumer prices rose over 2011–2017, the cap's bite in real terms exceeded the nominal compression; Figure 3b shows the inflation-adjusted path, in which capped raises produced periods of negative real wage growth. I report nominal magnitudes in the text and defer the full real-terms treatment of inflation to that figure.

running from gains of about 3 percentage points to declines of about 7 to 8, so while the point estimates are consistently negative across estimators and comparison groups, the magnitude remains undetermined.

New Jersey provides an unusually rich setting for studying these dynamics. The state has required binding interest arbitration for police and firefighter contract disputes since 1977, creating a long institutional history. Final offer arbitration governed these disputes through 1996, the period Mas (2006) studies; since the 1996 reform, arbitrators have used conventional arbitration, setting awards on a continuum rather than choosing one party's package. Unlike states where arbitration is occasional, New Jersey sees frequent use, with approximately 40 percent of police contract renewals proceeding to arbitration rather than voluntary settlement. I construct a novel dataset combining arbitration records spanning 1990–2020, obtained from the New Jersey Public Employment Relations Commission through public records requests and supplemented with contracts from the New Jersey State Policemen's Benevolent Association. Each contract was manually reviewed and coded to extract final offers from both parties, arbitrator awards, contract expiration dates, wage schedules, and negotiation delays. I match these records to clearance and crime data from the FBI's Uniform Crime Reporting Program (Kaplan, 2021) and supplement with enforcement data from the New Jersey Attorney General's public portal, creating the first systematic dataset linking arbitration outcomes to police performance over an extended period.

My primary identification strategy exploits within-state variation in contract expiration timing through staggered difference-in-differences, using estimators robust to treatment effect heterogeneity (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). Agencies with contracts expiring in 2011 faced the cap immediately, while those with contracts expiring later operated under the old regime for several additional years, and each agency enters as its own treated unit compared against the agencies not yet treated. I supplement this with synthetic difference-in-differences at the state level (Arkhangelsky et al., 2021), comparing New Jersey's clearance rate trajectory to a weighted combination of control states without

arbitration caps. Because the synthetic control approach aggregates New Jersey into a single treated series, it does not exploit the agency-level timing variation that the staggered design uses, so I read it as an aggregate check on the within-state estimates rather than as the primary evidence.

The state-level synthetic difference-in-differences comparison, reported in Appendix 8.1, points in the same direction, with a small and statistically insignificant decline of about 1.4 percentage points in clearance rates.

The cap plainly bound. Using the negotiated base salary schedules of 460 municipal agencies, not only those that reached arbitration, I find that it reduced annual scheduled increases by 1.99 percentage points against a pre-cap mean of 3.6 percent, and raised the probability that an agency's increase fell below 2 percent by 71 percentage points, with no detectable pre-trend. The null result on clearance is therefore not the null of a policy that failed to take hold.

One possible reason the response was small is that officers were not very upset by a cap they saw coming. Union final offers, which averaged 4.8 percent before the cap, fell to 2.4 percent afterward, and 93 percent of post-cap awards fell at or below the 2 percent ceiling, against 10 percent before the cap. Unions lowered their demands toward the constraint. When raises are known in advance to be capped, officers may be less likely to experience the surprise and perceived unfairness that drive the effort withdrawals documented after idiosyncratic arbitration losses.

I test this explanation in two ways, and neither supports it. If the size of the felt loss governs the response, agencies that lost more should respond more. Agencies whose capped raise fell far below the raises of uncapped neighbors in the same county show no larger decline in clearance than agencies whose raise was close to their neighbors'. Among arbitration cases, awards that fell further below the union's final offer are not followed by larger declines. Both tests have limited power and cannot rule out a modest gradient. The comparison with individual arbitration losses points the same way. An award closer to the

employer's offer is associated with a clearance rate 2.2 percentage points lower, about 7.7 percent of baseline and only marginally significant, which is no larger than the 9 to 13 percent declines estimated for the cap. Both estimates are imprecise and come from different designs, but the evidence does not show that a predictable cap produced a smaller response than an unexpected individual defeat.

This setting does not provide direct measures of officer expectations, so the expectation account cannot be tested more sharply here. The small response is also consistent with compensation not having fallen in net at all. If the value squeezed out of base salary reappeared elsewhere in the contract, officers were never made worse off, and no performance response would be expected. A related possibility works through capacity rather than effort. If municipal budgets kept growing while base-salary growth was capped, the freed funds could support more officers or more paid hours, so aggregate clearance could hold steady even if individual officers withdrew some effort. Distinguishing these explanations requires looking inside the contract and at the wider budget, and I take up the within-contract question directly in Section 5.

Using the same contract panel and identification but a new family of outcomes drawn from the structure of the compensation package, I find little sign of that reallocation. The cap compresses the entry wage written into contracts by about 4.5 percent, the margin it was designed to bind. That compression does not flow into the uncapped non-salary categories, such as uniform allowance, tuition stipends, and leave, where the estimates are within their standard errors of zero and the null persists when missing benefit values are treated as continuing under the contracts' continuation clauses rather than as removed. A simple accounting exercise puts at least four fifths of the value the cap removed outside every margin the contract records. Supervisory wage levels rise modestly, but those estimates are sensitive to small departures from parallel trends, so I read them as suggestive. Officers were therefore not plausibly made whole through the contract's uncapped margins, which narrows the explanations for the small and imprecise performance response to the ones the contract

data cannot see.

These findings contribute to the literature in four ways. First, I study police arbitration under a different rule. Since 1996, New Jersey has resolved police contract disputes through conventional arbitration, in which the arbitrator may set the award anywhere between the two offers, and from 2011 to 2017 a statute limited those awards to two percent per year. I examine how officers respond when the arbitration rule itself changes. I measure the response with the clearance rate, the share of reported crimes that officers solve, alongside clearances per capita, because the rate conditions out exogenous movements in crime volume and isolates the investigative-effort margin that officers actually control. This complements recent work showing that collective bargaining rights adoption increases violent incidents of officer misconduct (Dharmapala et al., 2022), affects policing practices and civilian deaths (Cunningham et al., 2021), and influences accountability mechanisms (Goncalves, 2021). My focus is on how compensation is determined, rather than on whether unions exist or contracts are generous.

Second, I add to our understanding of how workers respond to wage changes. Economic theory predicts that pay cuts reduce effort (Akerlof and Yellen, 1990; Fehr et al., 2009; Shapiro and Stiglitz, 1984), and plenty of evidence supports this prediction (Mas, 2008). My findings are consistent with this, since the point estimates are negative, although none is statistically significant. Prior work suggests that the size of the response may depend on whether workers see the cuts coming. Announced, uniform cuts often have modest effects (Kaur, 2019; Bewley, 1999), while idiosyncratic or surprising cuts generate strong negative responses (Card et al., 2012; Breza et al., 2018; Kube et al., 2012). In this setting I do not find that distinction. The cap and individual arbitration losses produce declines of similar size, and the response does not grow with the size of the visible loss. A gap between how a contract term is perceived and how it actually bears on officers also appears in Bollman et al. (2024), who find that mandatory overtime, long contested by police unions as a costly imposition, produces no measurable deterioration in officer wellness and only modest reductions in low-

level enforcement. That a provision's contested valuation can diverge from its measured effect reinforces the case for studying perceived, not merely objective, compensation outcomes.

Third, I provide detailed evidence about how public-sector arbitration actually works in practice. Building on early work by Zax and Ichniowski (1988), who documented that public-sector unionism substantially influenced municipal pay, employment, and expenditures (see also Valletta, 1993), the contract data I assembled allowed me to identify the precise timing of when each agency's contract expired, creating the variation between early and late adopters that drives my empirical strategy. Using thirty years of these records, I document patterns in arbitration use and award characteristics. The same records let me measure the cap's first stage on all agencies in the sample, not only those that reached arbitration. This institutional detail matters because bargaining processes, not just outcomes, may shape organizational performance. Freeman (1986) distinguished unions' monopoly face, which raises wages, from their voice face, which facilitates communication (Freeman and Medoff, 1984), while Anzia and Moe (2015) examines how public-sector unions shape resource distribution through institutional design.

Fourth, I add to work on how public-sector compensation absorbs fiscal constraints (Gittleman and Pierce, 2012; Clemens and Cutler, 2014). Standard compensation-spillover reasoning predicts that when a binding constraint lowers one priced margin while many adjacent margins remain free, bargaining should redirect value toward the unconstrained margins (Rosen, 1986; Woodbury, 1983; Gruber, 1994). New Jersey police contracts are close to an ideal setting for this test, because the cap binds the base-salary line and leaves uniform allowance, tuition stipends, and leave accrual untouched. The predicted substitution does not appear, and the confidence intervals are tight enough to rule out anything beyond small movements.

Taken together, my findings trace how a specific institutional constraint operates, from whether it bound to where the compressed pay went. This relates to recent work showing that predictable, rule-based changes in public employment generate different responses than

discretionary decisions (Biasi, 2021), and that institutional context mediates union impacts on productivity (Frandsen, 2021; Hirsch, 2004). More broadly, how fiscal constraints are designed may matter as much as whether they are implemented, a distinction with important implications for public sector labor markets facing budget pressures. This paper shows how to measure whether such a constraint bound and what it did to the rest of the compensation package.

# 2 Data

## 2.1 Crime and Clearance Data

The study draws on clearance and crime data from the Federal Bureau of Investigation's Uniform Crime Reporting (UCR) Program and the New Jersey Office of the Attorney General. For this project, I use a cleaned version of the UCR clearance dataset prepared and shared by Jonathan Kaplan. This dataset reports computed clearance rates at the agency-year level and includes both violent and property crimes.

I merged the UCR data with the union contract and arbitration datasets using unique Originating Agency Identifier (ORI) codes assigned to police agencies. Because ORI codes in the contract and arbitration data were not always standardized, I constructed a concordance table to align inconsistent naming conventions and correct typographical differences, ensuring accurate merges across datasets.

In addition to UCR data, I obtained enforcement data from the New Jersey Attorney General's public portal. This data includes incident-level reports on police activity and is available for more recent years. I used these records to validate the consistency of arrest and clearance patterns derived from the UCR and to construct additional enforcement outcomes where available.

## 2.2 Union Contract Data

I collected a novel dataset from multiple sources to ensure comprehensive coverage of police officer-level union contracts across 480 agencies in New Jersey. The primary source was the Public Employment Relations Commission (PERC), the official state agency that oversees labor relations. I scraped and downloaded all publicly available contract PDFs from the PERC website, covering municipal law enforcement agencies across years. Each document represented an individual bargaining agreement, often over 40 pages, formatted differently by municipality and bargaining unit.

To supplement and verify this collection, I contacted Michael Freeman, Executive Vice President of the New Jersey State Policemen's Benevolent Association and a central figure in police union negotiations. Officer Freeman shared his personal collection of union contracts, which he had assembled independently from agencies. I cross-referenced the PERC downloads with Freeman's dataset to identify missing contracts. Officer Freeman confirmed that the observed omissions were due to municipalities failing to submit documents, not from systematic exclusion.

Each contract PDF was examined line-by-line by trained research assistants. Key contract provisions were recorded in structured templates, including wage steps by year of service and rank, leave policies by tenure, education incentives, and longevity pay. Extracted data were entered into a shared database and reviewed for consistency. When discrepancies emerged, original contracts were reexamined. Each contract record was linked to metadata such as city, union name, effective dates, and whether the agreement was reached voluntarily or through arbitration.

An additional analytic step involved measuring negotiation delays. I calculated the number of months between the expiration date of a previous contract and the execution date of a new agreement. This measure was used to construct a continuous indicator of bargaining delay, which I document as part of the dataset's institutional detail; it is not used to estimate effects on enforcement activity in this paper.

The data underwent a multi-stage validation process. Two independent coders reviewed overlapping subsets of contracts. A random sample of records was audited by a third reviewer. Discrepancies in variables such as wage increases, education stipends, and longevity bonuses were resolved through direct consultation of the original contracts. Formatting inconsistencies were corrected by standardizing monetary values and converting percentage-based incentives to equivalent annual dollar amounts when possible.

The final dataset includes detailed officer-level contract provisions such as starting salary, step raises, promotional wage scales (e.g., sergeant, lieutenant, captain), vacation and sick leave schedules, uniform allowances, education incentives (for associate, bachelor, master's, and doctorate degrees), and longevity payments (monetary bonuses or salary percentage increases awarded at milestones such as 4, 8, 12, 16, 20, and 24 years of service). The dataset also includes variables indicating contract duration, method of settlement, and timing of agreement.

## 2.3 Contract Outcome Panel

For the analysis of the compensation package in Section 5, I collapse the contract-level rows to one observation per agency-year over 1996–2024, retaining the most recently signed contract that governs each agency-year. The resulting panel is unbalanced, covering 480 agencies and 10,103 agency-years; not every agency is observed in every year. It is larger than the clearance sample of Section 3 because it does not require a match to the clearance data: all 463 agencies in the clearance analysis are in this panel, and the remaining 17 have contracts but no matched clearance records. Of these agencies, 432 are treated and 48 are within-panel never-treated. All dollar outcomes are deflated to 2010 prices using the BLS CPI series and winsorized at the 1/99 percentile per outcome to limit the influence of coding errors at the tails. Table 11 reports non-missingness rates by outcome, and Table 12 reports pre-cap and post-cap means by treatment status.

A consequential feature of this panel is how missing data should be interpreted. Random

spot checks of contract PDFs reveal that nearly every New Jersey municipal police collective bargaining agreement contains an explicit "Continuation of Benefits" or "Past Practice" clause stating that benefits previously enjoyed but not restated in the current contract shall continue. This reflects a binding legal constraint under N.J.S.A. 34:13A-5.3 and Public Employment Relations Commission practice: employers cannot unilaterally remove benefits established through prior bargaining without negotiating their removal. A missing value for an agency that previously coded a provision should therefore be read as the provision remaining in effect, not as the provision having been removed. I use two treatments of missing data and rule out a third. The main results report the more conservative missing-at-random (MAR) estimates, which discard inherited values and leave wider confidence intervals, so that the headline numbers do not rely on imputation. As a robustness check, last-observation-carried-forward (LOCF) imputation within agency operationalizes the continuation clause directly by carrying a previously coded provision forward. I rule out treating missing data as "provision removed," because that would conflict with the institutional default and would generate misleading inferences about the cap triggering provision removals when the underlying mechanism is inheritance, not removal.

## 2.4 Civil-Service Governance and Personnel Data

The New Jersey Law Enforcement Diversity Data (New Jersey Office of the Attorney General, 2024), released by the Office of the Attorney General under N.J.S.A. 52:17B-4.10, provides agency-level structural variables including an indicator for whether hiring is governed by civil service regulations. Civil-service agencies face statutory promotion rules tying promotional wage steps to the civil-service rank structure; non-civil-service agencies retain discretion to negotiate the full wage scale. I crosswalk the diversity-data agency list to the contract panel by fuzzy name matching, matching 465 of 480 panel agencies (97 percent), of which 279 are non-civil-service and 186 are civil-service, and I treat the flag as time-invariant. The 2021–2025 stacked release of the same data also records officer-level promotion outcomes

and workforce composition, which I use descriptively in Appendix 9.2.

### 2.5 Outcome Measures

I examine three outcome measures capturing different dimensions of police effectiveness. The total clearance rate measures the percentage of all reported crimes, both violent and property offenses, that result in an arrest or exceptional clearance, measured in percentage points. The property clearance rate measures the percentage of reported property crimes, including burglary, larceny-theft, and motor vehicle theft, that are cleared, also measured in percentage points. The theft clearance rate focuses specifically on the percentage of larceny-theft offenses that are cleared, again measured in percentage points.

Clearance rates represent a standard measure of police investigative effectiveness and enforcement intensity. Unlike arrest rates, which may respond mechanically to crime incidence, clearance rates capture the probability that a reported crime is successfully resolved through investigative effort. These outcomes are plausibly responsive to morale or resource effects stemming from compensation negotiations: if the arbitration cap dampens officer expectations or strains municipal budgets, we might observe declines in discretionary enforcement activities reflected in lower clearance rates. Conversely, if the cap stabilizes bargaining dynamics and reduces contentious arbitration proceedings, clearance rates may remain stable or even improve.

## 3 Agency-Level Analysis: Staggered Difference-in-Differences

The staggered difference-in-differences design treats each agency's contract expiration as an independent treatment event, using the full variation in treatment timing across agencies rather than comparing a subset of early adopters to a subset of late adopters. This is the paper's primary design for the cap.

The staggered DiD framework uses all not-yet-treated agencies as controls at each point

in time, substantially increasing the effective sample size from 294 agencies to 463 agencies and incorporating all available variation in treatment timing. This approach is particularly well-suited to the institutional setting, where the cap's binding constraint varies precisely with contract expiration dates. Goodman-Bacon (2021) shows that the conventional two-way fixed-effects estimator recovers a variance-weighted average of treatment effects that can be biased when treatment timing is staggered and effects are heterogeneous, because already-treated units serve as controls. To avoid this bias, I employ two modern estimators designed for staggered treatment adoption: the Callaway and Sant'Anna (2021) estimator and the Sun and Abraham (2021) estimator. These methods provide independent validation of the key findings while offering greater statistical power to detect treatment effects.

## 3.1 Empirical Strategy

The state-level analysis treats all New Jersey agencies as uniformly exposed to the cap beginning in 2011. However, the cap's constraints bind only when agencies renegotiate contracts. Agencies whose contracts did not expire during the cap period (2010–2017) were never directly exposed, while those with expiring contracts faced treatment at different times depending on their contract expiration dates.

This variation in treatment timing motivates a staggered difference-in-differences design. Treatment is defined by contract expiration during the cap period. Let $G_i$ denote the year agency $i$'s contract expired during the cap window, with $G_i = \infty$ for agencies whose contracts never expired between 2010 and 2017. Agency $i$ is treated in year $t$ if $t \geq G_i$ and $G_i \leq 2017$.

I estimate treatment effects using the full sample spanning 1996–2020 and including 463 agencies. This long panel provides the maximum sample size and allows for robust pre-treatment diagnostics with extensive pre-cap data from 1996–2010. The longer time series strengthens identification by providing more pre-treatment observations to validate the parallel trends assumption and more post-treatment observations to estimate dynamic treatment effects.

The key identification assumption is that agencies whose contracts expired at different times would have followed parallel trends in clearance rates absent the policy. I assess this assumption using pre-treatment event-study coefficients and formal statistical tests for differential trends between early adopters and late adopters.

## 3.2 First Stage: Did the Cap Bind?

This design assigns treatment by contract expiration, so before estimating effects on clearance it is worth asking whether a contract expiring under the cap in fact constrained compensation. Evidence from individual arbitration awards, examined in Section 4, speaks only to the small set of contracts that reached arbitration. Because the cap applies to every municipal police contract in the state, its first-stage effect can be measured on all agencies in the sample rather than on arbitrating agencies alone. I construct an annual measure of negotiated compensation growth from the collected bargaining agreements, defined as the year over year percent change in the entry step of the base salary schedule. The resulting panel covers 460 agencies and 5,077 agency-year observations between 2005 and 2020. These are the 463 agencies in the clearance sample, less three that have no consecutive-year salary schedules in that window. I apply the estimator, the not-yet-treated comparison group, and the cohort definitions set out above, so that the first stage and the clearance estimates rest on identical identifying variation.

Table 3 reports the results. The cap reduced negotiated salary schedule increases by 1.99 percentage points (standard error 0.20), against a pre-cap mean of 3.6 percent, and raised the probability that an agency's annual increase fell below 2 percent by 71 percentage points (standard error 0.05). Both estimates are precisely determined. The event study shows no pre-trend, with an average pre-period coefficient of 0.22 (standard error 0.32), a sharp break in the year a contract first falls under the cap, and an effect that is stable at roughly two percentage points for the remainder of the sample. Raw means move in the same way, with the share of agency-years below 2 percent rising from 0.16 before the cap to 0.67 after it.

Two qualifications are worth stating. The measure tracks the negotiated schedule rather than realized payroll, so it does not capture step movement, longevity, or changes in rank composition, and it therefore speaks to the contractual terms officers bargained over rather than to take-home pay. It also reflects the statutory ceiling operating alongside the recession and the 2010 property tax levy cap, which the not-yet-treated comparison absorbs only to the extent that those forces affected early and late cohorts similarly.

Subject to those caveats, the first stage is unambiguous: the constraint bound, and it bound on the margin officers bargain over. It also supports the treatment definition itself, since compensation growth breaks in the year a contract expires rather than in 2011, which is what assigning treatment by expiration date requires. The clearance estimates that follow can therefore be read as the response to a policy that demonstrably took hold.

## 3.3 Results

### 3.3.1 Pre-Treatment Diagnostics

The staggered DiD design exploits variation in contract expiration timing across agencies, requiring the parallel trends assumption: agencies with different contract timing would have followed similar trajectories absent the cap. Diagnostic tests using the long-panel sample validate this assumption. All four complementary tests yield p-values exceeding 0.10 across property, theft, and violent clearance rates, comparing trends between early adopters (contracts expiring 2011–2012), late adopters (2013+), and never-treated agencies during 2000–2010. Agencies with different contract timing followed parallel trajectories before the cap's implementation.

Figures 9, 10, and 11 focus specifically on pre-treatment periods, showing early adopters before 2011 and late adopters before 2013. These figures demonstrate that both groups exhibited similar patterns during their respective pre-treatment windows, with no systematic divergence that would violate the parallel trends assumption. The series track closely across all three crime types, showing comparable levels and trends during the periods before each

group experienced contract expiration.

Figures 6, 7, and 8 display the complete time series from 1996–2020 with treatment timing marked by vertical dashed lines. Both early and late adopters track closely before 2011, when the cap was announced and implemented statewide, supporting the parallel trends assumption. After 2011, the series show similar patterns with no dramatic divergence, consistent with the small estimated effects found in the formal analysis. The long-panel diagnostics validate parallel trends during the pre-cap period, supporting identification for the staggered design. Any apparent divergences after 2011 likely reflect anticipatory adjustments or actual treatment effects rather than pre-existing differential trends.

Figures 6, 7, and 8 display complete time series from 1996–2020 with treatment timing marked by vertical dashed lines. Both early and late adopters track closely before 2011, supporting the parallel trends assumption. After 2011, the series show similar patterns with no dramatic divergence, consistent with the small estimated effects found in the formal analysis.

### 3.3.2 Treatment Effect Estimates

Table 4 reports staggered DiD estimates using 10,511 agency-year observations across 463 agencies over 1996–2020. Panel A holds the comparison group fixed at never-treated agencies so that the two estimators are directly comparable. Both yield negative effects on every outcome. The Callaway-Sant'Anna estimator gives declines of 2.5 percentage points for property clearances, 3.3 percentage points for theft, and 1.9 percentage points for total clearances; the Sun-Abraham interaction-weighted estimator gives 1.8, 1.6, and 1.5 percentage points respectively. The two estimators therefore agree in sign and in order of magnitude once the comparison group is held fixed. None of the estimates is statistically distinguishable from zero. Panel B reports the preferred specification, which uses all not-yet-treated agencies as controls: property clearances decline 2.3 percentage points, theft 2.6 percentage points, and total clearances 1.9 percentage points, with standard errors roughly fifteen percent smaller

than under the never-treated comparison. Measured against pre-cap baseline clearance rates of 19.1, 19.4, and 22.7 percent respectively, these correspond to relative declines of about 12, 13, and 9 percent. Panel C repeats the preferred specification weighting agencies by offense volume, which halves the standard errors and leaves the point estimates of similar magnitude at 1.7, 2.3, and 1.3 percentage points. The point estimates are stable across the two comparison groups, indicating that the result does not depend on the small set of agencies whose contracts never expired during the cap window.

Both estimators show considerable uncertainty, with confidence intervals that include zero and substantial effects in both directions but consistency in finding small-to-moderate negative point. The Callaway-Sant'Anna estimates suggest larger negative effects but come with substantial uncertainty, while Sun-Abraham produces estimates much closer to zero. Importantly, both methods rule out large positive effects and provide no evidence of catastrophic performance declines despite significant wage compression.

### 3.3.3 Interpretation

Both the Callaway-Sant'Anna and Sun-Abraham estimators yield negative point estimates, none of which reaches statistical significance. This pattern is consistent across all three crime types examined. What emerges most clearly from this analysis is that the arbitration cap did not generate substantial performance declines. Even the Callaway-Sant'Anna estimates, which suggest the largest negative effects, have confidence intervals that rule out declines worse than approximately 8-10 percentage points for most crime types. The Sun-Abraham intervals are of similar width, excluding declines larger than roughly 6 to 7 percentage points.

The two estimators agree closely once the comparison group is held fixed. In Panel A both use never-treated agencies as controls, and the gap between them is well inside the sampling error of either estimate: 0.6 percentage points for property clearances against standard errors of roughly 2.5 to 2.9. What remains reflects how each procedure aggregates

cohort-specific effects rather than any disagreement about which agencies serve as controls. Callaway-Sant'Anna averages group-time treatment effects directly, while the Sun-Abraham interaction-weighted estimator recovers event-time coefficients that are then averaged using the same event-time weights. Both are robust to the negative-weighting problems that affect two-way fixed-effects event studies under heterogeneous and staggered adoption (Borusyak et al., 2021; Roth et al., 2023). Their agreement is therefore informative in its own right: the muted response is not an artifact of a particular aggregation scheme, and it does not depend on which of the two heterogeneity-robust procedures is used.

As an additional check on the pre-trends assumption, I apply the Rambachan and Roth (2023) sensitivity analysis to the Callaway-Sant'Anna estimates. For each outcome, this procedure asks: how large would the post-treatment deviation from parallel trends have to be, relative to the largest pre-treatment violation, before the 95% confidence interval no longer excludes zero? Across the three crime outcomes the breakdown $\bar{M}$ values are in line with what would be considered modest relative violations, consistent with the conclusion that the estimated effects are small but not robust to substantial unobserved differential trends.

This pattern aligns with expectation adjustment theory. When officers anticipate constrained wage outcomes through a transparent, binding cap, they may adjust expectations rather than experience the disappointment-driven effort withdrawals that emerge when outcomes unexpectedly fall short. The cap compressed wage growth from approximately 4 percent annually to 1 percent, yet clearance rates showed no corresponding collapse. Multiple factors counsel against strong conclusions about the precise magnitude of effects. The Callaway-Sant'Anna estimates have large standard errors and show considerable variability across crime types, suggesting residual noise in the estimation. The Sun-Abraham estimates, while more precise, hover very close to zero with modest negative point estimates. Theoretically, wage compression might be expected to reduce discretionary effort absent other compensating mechanisms.

Two readings of why the response is muted deserve emphasis, because they bear on how the null should be interpreted. The first concerns the channel. Mas (2006) identifies the effect of losing an arbitration, a relative defeat against a specific counteroffer, whereas the cap lowers the magnitude of raises uniformly without staging such a defeat. If the reciprocal effort response is triggered by the experience of losing rather than by the level of the raise, a muted response to the cap is what the reference-dependence account predicts. The second concerns aggregation. Even holding per-officer effort fixed, if municipal budgets continued to grow while base-salary growth was capped, the freed funds could expand staffing or paid hours, so agency-level clearance could hold steady through greater capacity. Neither reading licenses treating the insignificant estimates as evidence of no effect. The defensible statement is narrower still. The estimates are small and negative, but they are too imprecise to separate a true zero from a decline of the magnitude Mas-style disappointments would predict, which remains inside the confidence interval of every specification, including the weighted one. What the intervals exclude is deterioration far larger than that.

Two heterogeneity-robust estimators applied to the same comparison group produce negative point estimates of similar magnitude, and the estimates are stable across the never-treated and not-yet-treated comparison groups. The estimates are not precise enough to distinguish moderate declines from zero, so the finding is best read as an absence of evidence for large effects rather than as evidence of their absence.

While the precise magnitude remains uncertain, estimates range from small negative to near-zero. All approaches exclude the substantial performance declines that might be expected from significant wage compression. From a policy perspective, the key finding is clear. Fiscal savings from the arbitration cap did not come at the cost of catastrophic performance declines. This addresses the primary concern that binding compensation constraints might trigger widespread effort reductions among police officers. The consistency of this finding across multiple identification strategies and crime types suggests it is robust to different methodological choices and assumptions.

### 3.3.4 Does the Response Scale with the Visible Loss?

A uniform cap lowers every agency's raise, but it does not lower every raise by the same amount relative to what officers can observe nearby. Because contracts expire at different times, an agency renegotiating under the cap may sit among neighbors still operating under pre-cap agreements. New Jersey's interest-arbitration statute directs arbitrators to weigh comparability with nearby jurisdictions, so a neighboring settlement is a benchmark both parties already use.

I measure this visible gap as the difference between the average raise received by same-county agencies whose contracts had not yet expired and the agency's own raise at renegotiation. The measure is defined for 317 of the 463 agencies. The remainder are excluded mechanically rather than by selection, since agencies renegotiating late in the cap window have no uncapped neighbors left and agencies whose contracts never expired do not renegotiate at all. The median agency saw uncapped neighbors receiving 1.4 percentage points more than itself, and the ninetieth percentile saw 5.9 points more. Almost all of this variation occurs within rather than across expiration cohorts, so the measure is not a restatement of when an agency happened to renegotiate.

Splitting at the median gap produces two groups with similar cohort composition. Agencies in the high-gap half faced neighbors receiving 4.0 percentage points more on average, while those in the low-gap half faced neighbors receiving 0.9 points less. Estimated separately, and weighting by offense volume, the effect on property clearance rates is $-0.07$ percentage points (SE 1.31) in the high-gap half and $+0.50$ percentage points (SE 2.20) in the low-gap half. The difference between the two is small relative to either standard error.

These estimates do not show the performance response scaling with the size of the visible loss. The exercise tests one channel and does not establish that officers were indifferent to the cap. The confidence intervals admit effects of roughly two and a half percentage points in either direction, so a modest gradient would not be detected. What the results indicate is that the most salient available measure of relative loss does not organize the variation in

measured performance.

# 4 Pay Disappointment and Discretionary Effort

## 4.1 An Individually Observable Pay Shock

The cap tightened compensation sharply, yet clearance rates moved little. One reading of that pattern is that officers simply do not adjust effort in response to pay. A second is that they do, but that a pre-announced, statewide ceiling is the wrong kind of pay shock to elicit it. Distinguishing the two requires a setting in which an individual agency's officers receive an adverse compensation outcome that is specific to them and not known in advance.

Interest arbitration provides one. This section uses the outcome of individual arbitration cases to ask whether officers withdraw discretionary effort when a pay decision goes against them, and reads the answer against the cap estimates of Section 3.

The variation comes from individual interest arbitration awards. The setting differs from the one Mas (2006) studies in an important way. During his 1978–1996 sample, New Jersey used final offer arbitration, under which the arbitrator selects one party's final offer in its entirety, so that conditional on the offers, whether the employer or union prevails is plausibly as good as random. Since the 1996 reform of the arbitration statute, these disputes have instead been resolved by conventional arbitration, in which the arbitrator may set the award anywhere on a continuum. In my data, 139 of the 140 dockets are conventional awards, and 109 of the 140 awards fall strictly between the two parties' final offers. I therefore define an employer win as an award that lands closer to the employer's final offer than to the union's.

This changes what the comparison can claim. A conventional arbitrator chooses a point on a continuum by weighing statutory criteria, so which offer the award lands nearer partly reflects the merits of the case rather than a coin flip between two packages. The binary is also fragile, because about half of the awards fall within 20 percent of the midpoint between the two offers, and the offer figures flatten multi-year, multi-dimensional packages into a single

percentage that is not always on a common cost basis for the two parties. I therefore read the estimates below as associations between adverse awards and subsequent performance, conditional on the offers, rather than as quasi-experimental effects.

The theoretical logic underlying this approach centers on officer expectations and disappointment. When arbitrators issue decisions that fall short of officers' expectations, officers may reduce discretionary enforcement activities such as arrests and investigative effort. Before the cap, arbitration awards varied widely, ranging from zero to over six percent annually (Figure 1). Officers might reasonably expect awards of 4 to 5 percent based on recent patterns, meaning a 2 percent award would represent significant disappointment.

The two percent cap fundamentally altered this dynamic by constraining the maximum possible award, making outcomes highly predictable. The cap's effect on officer behavior depends critically on expectation adjustment. If unions and officers lower their expectations in response to the known constraint, the disappointment effect should be muted or eliminated. Officers who expect 2 percent and receive 2 percent experience no disappointment. However, if officers view the cap itself as an unfair constraint on theircompensation growth, the policy could generate persistent effort reductions.

Figure 1 provides evidence consistent with expectation adjustment. Average union final offers dropped from 4.8 percent before the cap to 2.4 percent afterward, and the gap between union proposals and final awards narrowed considerably. This convergence suggests that unions internalized the constraint and adjusted their demands accordingly. The cap's primary effect may have been to eliminate the variance in arbitration outcomes that previously generated behavioral responses, rather than creating uniform disappointment. The award distribution shifted sharply under the cap. Before 2011, 71 percent of awards exceeded 3 percent and only 10 percent fell at or below 2 percent. Afterward no award exceeded 3 percent and 93 percent fell at or below the ceiling, with the mean falling from 3.4 to 1.3 percent. Awards did not, however, pile up immediately beneath the cap. The post-cap median was 1.4 percent and two thirds of awards fell below 1.8 percent, so the ceiling bound the

upper tail without becoming the modal outcome. This matters for interpretation, because it indicates that for most cases something other than the ceiling determined the award.

This logic implies that the cap should produce a smaller behavioral response than an individual arbitration defeat of comparable monetary size, which is the pattern the previous section reports. The test below asks whether the second half of that comparison holds in this setting: whether an individual defeat moves effort at all.

I estimate how individual arbitration outcomes affect police performance using the following specification. Let $\text{EmployerWin}_i$ be an indicator equal to one if the award in arbitration case $i$ lies closer to the employer's final offer than to the union's. I merge this measure with monthly agency-level panel data on police clearance rates spanning a 24-month window ($-12$ to $+12$ months) around arbitration decisions occurring between 2010 and 2017. The estimating equation is:

$$y_{it} = \alpha + \beta \cdot \text{Post}_{it} \times \text{EmployerWin}_i + \lambda_i + \tau_t + X_{it}'\Gamma + \varepsilon_{it}, \tag{1}$$

where $y_{it}$ represents the clearance rate for agency $i$ in month $t$, $\text{Post}_{it}$ indicates months after the arbitration decision, $\lambda_i$ and $\tau_t$ are agency and time fixed effects, and $X_{it}$ includes time-varying controls. The coefficient $\beta$ captures the effect of employer victories on subsequent clearance rates. Standard errors are clustered at the agency level.

## 4.2 Results

Table 1 presents two ways of measuring how arbitration outcomes affect police performance. Panel A uses clearances per 100,000 capita, the same measure Mas (2006) used. Panel B uses the clearance rate: the percentage of reported crimes that police successfully solve.

These two measures capture different dimensions of police performance. Clearances per capita reflect both the volume of crimes available to be cleared and the intensity of inves-

tigative effort officers devote to solving them. This measure can change either because police work harder or because crime itself changes. For example, if crime drops in a jurisdiction, clearances per capita will mechanically fall even if police effort stays constant, simply because there are fewer crimes to clear. The clearance rate, by contrast, isolates investigative efficiency by measuring what share of crimes get solved, holding the volume of crimes constant. By construction, the clearance rate is unaffected by exogenous fluctuations in crime incidence and instead directly measures the share of reported crimes that officers successfully clear through arrest or exceptional means. This makes it a more direct test of whether officers reduce discretionary investigative effort in response to wage disappointments.

Panel A shows that employer victories in arbitration are associated with 3.1 fewer clearances per 100,000 people (column 3, my preferred specification with arbitration window fixed effects). The estimate is not statistically significant at conventional levels, with a standard error of 2.1 ($p = 0.13$). The imprecision likely reflects several factors. First, my 24-month observation window is substantially shorter than Mas's multi-year panels, reducing statistical power. I restrict the window to maintain a balanced panel around arbitration decisions and avoid contamination from subsequent arbitrations or contract renegotiations, which typically occur on 2 to 4 year cycles. Second, monthly clearance counts exhibit high variance, particularly for smaller agencies where zero-crime months can create missing observations. Third, population denominators introduce additional measurement error, as intercensal estimates are less precise at the municipal level.

Panel B presents results using the clearance rate as the outcome. Column (6), my preferred specification, shows that employer victories are associated with clearance rates 2.2 percentage points lower (standard error = 1.2, $p = 0.06$). This effect is economically meaningful, representing a 7.7 percent decline relative to the baseline clearance rate of 28.5 percent. The estimate is marginally significant and more precisely estimated than the per capita specification in Panel A. Because the clearance rate holds crime volume constant, this result directly measures whether officers reduce investigative effort following disappointing pay

outcomes. Officers retain substantial discretion over how intensively they pursue cases, how thoroughly they follow up on leads, how many witnesses they interview, and how much time they allocate to case investigation. The estimate is consistent with some withdrawal of effort on these margins, though at conventional thresholds it is marginal and I read it as suggestive.

For scale, Mas (2006) reports declines on the order of 5 to 8 percent in clearances following employer victories, so the magnitude here is of a similar order. I treat that as a point of comparison rather than as corroboration: the designs differ in sample, period, and outcome construction, and the estimate here is only marginally significant.

The two measures point in the same direction, toward a modest reduction in discretionary effort following an adverse arbitration outcome, but only the clearance-rate estimate approaches conventional significance and neither establishes the effect decisively. The clearance rate offers a complementary perspective by isolating changes in investigative intensity from variation in crime volume. When crime fluctuates for reasons unrelated to officer effort such as seasonal patterns, economic conditions, or demographic shifts, the per capita measure incorporates this noise, while the clearance rate conditions it out. The fact that Panel B yields more precise estimates suggests this conditioning helps isolate the behavioral response of interest. This does not diminish the value of the per capita specification, which remains essential for understanding absolute magnitudes and comparing across studies. Rather, the two measures provide complementary insights: per capita captures the total impact on clearance activity, while the clearance rate isolates the efficiency channel.

One limitation of this specification is that it treats arbitration outcomes as binary, an employer win or a union win, and so collapses the intensive margin of how far short of expectations an award fell. Reference dependence predicts that the size of the shortfall, not only its sign, should govern the effort response. I estimated that specification directly, interacting the employer-win indicator with the magnitude of the shortfall between the union's position and the award. The design cannot resolve a gradient at the precision this sample allows, so I report the binary estimates here as an average over that heterogeneity.

Returning to the question this section opened with, the evidence does not cleanly separate the two readings. An individual arbitration defeat is associated with a 7.7 percent relative decline in clearance rates, and the cap with relative declines of roughly 9 to 13 percent, so the individual shock does not produce the visibly larger response that the expectation-adjustment account would lead one to expect. Both estimates are imprecise, the two designs differ in sample, window, and baseline, and the comparison is therefore suggestive at best. What can be said is that neither setting produces evidence of a large effort withdrawal, and that the muted response to the cap documented in Section 3 is not explained by officers being unresponsive to pay in general, since the individual-defeat estimate is if anything the weaker of the two.

# 5 Where Did the Compressed Pay Go?

The cap compressed negotiated salary growth sharply, yet measured performance moved little. One natural reconciliation is that officers were not in fact made worse off. If the value squeezed out of base salary reappeared in the contract's many uncapped provisions, effective compensation changed little and the weak performance response follows directly. Whether that reallocation occurred is a question about contract structure, and it is the question this section takes up.

The muted performance response admits several explanations, and it is worth setting them out before narrowing to the one the contract panel can test. First, officers may not have experienced the cap as a loss, because a uniform, pre-announced ceiling resets expectations in a way that an idiosyncratic arbitration defeat does not (Koszegi and Rabin, 2006). Second, the compressed base-salary value may have reappeared inside the contract, loaded onto the many uncapped non-salary margins, leaving effective compensation little changed. Third, it may have escaped into budgetary channels the contract does not record, such as overtime, court time, and shift differentials, or funded expanded staffing and hours so that agency

output held up through capacity rather than effort. Fourth, working conditions may have improved as wages fell, a compensating-differentials adjustment (Rosen, 1986) that offsets the cut in officers' overall package. Fifth, the incidence of the cut may be narrow, concentrated on new entrants at the capped entry-salary line while incumbents are largely insulated, so that any aggregate effort effect is small by construction. This section takes up the second explanation directly, because it is the one the contract panel can adjudicate.

## 5.1 Scope of the Cap and Empirical Approach

The statute's scope defines the central distinction. The cap binds the aggregate increase in base-salary items over the term of a contract, where base-salary items are defined to include salary-guide wages, step increments, and longevity payments (N.J.S.A. 34:13A-16.7). Non-salary economic issues, including uniform allowance, tuition stipends, and holiday and leave accrual, lie outside the cap. The operative margin is therefore capped versus uncapped rather than wage versus non-wage. In particular, the salary-guide wages of supervisory ranks are themselves base-salary items, so I treat them as capped components.

I organize the outcomes by the kind of contract response they capture. The directly capped outcome is the entry wage, the starting salary at year of service one, which is the line the cap was written to bind. Categorical outcomes ask whether bargained value flows into the contract's other compensation categories, and they cover both forms such a flow could take: raising the dollar value of an existing category, measured in log form for education stipends at the associate, bachelor's, master's, and doctoral tiers, the tuition reimbursement value, uniform allowance, holiday count, sick leave allowance, and vacation accrual at the five-year tenure milestones; and opening a category an agency did not previously offer, coded as indicators for whether the agency has a tuition incentive, a tuition-increment provision, or a tuition reimbursement provision, together with first-introduction indicators. Longevity payouts are also coded, but as base-salary items inside the cap they are not part of the uncapped family. The promotional wages at the Sergeant, Lieutenant, and Captain ranks

are themselves capped base-salary items, and I carry them in log form to ask whether value moved up the rank ladder within the capped aggregate.

The design is the one used for clearance in Section 3: the Callaway and Sant'Anna (2021) estimator with not-yet-treated agencies as the comparison group, treatment cohort $G_i$ defined by the year of first contract expiration in the 2010–2017 cap window, and standard errors clustered at the agency level. Because the outcome family has twenty-seven members, I report Holm step-down adjusted $p$-values (Holm, 1979), which control the family-wise error rate under arbitrary dependence across outcomes. Pooling all twenty-seven outcomes into one family is deliberately conservative; Appendix 9.1 explains why neither conclusion below depends on how the family is drawn. Because failing to reject parallel pre-trends does not establish that they hold (Roth et al., 2023), I also report the Rambachan and Roth (2023) relative-magnitude sensitivity analysis for the headline outcomes.

## 5.2 The Cap Compresses Entry Wages

The cap binds where it was designed to bind. Table 5 reports a Callaway–Sant'Anna ATT on log starting salary of $-0.0465$ (SE 0.0079, Holm $p < 0.001$): agencies whose contracts expired during the cap window pay entry wages about 4.5 percent lower in real terms than not-yet-treated comparison agencies, the largest single-outcome effect in the twenty-seven-outcome family after the Holm correction. This is the level counterpart of the first stage in Section 3.2, not a re-estimate of the same quantity: that section measures the cap's effect on the annual growth of the negotiated salary schedule, which fell by about two percentage points, whereas the estimate here measures its effect on the level of the entry wage written into contracts.[2] The statutory limit on salary awards passes through to the entry wage that agencies write into their contracts.

[2]The two magnitudes are related but not directly comparable, and they agree in the way one would expect. A raise rate cut by roughly two percentage points per year, compounded over a typical two- to three-year New Jersey contract cycle, implies a cumulative entry-wage level about 4 to 6 percent below the not-yet-treated comparison path, bracketing the 4.5 percent estimated here.

## 5.3 No Spillover into Other Compensation Categories

Where does the compressed value go? Not, it appears, into the contract's other priced compensation categories. Table 5 reports Callaway–Sant'Anna ATTs on the uncapped non-salary categories: education stipends, tuition reimbursement value, uniform allowance, holiday count, sick leave, and vacation accrual at the 5-, 10-, 15-, and 20-year milestones. Every estimate lies within its standard error of zero, with Holm-adjusted $p$-values above 0.99 and confidence intervals tight enough to exclude effects larger than a few percent. Only two categories are even marginally distinguishable from zero before correction, and both move the wrong way for a substitution story: sick leave falls 2.2 percent and the master's-tier tuition stipend falls 2.0 percent. If anything, cap-bound agencies trimmed these categories slightly rather than loading value into them. Longevity payouts are excluded from this comparison. The statute places them inside the cap, and they are in any case estimated too imprecisely to sign.

The same holds for adding a compensation category an agency did not previously offer. A binding constraint on the wage line could plausibly push agencies not only to raise existing perks but to introduce new ones, and tuition money is the natural place to look on both margins. The indicators for whether an agency has a tuition incentive, has a tuition-increment provision, newly introduces a tuition incentive, or newly introduces a tuition reimbursement provision show no statistically distinguishable post-treatment movement. Agencies neither raised tuition money nor began offering it after the cap.

Before reading these nulls as real, consider whether they are an artifact of missing data. Because of the continuation clauses described in Section 2.3, increasing reliance on inherited language could mechanically depress later-period adoption under missing-at-random handling and manufacture a null. Two checks address this. First, the cap does not itself generate missingness: a difference-in-differences on the agency-year missingness rate, with agency and year fixed effects and standard errors clustered at the agency level, returns an ATT of +0.36 percentage points (SE 0.77) against a baseline missingness rate of 36 percent.

Second, Table 8 re-estimates the non-wage outcomes under last-observation-carried-forward imputation, which operationalizes the continuation clause directly. Coverage rises substantially, for example longevity at 16 years from 73.8 to 89.5 percent of treated agency-years and uniform allowance from 77.1 to 90.4 percent, yet the point estimates and significance pattern are essentially unchanged across all seventeen non-wage categories. Of the two outcomes marginally distinguishable from zero under missing-at-random, the master's-tier tuition stipend is no longer even marginal under imputation and sick leave stays marginally negative, and no outcome gains significance. The categorical null is a feature of the data, not of how missing values are treated.

A simple accounting exercise makes the magnitude concrete. Table 6 translates the log-point ATTs into approximate dollars per officer-year. The cap removes roughly $3,100 per officer-year from the wage scale, of which only about $290, under one-tenth, reappears in any observable category. Taking the upper end of the 95 percent interval on every observable channel at once, the most that could resurface is about 19 percent of the compression. At least four fifths of the value the cap removed is therefore unaccounted for on the margins the contract records, under a reading generous to the substitution hypothesis at every step. Because the dollar weights rest on assumed workforce and tenure shares rather than agency head counts, I read the exercise as illustrative rather than as a precise valuation.

The accounting also raises a question of incidence. Because the cap binds the entry-salary line most directly, the compression falls disproportionately on new and early-career officers moving up the front of the step schedule, while officers already near the top of the guide are more insulated. If the burden is concentrated on a minority of the workforce in any given year, the aggregate effort response can be small even where the per-affected-officer cut is not, which offers a further reconciliation with the muted clearance response in Section 3. A direct decomposition of the compression across the tenure and rank distribution is left to work with officer-level payroll data the contract panel does not contain.

## 5.4 Promotional Wage Levels and Civil-Service Governance

The promotional wage levels point toward substitution up the rank ladder, but the evidence is suggestive rather than identified. The log Sergeant wage rises by 1.3 percent (Holm $p = 0.018$) and the log Captain wage by 1.9 percent (Holm $p = 0.020$), while the Lieutenant wage shows no distinguishable movement. Two features keep me from reading these as identified substitution. Coverage is partial and selective: the Sergeant wage is recorded for 70 percent of agency-years, the Lieutenant wage for 44 percent, and the Captain wage for only 27 percent, with the missing observations concentrated in smaller agencies that do not maintain those ranks. And these wage-level estimates are sensitive to assumed violations of parallel trends, as Section 5.5 shows. Because supervisory wages are themselves base-salary items inside the cap, even taken at face value these movements would show only a reallocation within the capped aggregate, not value escaping the cap, and the contract data do not reveal whether the aggregate is computed across rank lines.

The response also differs by civil-service governance (Table 7). The entry-wage compression is concentrated in civil-service agencies, where the starting-salary ATT is $-5.7$ percent, while in non-civil-service agencies it is $+0.2$ percent and indistinguishable from zero. The promotional wage levels run the other way: the Sergeant wage rises 1.8 percent and the Captain wage 2.7 percent in non-civil-service agencies and both are flat in civil-service ones. The contrast is consistent with non-civil-service agencies, free to renegotiate the full scale, allocating wage growth toward supervisory tiers, while civil-service agencies, bound to a statutory structure, absorb the cap on the entry wage. Because the wage-level contrasts rest on the fragile Sergeant and Captain estimates, I read the split as consistent with an institutional-substitution mechanism rather than as establishing it. Appendix 9.2 reports the post-cap promotion and workforce-composition evidence, which does not show a corresponding personnel response.

## 5.5 Sensitivity to Parallel-Trends Violations

For each of the three headline outcomes I take the event-study estimates from the staggered design and ask how large a post-period violation, expressed as a multiple $\bar{M}$ of the largest pre-period violation, the inference tolerates before the robust 95 percent confidence interval includes zero.[3]

Under exact parallel trends ($\bar{M} = 0$) all three effects are significant. The outcomes differ sharply in how quickly that significance dissolves as $\bar{M}$ grows. The starting-salary compression is the more robust, with its interval excluding zero up to about $\bar{M} = 0.15$. The promotional wage levels are fragile: the Sergeant and Captain wage intervals include zero once $\bar{M}$ exceeds roughly 0.05. This ordering is why the entry-wage compression anchors the evaluation while the promotional wage movements are reported as suggestive. The central result of this section, the precisely estimated non-salary null, is a statement about the absence of an effect and does not rest on this sensitivity analysis.

## 5.6 Implications for the Performance Results

Read against the menu of explanations for the muted performance response, the contribution of this section is to close off one branch. Officers were not plausibly made whole through the contract's uncapped margins, so the reconciliations that remain live are the ones the contract panel cannot see: an expectation reset that dulled the sense of loss, value flowing into budgetary channels or expanded capacity, working-conditions improvements that would act as compensating differentials, and an incidence concentrated on early-career officers at the entry-salary line. The contract panel does not record overtime, court time, or shift differentials. These are bargaining margins paid through municipal budgets rather than written into the salary guide, so the analysis cannot rule out substitution toward budgetary channels that the contract data do not observe.

[3]The sensitivity analysis targets the equally weighted average of the post-period event-study coefficients, which differs slightly from the cohort-weighted simple ATT reported in Table 5; for starting salary the event-study post-period average is approximately $-0.028$ against the simple ATT of $-0.047$.

# 6 Conclusion

This paper examines whether fiscal constraints on public sector compensation reduce service quality. I analyze New Jersey's two percent arbitration cap on police salary awards, which cut average annual raises from 3.9 percent to 1.85 percent. The setting offers a test of reciprocity and behavioral responses to wage constraints in an occupation where effort is discretionary and performance is difficult to monitor.

The cap bound, and it bound on the margin officers bargain over. Using the negotiated base salary schedules of 460 municipal agencies, not only those that reached arbitration, I find that it reduced annual scheduled increases by 1.99 percentage points against a pre-cap mean of 3.6 percent, and raised the probability that an agency's increase fell below 2 percent by 71 percentage points. The event study shows no pre-trend and a sharp break in the year a contract first falls under the cap.

Against that first stage, the performance results are small and imprecise. Under the preferred specification, clearance rates decline by 1.9 to 2.6 percentage points, or about 9 to 13 percent of pre-cap baselines. The point estimates are negative for every crime type, under both heterogeneity-robust estimators, and under both the never-treated and not-yet-treated comparison groups. None is statistically distinguishable from zero. The confidence intervals run from gains of about 3 percentage points to declines of about 7 to 8, and weighting agencies by offense volume, which produces the most precise estimates in the paper, leaves the point estimates similar without moving them clear of zero. A state-level synthetic difference-in-differences comparison, reported in the appendix as an aggregate check, is also small and insignificant.

The honest reading is that a policy which demonstrably compressed pay produced no detectable response in measured police performance. I do not read the uniformly negative signs as evidence of a cost. The estimators share the same data and the outcomes are correlated, so their agreement is not independent confirmation and does not narrow the interval. Nor do I read the insignificance as evidence of no effect. Declines of the magnitude

reported for individual arbitration disappointments sit inside every interval in the paper, including the weighted one. What the data exclude is deterioration much larger than that, and the catastrophic outcomes sometimes invoked in debates over compensation limits.

Two readings of the small response deserve emphasis. The first concerns the channel. A uniform, pre-announced ceiling lowers the level of everyone's raise without staging a defeat against a specific counteroffer, so if the reciprocal effort response is triggered by the experience of losing rather than by the size of the raise, a muted response is what reference dependence predicts. The data give this reading little direct support, since the response does not grow with the size of the visible loss or with how far an award fell below the union's offer, although these tests have limited power. The second concerns aggregation. If municipal budgets continued to grow while base salary growth was capped, freed funds could expand staffing or paid hours, and agency-level clearance could hold steady through capacity even with per-officer effort unchanged or lower.

One further reading can be set aside. If the value squeezed out of base salary had reappeared elsewhere in the contract, officers would never have been made worse off. Section 5 shows that it did not. The cap lowered the entry wage written into contracts by about 4.5 percent, and that compression did not resurface in the uncapped non-salary margins, such as uniform allowance, tuition stipends, and leave, where every estimate is within its standard error of zero. At least four fifths of the value the cap removed is unaccounted for on the margins the contract records. Supervisory wage levels rise modestly, but they are themselves base-salary items inside the cap and the estimates are fragile to small violations of parallel trends.

The data do not adjudicate between the two readings that remain. I observe contracts, awards, and clearance, not officer expectations, perceived fairness, morale, or staffing rosters. Distinguishing expectation adjustment from capacity expansion, or from compensation paid through budgetary channels the contract does not record, such as overtime and court time, would require survey measurement of officer attitudes or richer staffing and payroll data.

A further caveat concerns selection into arbitration. About forty percent of New Jersey contract renewals proceed to arbitration and the remainder settle voluntarily. The analysis treats agencies as treated when their contracts expire during the cap window, regardless of settlement type. If voluntary settlements after 2011 also bargained in the shadow of the cap, then both groups incorporate some cap-related adjustment and the estimates are biased toward zero. That possibility is a reason to read the small magnitudes with caution rather than as a settled null.

The paper makes four contributions. First, it studies police arbitration under a different rule, conventional arbitration with a statutory cap on awards, and measures performance with the clearance rate as well as clearances per capita. Second, it compares a binding, transparent, uniformly applied compensation constraint with idiosyncratic pay disappointments in the same state and finds declines of similar size under both, neither precisely estimated, so the evidence does not support the view that a predictable constraint is less costly. Third, it assembles thirty years of New Jersey police contracts and arbitration records. Their expiration dates supply within-state variation in treatment timing that a single aggregate comparison discards, and they allow the cap's first stage to be measured on all agencies in the sample rather than on the arbitrating subset. Fourth, it shows that a binding cap on one priced margin of a public-sector contract need not push value into the adjacent margins it leaves free, in a setting where compensation-spillover reasoning predicts exactly that.

For policymakers the implication is narrower than the sharpness of the reform might suggest. The cap achieved its fiscal objective, compressing negotiated wage growth substantially and durably. Whether it carried a performance cost is not settled by these data. The point estimates lean negative and the intervals are wide enough to accommodate both a true zero and a decline worth caring about. Where performance is hard to monitor and the constraint is announced in advance and applied to everyone, the presumption that pay compression translates directly into service quality losses deserves more evidence than it usually receives.

# 7 Tables and Figures

Table 1: Arbitration Outcomes and Clearance: Clearances per 100k vs. Clearance Rate (Balanced Sample)

| | Panel A: Per 100,000 capita | | | Panel B: Clearance rate (%) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Postarbitration × Employer win | -9.450 | -0.821 | -3.146 | -2.082 | -3.742 | -2.192 |
| | (11.255) | (3.047) | (2.078) | (2.171) | (1.468) | (1.163) |
| Arbitration window FE | No | Yes | Yes | No | Yes | Yes |
| Month × Year FE | No | Yes | No | No | Yes | No |
| Observations | 3,297 | 3,297 | 3,297 | 3,297 | 3,297 | 3,297 |
| $R^2$ | 0.002 | 0.669 | 0.630 | 0.001 | 0.316 | 0.252 |
| Mas (2006) range | -6.79 to -9.75 | | | – | | |

*Notes:* Standard errors clustered on arbitration window in parentheses. Balanced sample includes only observations with non-missing values for both measures. Panel A dependent variable is clearances per 100,000 capita. Panel B dependent variable is clearance rate = (total clearances / total index crimes) × 100. Mas (2006) reported -6.79 to -9.75 clearances per 100k.

Table 2: Synthetic Difference-in-Differences at State Level Results

| Statistic | Value |
|---|---|
| Average Treatment Effect (ATT) | -0.013862 |
| Standard Error | 0.062610 |
| 95% Confidence Interval | [-0.136577, 0.108853] |
| Pre-treatment period | 1996–2010 |
| Post-treatment period | 2011–2018 |
| Number of donor states | 48 |
| Time placebo p-value | 0.0002 |
| Space placebo p-value | 0.8163 |

*Notes:* This table summarizes the revised SDID estimates after removing states with poor pre-treatment fit (HI, KS). The ATT represents the average difference in clearance rates between New Jersey and its synthetic control during the post-treatment period, in proportion units ($-0.0139$ is a decline of about 1.4 percentage points). Standard errors are computed using placebo inference. Time placebo p-value tests temporal specificity; space placebo p-value tests whether the effect magnitude is unusual relative to cross-state variation.

Table 3: First Stage: Effect of the 2 Percent Cap on Negotiated Salary Schedules

| Outcome | ATT | Std. Error | 95% CI | Pre-cap mean |
|---|---|---|---|---|
| *Panel A: Not-yet-treated comparison group (preferred specification)* | | | | |
| Annual increase (%) | -1.987*** | (0.197) | [-2.37, -1.60] | 3.55 |
| Pr(increase $<$ 2%) | 0.714*** | (0.054) | [ 0.61, 0.82] | 0.155 |
| *Panel B: Never-treated comparison group* | | | | |
| Annual increase (%) | -1.672*** | (0.168) | [-2.00, -1.34] | 3.55 |
| Pr(increase $<$ 2%) | 0.534*** | (0.042) | [ 0.45, 0.62] | 0.155 |
| *Panel C: Event-study averages (not-yet-treated), annual increase* | | | | |
| Pre-period average | 0.221 | (0.319) | [-0.40, 0.85] | |
| Post-period average | -1.978*** | (0.235) | [-2.44, -1.52] | |

**Notes:** *** p<0.01, ** p<0.05, * p<0.1. The outcome is the year over year percent change in the entry step of the base salary schedule, constructed from collected collective bargaining agreements. The estimation sample covers 460 agencies and 5,077 agency-year observations from 2005 to 2020; the estimator uses the 4,317 pair-balanced observations in Panels A and C and the 3,942 pair-balanced observations in Panel B. All estimates use the Callaway & Sant'Anna (2021) estimator with regression adjustment, treatment timing given by the year an agency's contract first falls under the cap, and the same cohort definitions used for the clearance outcomes. Panel C reports the pre- and post-period averages from the event-study aggregation; none of the eleven individual pre-period coefficients is statistically distinguishable from zero. The pre-cap mean column reports the raw mean of each outcome across agency-years before an agency is capped.

Table 4: Treatment Effects on Clearance Rates: Method Comparison

| Crime Type | Method | ATT | Std. Error | 95% CI | P-value | N Agencies |
|---|---|---|---|---|---|---|
| *Panel A: Never-treated comparison group (both estimators)* | | | | | | |
| Property | CS | -2.490 | (2.923) | [-8.22, 3.24] | 0.394 | 463 |
| | Sun-Abraham | -1.845 | (2.542) | [-6.83, 3.14] | 0.468 | 463 |
| Theft | CS | -3.292 | (3.140) | [-9.45, 2.86] | 0.294 | 463 |
| | Sun-Abraham | -1.575 | (2.743) | [-6.95, 3.80] | 0.566 | 463 |
| Total | CS | -1.946 | (2.913) | [-7.65, 3.76] | 0.504 | 463 |
| | Sun-Abraham | -1.512 | (2.477) | [-6.37, 3.34] | 0.542 | 463 |
| *Panel B: Not-yet-treated comparison group (preferred specification)* | | | | | | |
| Property | CS | -2.303 | (2.503) | [-7.21, 2.60] | 0.357 | 463 |
| Theft | CS | -2.583 | (2.660) | [-7.80, 2.63] | 0.331 | 463 |
| Total | CS | -1.949 | (2.489) | [-6.83, 2.93] | 0.433 | 463 |
| *Panel C: Not-yet-treated, weighted by offense volume* | | | | | | |
| Property | CS | -1.715 | (1.176) | [-4.02, 0.59] | 0.145 | 463 |
| Theft | CS | -2.286 | (1.423) | [-5.08, 0.50] | 0.108 | 463 |
| Total | CS | -1.337 | (1.014) | [-3.32, 0.65] | 0.187 | 463 |

**Notes:** *** p<0.01, ** p<0.05, * p<0.1. CS refers to Callaway & Sant'Anna (2021) with bootstrap standard errors and bootstrap confidence intervals. Sun-Abraham refers to the interaction-weighted estimator of Sun & Abraham (2021), implemented in `eventstudyinteract`, aggregated to an overall ATT using the same event-time weights as the Callaway-Sant'Anna aggregation, with standard errors clustered by agency. Panel A holds the comparison group fixed at never-treated agencies so that the two estimators are directly comparable; the Sun-Abraham estimator requires a never-treated or last-treated comparison cohort and cannot use a rolling not-yet-treated control group. Panel B reports the preferred specification, which uses all not-yet-treated agencies as controls at each point in time as described in Section 3.1. Panel C repeats the preferred specification weighting each agency by its mean annual index-property offense count, since clearance rates computed on few offenses are mechanically noisy. Weights enter through the `iweight` option, which `csdid` supports directly. Sample spans 1996–2020 with 463 agencies and 10,511 agency-year observations (10,517 in Panels B and C).

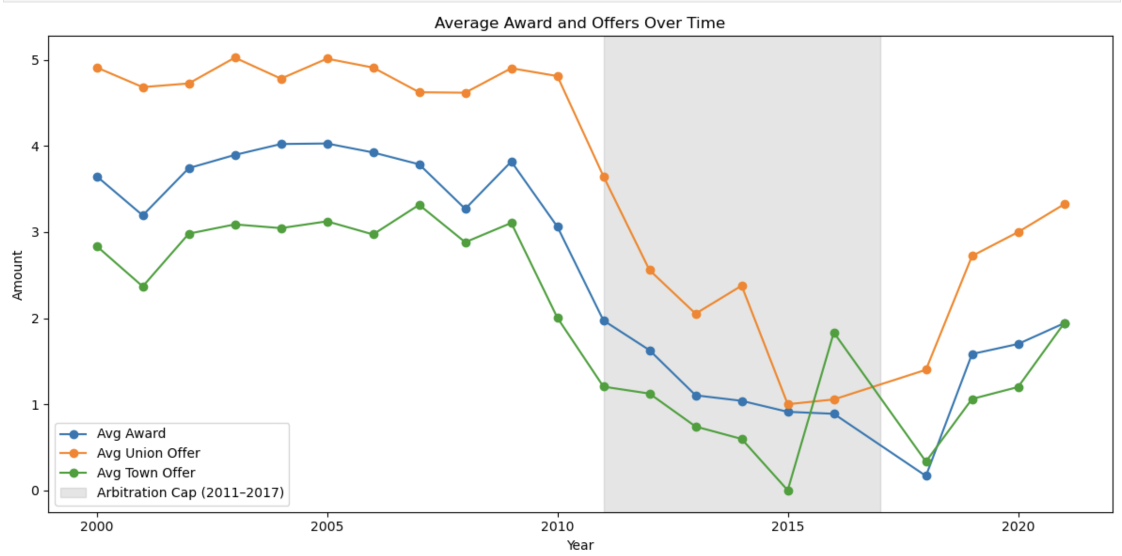



Figure 1: Average Arbitration Awards and Final Offers Over Time

*Notes:* This figure plots average arbitration awards alongside union and municipality final offers over time. During the cap period (2011–2017, shaded), the average arbitration award dropped alongside both the union and town offers, as unions lowered their demands toward the constraint. Values are average annual percentage increases.

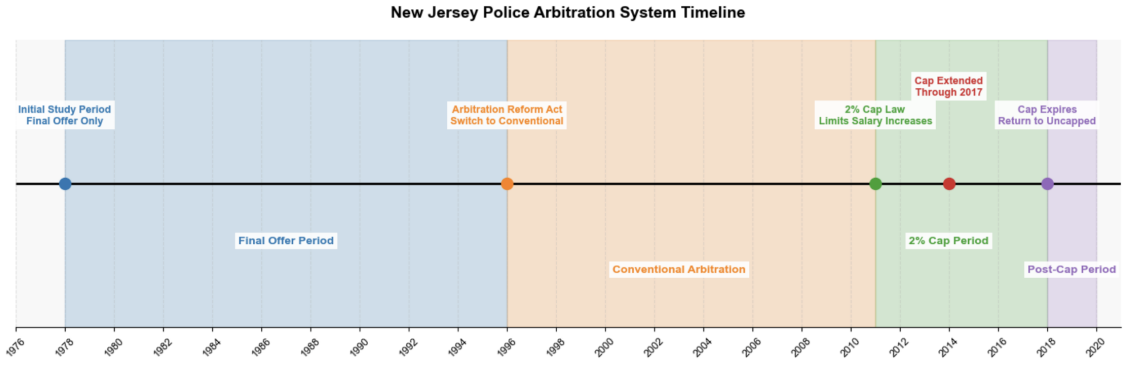
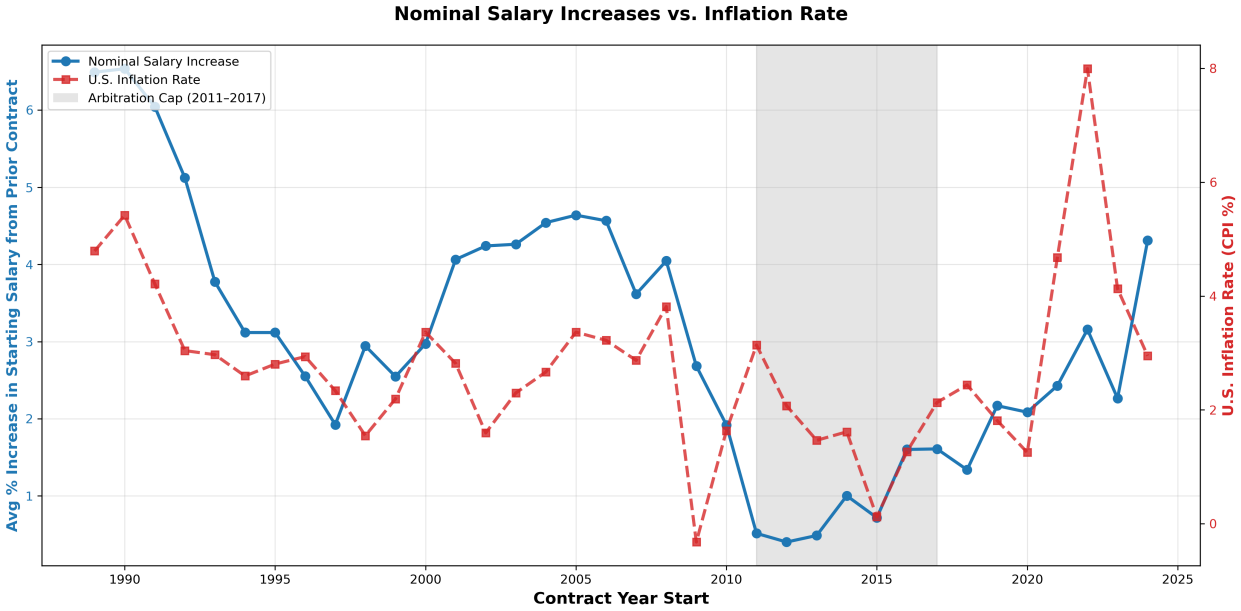



Figure 2: Arbitration Cap Timeline and Structure

*Notes:* This figure illustrates the timeline and policy structure of New Jersey's two percent arbitration cap, which was in effect from 2011 through 2017. The cap imposed a statutory ceiling on salary increases that arbitrators could award in police contract disputes during this period.




(a) Nominal Salary Increases vs. Inflation Rate




(b) Real Wage Growth (Inflation-Adjusted)

*Notes:* Panel (a) displays the average percentage increase in starting salaries from the prior contract (blue line) compared to the U.S. inflation rate (red dashed line) across New Jersey police agencies over time. Panel (b) shows the corresponding real wage growth, calculated as salary increases minus inflation. Before 2011, nominal raises often exceeded inflation, resulting in positive real wage growth. From 2011 to 2017 (gray shaded area), during the arbitration cap period, nominal increases mostly stayed at or below about one percent while inflation was higher in most years, resulting in periods of negative real wage growth. This suggests that the cap served as a binding constraint on compensation growth, effectively reducing real purchasing power for many officers during this period.

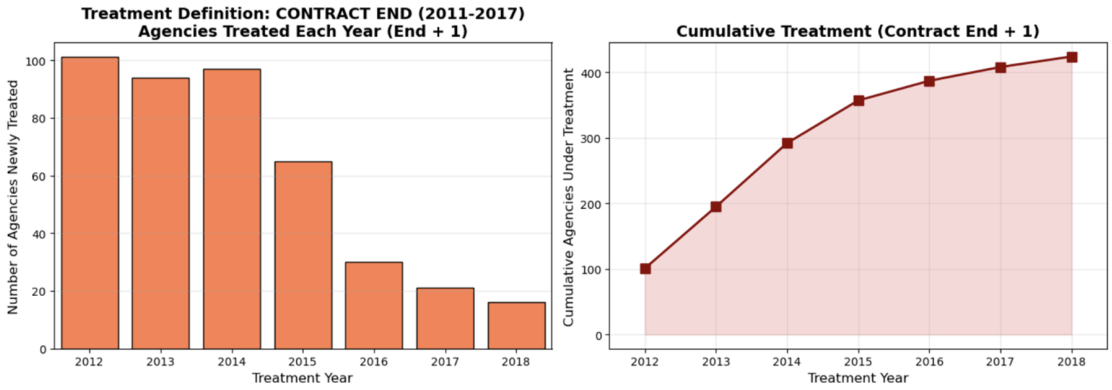



Figure 4: Contract End

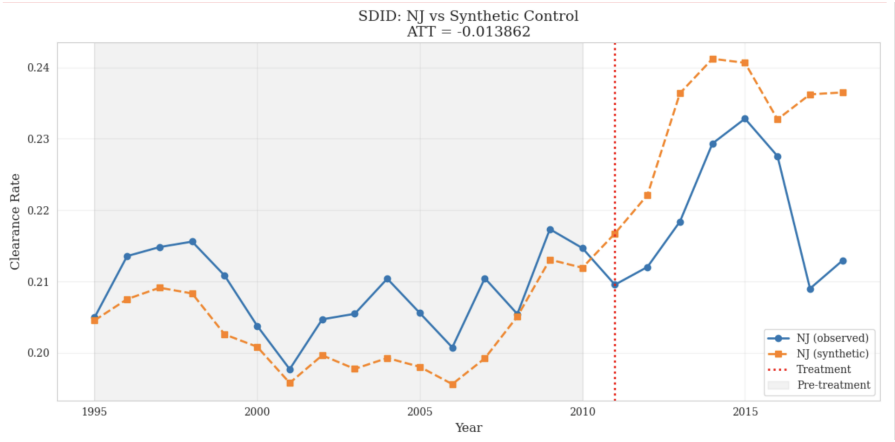



Figure 5: State-Level Synthetic Difference-in-Differences: New Jersey vs. Synthetic Control

*Notes:* This figure displays observed clearance rates for New Jersey (solid blue line) compared to the synthetic control (dashed orange line) constructed using weighted donor states. The vertical red dotted line marks policy implementation in 2011. The shaded gray area indicates the pre-treatment period used to construct synthetic control weights.

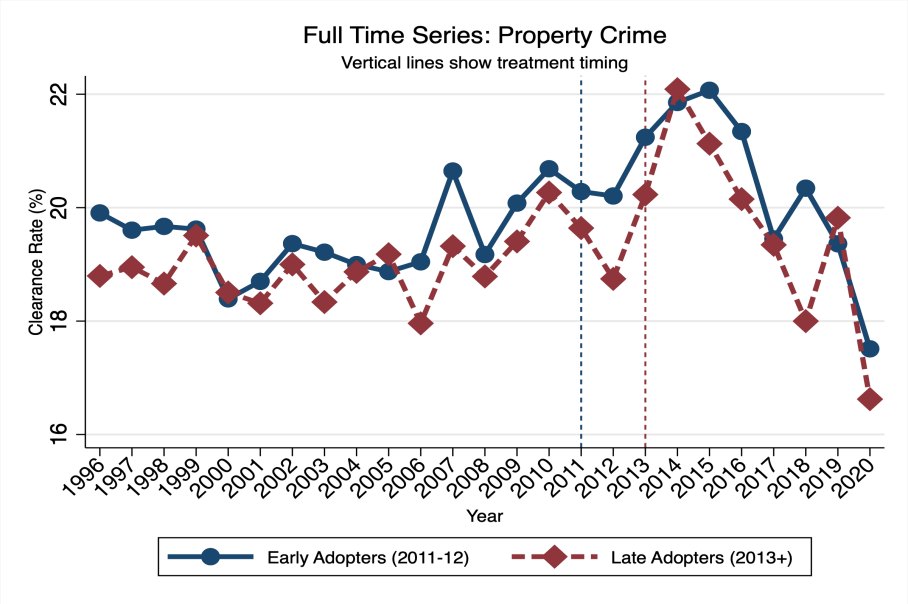



Figure 6: Full Time Series: Property Crime Clearance Rates

*Notes:* This figure shows the complete time series of property crime clearance rates for early adopters (contracts expiring 2011–2012) and late adopters (contracts expiring 2013+). Vertical lines mark treatment timing for each group. The figure displays trends both before and during treatment to assess parallel trends and post-treatment divergence.

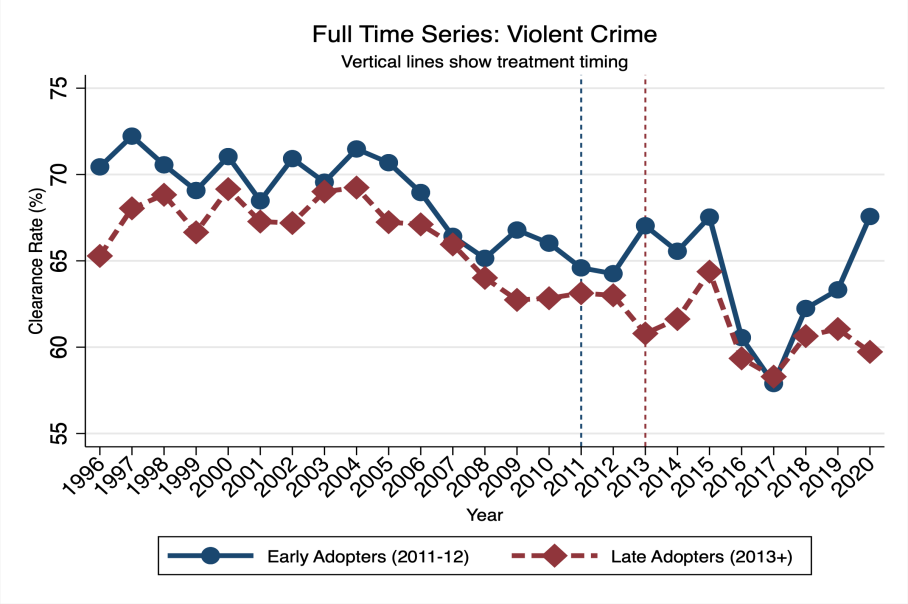



Figure 7: Full Time Series: Violent Crime Clearance Rates

*Notes:* This figure shows the complete time series of violent crime clearance rates for early and late adopters. Vertical lines indicate when each group first faced the arbitration cap. Visual inspection helps assess whether parallel trends held before treatment and whether clearance rates diverged afterward.

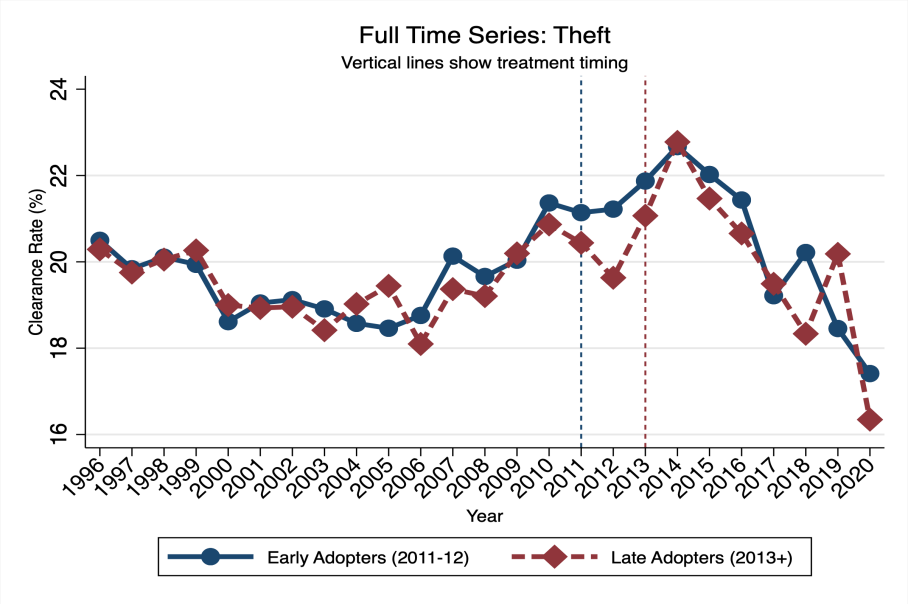



Figure 8: Full Time Series: Theft Clearance Rates

*Notes:* This figure shows the complete time series of theft clearance rates for early and late adopters, with vertical lines marking treatment timing. The full time series provides context for interpreting pre-treatment parallel trends and post-treatment effects.

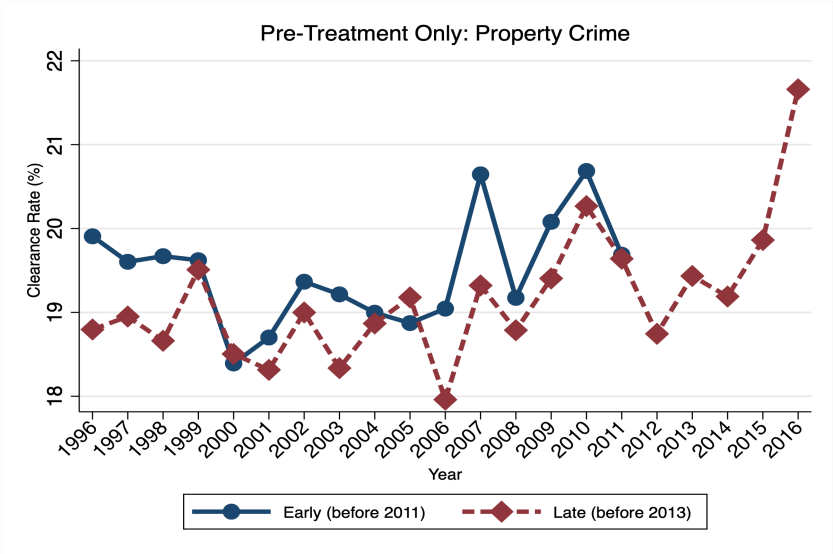



Figure 9: Pre-Treatment Trends: Property Clearance Rates by Adoption Timing

*Notes:* This figure plots mean property clearance rates for early and late adopters during their respective pre-treatment periods only. Early adopters are shown before 2011; late adopters before 2013. Parallel lines would support the parallel trends assumption required for difference-in-differences identification. Any divergence suggests differential pre-treatment trajectories that may complicate causal interpretation.

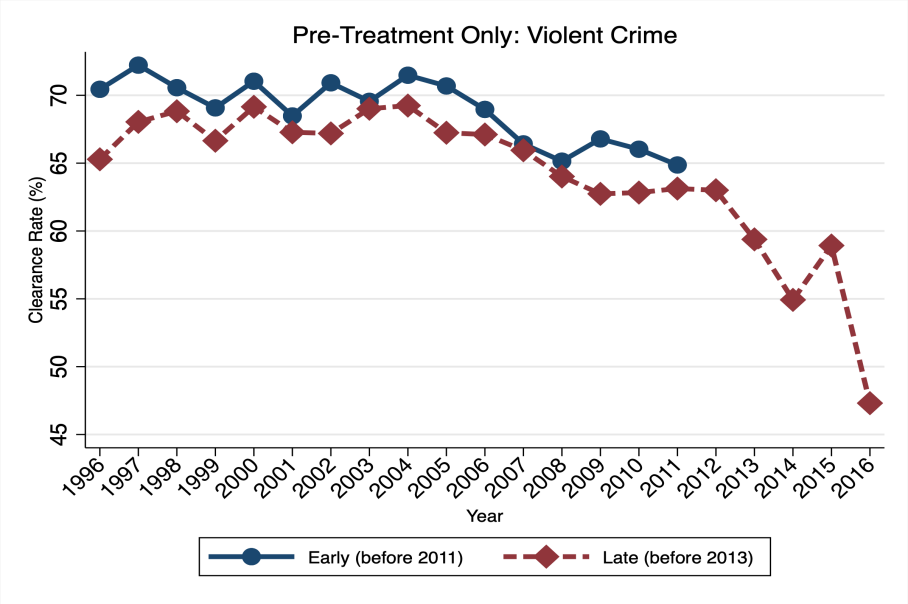



Figure 10: Pre-Treatment Trends: Violent Crime Clearance Rates by Adoption Timing
*Notes:* This figure plots mean violent crime clearance rates for early and late adopters during their respective pre-treatment periods. The figure isolates pre-treatment trends to assess whether the parallel trends assumption is satisfied. Divergent trends would indicate that treatment and control groups were on different trajectories even before policy exposure.

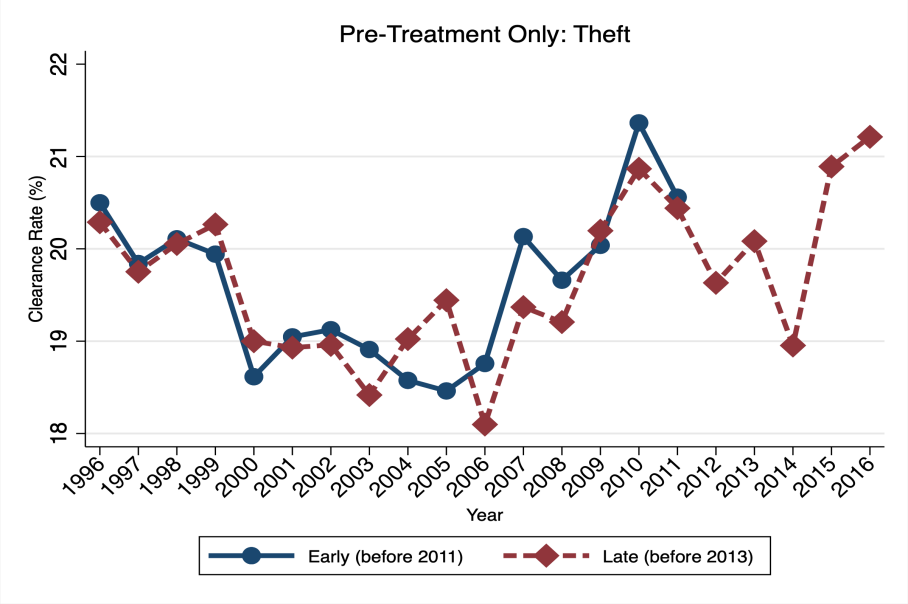



Figure 11: Pre-Treatment Trends: Theft Clearance Rates by Adoption Timing
*Notes:* This figure plots mean theft clearance rates for early and late adopters during pre-treatment periods. Parallel pre-treatment trends are essential for valid difference-in-differences estimation. Visual inspection and formal tests help determine whether this assumption holds.

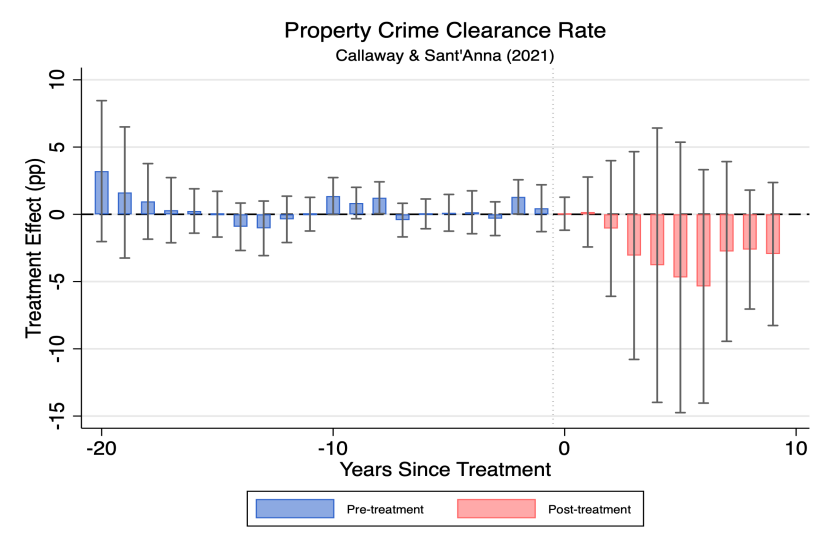



Figure 12: Event-Study Estimates: Callaway–Sant'Anna (Property Clearance Rate)
*Notes:* This figure displays dynamic treatment effects for property clearance rates estimated using the Callaway and Sant'Anna (2021) method. Blue bars represent pre-treatment periods; red bars represent post-treatment periods. Error bars indicate 95% confidence intervals. Years since treatment are defined relative to contract expiration during the arbitration cap period. Pre-treatment coefficients test for differential trends; post-treatment coefficients estimate dynamic treatment effects.

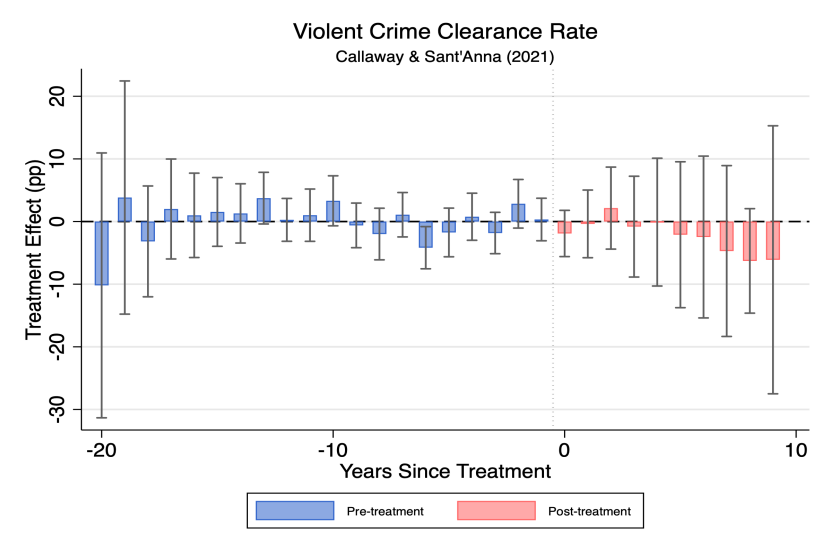



Figure 13: Event-Study Estimates: Callaway–Sant'Anna (Violent Crime Clearance Rate)
*Notes:* This figure displays dynamic treatment effects for violent crime clearance rates using the Callaway and Sant'Anna (2021) method. The estimator aggregates cohort-specific average treatment effects using inverse-probability weights. Pre-treatment coefficients help assess parallel trends; post-treatment coefficients show how effects evolve over time.

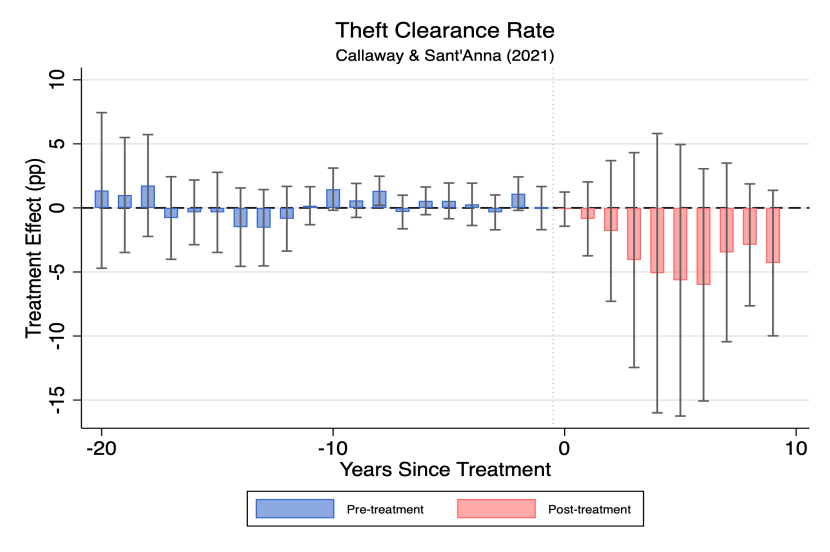



Figure 14: Event-Study Estimates: Callaway–Sant'Anna (Theft Clearance Rate)
*Notes:* This figure displays dynamic treatment effects for theft clearance rates using the Callaway and Sant'Anna (2021) method. The event study design allows treatment effects to vary across time since treatment, capturing heterogeneous dynamics in policy response.

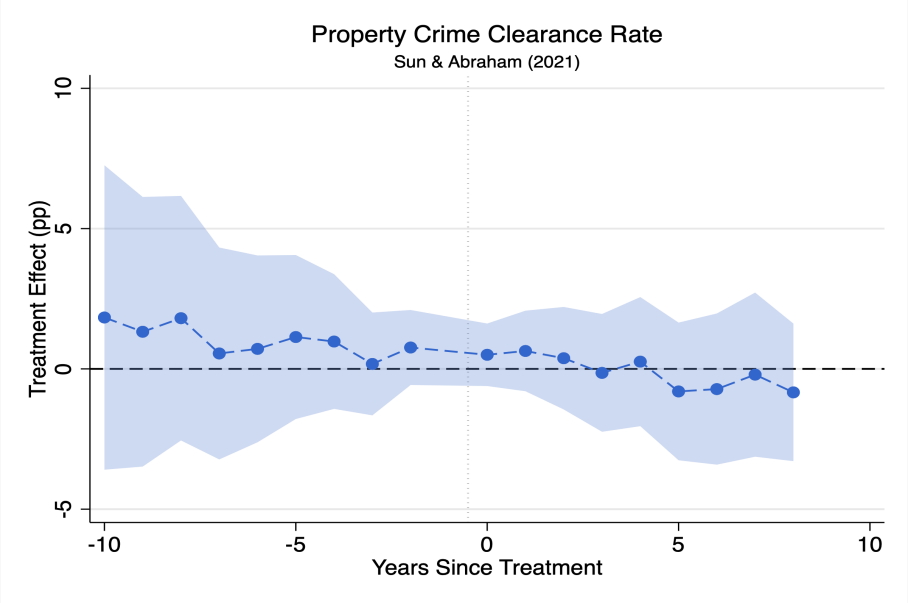



Figure 15: Event-Study Estimates: Sun–Abraham (Property Clearance Rate)
*Notes:* This figure displays dynamic treatment effects for property clearance rates estimated using the Sun and Abraham (2021) interaction-weighted estimator. The shaded area represents the 95% confidence interval. The estimator uses the last-treated cohort as a comparison group and reweights to avoid contamination bias from already-treated units.

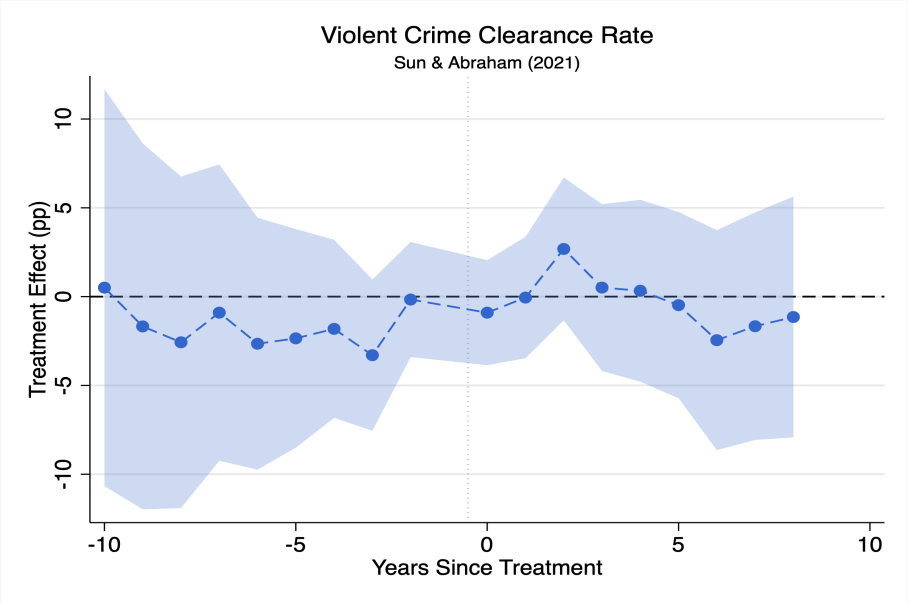



Figure 16: Event-Study Estimates: Sun–Abraham (Violent Crime Clearance Rate)
*Notes:* This figure displays dynamic treatment effects for violent crime clearance rates using the Sun and Abraham (2021) method. The interaction-weighted approach addresses concerns about heterogeneous treatment effects and negative weighting in standard two-way fixed effects models.

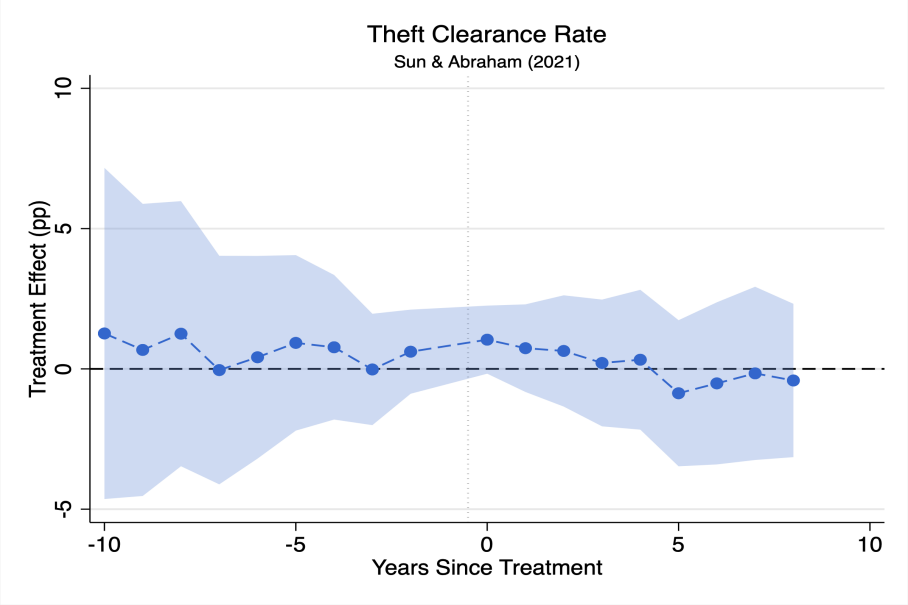



Figure 17: Event-Study Estimates: Sun–Abraham (Theft Clearance Rate)
*Notes:* This figure displays dynamic treatment effects for theft clearance rates using the Sun and Abraham (2021) method. Pre-treatment coefficients test the parallel trends assumption; post-treatment coefficients reveal how effects evolve in years following contract expiration.

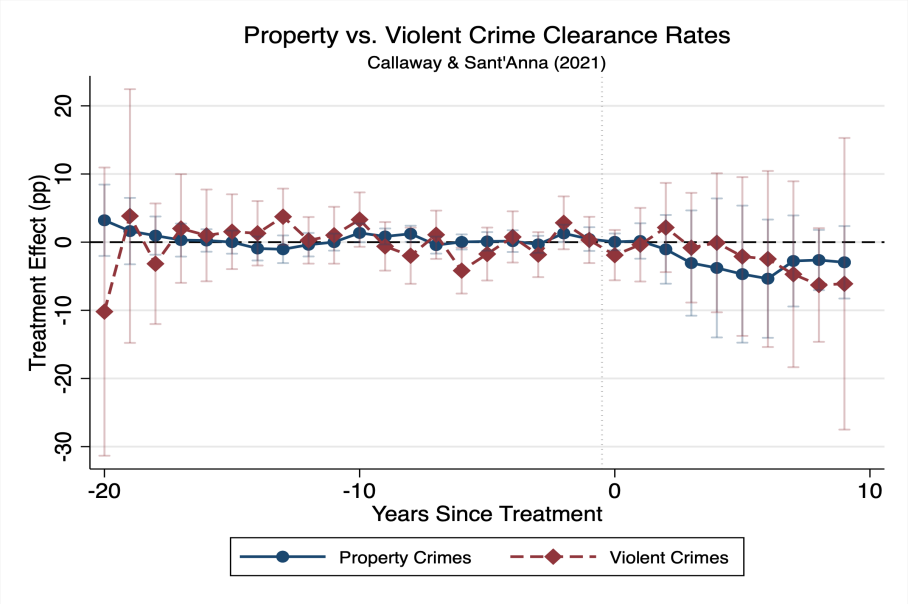



Figure 18: Property vs. Violent Crime Clearance Rates: Event-Study Comparison

*Notes:* This figure compares event-study estimates for property crimes (solid navy line) and violent crimes (dashed maroon line) using the Callaway and Sant'Anna (2021) method. Confidence intervals are shown with error bars. Violent crime estimates are less precise.

## Tables for Section 5

Table 5: Callaway–Sant'Anna ATT Estimates on Contract Components

| Outcome | ATT | SE | $\lvert t\rvert$ | $p$ | $p_{\text{Holm}}$ | $N$ |
|---|---|---|---|---|---|---|
| *Dollar value (log of CPI-deflated $)* | | | | | | |
| Starting salary (log) | -0.0465*** | (0.0079) | 5.88 | 0.000 | 0.000 | 9,006 |
| Sergeant wage (log) | 0.0130** | (0.0038) | 3.39 | 0.001 | 0.018 | 5,363 |
| Lieutenant wage (log) | 0.0057 | (0.0048) | 1.19 | 0.233 | 1.000 | 3,190 |
| Captain wage (log) | 0.0190** | (0.0057) | 3.35 | 0.001 | 0.020 | 1,870 |
| Tuition stipend, Associate (log) | -0.0055 | (0.0099) | 0.56 | 0.578 | 1.000 | 3,695 |
| Tuition stipend, BA (log) | 0.0015 | (0.0087) | 0.17 | 0.865 | 1.000 | 3,696 |
| Tuition stipend, MA (log) | -0.0198 | (0.0101) | 1.95 | 0.051 | 1.000 | 1,595 |
| Tuition reimbursement value (log) | 0.0041 | (0.0222) | 0.18 | 0.854 | 1.000 | 801 |
| Uniform allowance (log) | 0.0132 | (0.0165) | 0.80 | 0.424 | 1.000 | 6,371 |
| Holiday count (log) | -0.0075 | (0.0081) | 0.92 | 0.357 | 1.000 | 8,348 |
| Sick leave (log) | -0.0223 | (0.0131) | 1.70 | 0.089 | 1.000 | 8,759 |
| Vacation at 5 yrs (log) | 0.0038 | (0.0108) | 0.35 | 0.724 | 1.000 | 8,952 |
| Vacation at 10 yrs (log) | 0.0051 | (0.0170) | 0.30 | 0.765 | 1.000 | 7,885 |
| Vacation at 15 yrs (log) | 0.0045 | (0.0176) | 0.25 | 0.800 | 1.000 | 7,831 |
| Vacation at 20 yrs (log) | -0.0079 | (0.0120) | 0.66 | 0.511 | 1.000 | 7,444 |
| Longevity at 4 yrs (log) | 0.1522 | (0.1839) | 0.83 | 0.408 | 1.000 | 1,558 |
| Longevity at 8 yrs (log) | -0.0282 | (0.0761) | 0.37 | 0.711 | 1.000 | 6,459 |
| Longevity at 12 yrs (log) | -0.0593 | (0.0879) | 0.67 | 0.500 | 1.000 | 7,019 |
| Longevity at 16 yrs (log) | -0.0404 | (0.0917) | 0.44 | 0.660 | 1.000 | 6,931 |
| Longevity at 20 yrs (log) | -0.0634 | (0.0958) | 0.66 | 0.508 | 1.000 | 6,525 |
| Longevity at 24+ yrs (log) | 0.0132 | (0.1079) | 0.12 | 0.903 | 1.000 | 5,342 |
| *Presence and first introduction* | | | | | | |
| Has tuition incentive | -0.0256 | (0.0199) | 1.29 | 0.198 | 1.000 | 9,194 |
| Introduced tuition incentive | 0.0020 | (0.0119) | 0.17 | 0.869 | 1.000 | 8,984 |
| Has tuition increment | -0.0151 | (0.0106) | 1.43 | 0.154 | 1.000 | 9,199 |
| Introduced tuition increment | 0.0000 | (0.0000) | — | — | — | 9,197 |
| Has tuition reimbursement | -0.0093 | (0.0119) | 0.78 | 0.437 | 1.000 | 9,183 |
| Introduced tuition reimbursement | 0.0088 | (0.0043) | 2.06 | 0.039 | 0.938 | 8,989 |

*Notes:* Callaway–Sant'Anna (2021) staggered DID with not-yet-treated agencies as the comparison group, a set that nests the never-treated. Treated cohort $G_i \in [2010, 2017]$ from the year of first contract expiration during the cap window. Standard errors clustered at the agency level. $p_{\text{Holm}}$ is the Holm step-down adjusted $p$-value across the family of 27 outcomes. Stars from Holm: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Dollar outcomes in log of CPI-deflated 2010 $; presence and introduction outcomes are linear-probability.

Table 6: Compensation Accounting: Where Did the Cap's Bite Go?

| *Panel A: Cap's compression across the wage scale* | | | | | |
|---|---|---|---|---|---|
| Wage step | Pre-cap median ($) | ATT (log) | $ /officer/yr | Workforce share | Contribution ($) |
| Year 1 | 43,965 | -0.0465 | -1,999 | 0.10 | -200 |
| Year 5 | 70,911 | -0.0749 | -5,119 | 0.30 | -1,536 |
| Year 10 | 82,933 | -0.0295 | -2,408 | 0.25 | -602 |
| Year 16 | 80,067 | -0.0267 | -2,109 | 0.35 | -738 |
| **Total weighted compression** | | | | | **-3,076** |
| *Panel B: Observable substitution offsets* | | | | | |
| Channel | | ATT | Per-rank $ | Share | Contribution ($) |
| Sergeant wage | | +0.0130 | +1,115 | 0.20 | +223 |
| Lieutenant wage | | +0.0057 | +553 | 0.05 | +28 |
| Captain wage | | +0.0190 | +1,996 | 0.03 | +60 |
| Education stipend (BA) | | +0.0015 | +2 | 1.00 | +2 |
| Education stipend (MA) | | -0.0198 | -32 | 1.00 | -32 |
| Longevity (8 yr step) | | -0.0282 | -0 | 1.00 | -0 |
| Longevity (16 yr step) | | -0.0404 | -0 | 1.00 | -0 |
| Uniform allowance | | +0.0132 | +12 | 1.00 | +12 |
| Holiday count | | -0.0075 | -0 | 1.00 | -0 |
| Sick leave | | -0.0223 | -0 | 1.00 | -0 |
| **Total observable offsets** | | | | | **+292** |
| *Panel C: Balance sheet* | | | | | |
| Compression: weighted dollars lost across wage scale | | | | | **-3,076** |
| Offsets: observable substitution into other components | | | | | **+292** |
| Net change to officer compensation | | | | | **-2,784** |
| Unaccounted ($ compression not offset by observables) | | | | | **+2,784** |

*Notes:* Per-officer-year dollar magnitudes computed from Callaway–Sant'Anna ATT estimates and pre-cap (2005–2010) median wages or compensation values. Wage-step workforce shares are rough assumptions about the tenure distribution of a typical NJ municipal police agency. *Workforce share* for rank-level outcomes uses approximate shares (Sergeant 20%, Lieutenant 5%, Captain 3%). Contributions are point estimates times dollar base times share; longevity, holiday, and sick-leave components are listed at zero dollars. All non-salary components have Holm-adjusted $p$-values above 0.99 (see Table 5), so their contributions are not distinguishable from zero.

Table 7: Heterogeneity: Substitution Effects by Civil-Service Governance

| | Non-civil-service | | Civil-service | | Diff. |
|---|---|---|---|---|---|
| Outcome | ATT | SE | ATT | SE | $p$ |
| Starting salary (log) | +0.0017 | (0.0126) | -0.0571 | (0.0113) | 0.000 |
| Sergeant wage (log) | +0.0178 | (0.0044) | -0.0034 | (0.0068) | 0.009 |
| Captain wage (log) | +0.0271 | (0.0079) | -0.0049 | (0.0049) | 0.001 |

*Notes:* Callaway–Sant'Anna ATT estimated separately on the non-civil-service and civil-service subgroups, using the same contract-expiration-cohort treatment definition as the main results. Standard errors clustered at the agency level. The Diff. column reports the two-sided $p$-value for equality of subgroup ATTs, computed as $z = (\hat{\tau}^{\text{NCS}} - \hat{\tau}^{\text{CS}}) / \sqrt{\widehat{\text{Var}}(\hat{\tau}^{\text{NCS}}) + \widehat{\text{Var}}(\hat{\tau}^{\text{CS}})}$, which assumes independence of the subgroup estimates. The civil-service flag is from the NJ Law Enforcement Diversity Data, merged via fuzzy name matching within county (token-set ratio threshold 85; 97% match rate).

Table 8: Robustness: Standard MAR vs. LOCF Inheritance Imputation on Non-Wage Compensation Outcomes

| | MAR (main spec) | | | LOCF (inheritance) | | |
|---|---|---|---|---|---|---|
| Outcome | ATT | SE | $N$ | ATT | SE | $N$ |
| Tuition stipend, Associate | -0.0055 | (0.0099) | 3,695 | +0.0076 | (0.0110) | 4,126 |
| Tuition stipend, BA | +0.0015 | (0.0087) | 3,696 | +0.0134 | (0.0100) | 4,052 |
| Tuition stipend, MA | -0.0198* | (0.0101) | 1,595 | -0.0056 | (0.0119) | 1,706 |
| Tuition reimbursement value | +0.0041 | (0.0222) | 801 | +0.0269 | (0.0212) | 921 |
| Longevity at 4 yrs | +0.1522 | (0.1839) | 1,558 | +0.0546 | (0.1125) | 2,046 |
| Longevity at 8 yrs | -0.0282 | (0.0761) | 6,459 | +0.0047 | (0.0512) | 8,192 |
| Longevity at 12 yrs | -0.0593 | (0.0879) | 7,019 | -0.0646 | (0.0710) | 8,430 |
| Longevity at 16 yrs | -0.0404 | (0.0917) | 6,931 | -0.0342 | (0.0758) | 8,281 |
| Longevity at 20 yrs | -0.0634 | (0.0958) | 6,525 | -0.0579 | (0.0793) | 7,745 |
| Longevity at 24+ yrs | +0.0132 | (0.1079) | 5,342 | +0.0176 | (0.0785) | 6,363 |
| Uniform allowance | +0.0132 | (0.0165) | 6,371 | +0.0020 | (0.0157) | 7,312 |
| Holiday count | -0.0075 | (0.0081) | 8,348 | -0.0002 | (0.0081) | 8,969 |
| Sick leave | -0.0223* | (0.0131) | 8,759 | -0.0229* | (0.0128) | 8,866 |
| Vacation at 5 yrs | +0.0038 | (0.0108) | 8,952 | +0.0039 | (0.0107) | 9,051 |
| Vacation at 10 yrs | +0.0051 | (0.0170) | 7,885 | +0.0056 | (0.0170) | 7,955 |
| Vacation at 15 yrs | +0.0045 | (0.0176) | 7,831 | +0.0051 | (0.0176) | 7,904 |
| Vacation at 20 yrs | -0.0079 | (0.0120) | 7,444 | -0.0074 | (0.0119) | 7,518 |

*Notes:* Callaway–Sant'Anna ATT estimates on the log of CPI-deflated dollar outcomes. The *MAR* columns replicate the main specification with missing values dropped from the estimation sample. The *LOCF* columns apply last-observation-carried-forward within agency to fill missing values that follow a previously-coded positive value, operationalizing the continuation-clause interpretation of New Jersey's past-practice doctrine. Standard errors clustered at the agency level. Stars: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. No outcome is distinguishable from zero at the 5 percent level under either treatment of missing data. Sick leave is marginally negative under both, and the master's-tier stipend is marginal under MAR only, so neither points toward substitution into non-wage compensation.

# 8 Appendix A: Supplementary Tables

Table 9: Cohort-Specific Treatment Effects: Callaway & Sant'Anna (2021)

| Crime Type | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|---|---|
| Property | -1.953 | -1.532 | -0.706 | -7.293 | -5.032 | -2.047 | -0.496 |
| | (2.447) | (3.243) | (4.126) | (4.205) | (4.516) | (3.957) | (3.786) |
| Violent | -3.165 | -3.188 | -1.535 | 1.811 | 2.868 | 6.501 | 3.411 |
| | (5.374) | (5.487) | (6.820) | (6.205) | (7.116) | (9.497) | (8.578) |
| Theft | -4.556 | -2.761 | -0.380 | -6.688 | -4.779 | -1.004 | 2.725 |
| | (2.698) | (3.503) | (4.290) | (4.653) | (4.773) | (3.988) | (4.108) |

**Notes:** *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Each column represents the cohort of agencies whose contracts expired in that year (treatment timing under the arbitration cap).

Table 10: Cohort-Specific Treatment Effects: Sun & Abraham (2021)

| Crime Type | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|---|---|
| Property | -0.182 | -0.058 | 0.419 | -2.511 | -3.845 | -3.125 | -3.206 |
| | (2.289) | (2.818) | (3.209) | (3.713) | (4.196) | (3.097) | (2.518) |
| Violent | -0.663 | 0.128 | 4.064 | 5.406 | -0.829 | 5.882 | 1.305 |
| | (3.473) | (4.575) | (4.827) | (5.136) | (5.394) | (5.646) | (7.278) |
| Theft | -1.596 | -1.368 | 0.126 | -2.502 | -3.671 | -3.061 | -1.610 |
| | (2.416) | (2.977) | (3.294) | (3.982) | (4.379) | (3.070) | (3.598) |

**Notes:** *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Each column represents the cohort of agencies whose contracts expired in that year (treatment timing under the arbitration cap).

## 8.1 State-Level Synthetic Difference-in-Differences

This appendix reports the state-level synthetic difference-in-differences analysis. It compares New Jersey's aggregate clearance rate trajectory after 2011 with a weighted combination of control states. The design treats all New Jersey agencies as uniformly exposed from 2011 and therefore does not exploit the variation in contract expiration timing that identifies the staggered estimates in the main text. It is retained as an aggregate check rather than as a primary specification.

### 8.1.1 Empirical Strategy

Following Arkhangelsky et al. (2021), the SDID estimator can be expressed as a weighted double-difference. For unit weights $\omega \in \Omega$ and time weights $\lambda \in \Lambda$, the estimated average treatment effect on the treated is:

$$\hat{\tau}^{\text{sdid}}(\omega, \lambda) = \omega_{\text{tr}}^{\top} \mathbf{Y}_{\text{tr,post}} \lambda_{\text{post}} - \omega_{\text{co}}^{\top} \mathbf{Y}_{\text{co,post}} \lambda_{\text{post}} - \omega_{\text{tr}}^{\top} \mathbf{Y}_{\text{tr,pre}} \lambda_{\text{pre}} + \omega_{\text{co}}^{\top} \mathbf{Y}_{\text{co,pre}} \lambda_{\text{pre}}, \quad (2)$$

where $\mathbf{Y}_{\text{tr,post}}$ and $\mathbf{Y}_{\text{co,post}}$ denote the outcome matrices for treated and control units in post-treatment periods, and $\mathbf{Y}_{\text{tr,pre}}$ and $\mathbf{Y}_{\text{co,pre}}$ denote the corresponding pre-treatment matrices. The unit weights $\omega_{\text{co}}^{\text{sdid}}$ are chosen so that the weighted average of control units' pre-treatment outcomes tracks the treated unit's pre-treatment trajectory, while the time weights $\lambda_{\text{pre}}^{\text{sdid}}$ are chosen so that the weighted average of pre-treatment periods balances against the post-treatment period for control units. The standard difference-in-differences estimator is a special case in which all unit and time weights are uniform.

Equation (2) makes the double-differencing structure transparent: the treatment effect equals the weighted pre-to-post change for the treated unit minus the weighted pre-to-post change for the synthetic control.

To evaluate the aggregate effect of the arbitration cap, I compare New Jersey's clearance rate trajectory after 2011 to a synthetic control constructed from states without arbitration caps. The synthetic difference-in-differences (SDID) estimator (Arkhangelsky et al., 2021) improves upon traditional synthetic control methods, which can be biased when the pre-treatment fit is imperfect (Ferman and Pinto, 2020), by allowing both treated and control units to have different underlying trends rather than requiring perfect parallel trends in levels. SDID constructs a counterfactual for New Jersey by selecting optimal weights over both control states and pre-treatment time periods, creating a synthetic version of New Jersey that closely tracks the actual state's pre-2011 trajectory.

SDID allocates greater weights to states whose clearance rate trajectories most closely

align with New Jersey's throughout the pre-treatment period (1996–2010). This approach achieves better pre-treatment balance compared to conventional techniques while allowing for potential divergence of trends between New Jersey and control states. The treatment effect is measured as the difference between New Jersey's post-2011 outcomes and the trajectory of the weighted synthetic control.

Following Arkhangelsky et al. (2021), I conduct pre-treatment fit diagnostics to assess the quality of potential donor states. I compute two metrics for each donor: (1) the correlation between the donor's pre-treatment clearance rates and New Jersey's pre-treatment clearance rates (1996–2010), and (2) the difference in linear time trends during the pre-treatment period.

Two states exhibited poor pre-treatment fit and were excluded from the donor pool. Hawaii showed a negative pre-treatment correlation with New Jersey's clearance rates and a substantially divergent pre-treatment trend. Kansas similarly displayed a weak correlation and divergent trend. After removing these two states, the final donor pool consists of 48 states spanning 1996–2018. All remaining donors exhibit positive pre-treatment correlations with New Jersey, supporting the parallel trends assumption.

I assess the credibility of the SDID estimates using two complementary placebo tests. The time placebo test evaluates whether the estimated effect is temporally specific to the policy change. I reassign treatment to earlier years within the pre-treatment period and re-estimate the model for each fake treatment date. If the method falsely detects large effects at these placebo dates, this would indicate the divergence between New Jersey and its synthetic control reflects pre-existing differential trends rather than a policy impact. A low time placebo p-value indicates the observed divergence is unusual and specific to the actual treatment timing.

The space placebo test assesses whether New Jersey's estimated effect is unusually large relative to typical cross-state variation. I iteratively treat each of the 48 control states as if it had implemented an arbitration cap in 2011, computing a pseudo-treatment effect for each.

This generates a distribution of placebo effects under the null hypothesis of no treatment. I then compare New Jersey's estimated effect to this distribution. A high p-value suggests the effect is within the range of normal cross-state variation, while a low p-value would indicate New Jersey's divergence is extreme. Standard errors are computed using the space placebo distribution. The standard error equals the standard deviation of the 48 placebo treatment effects. This approach accounts for uncertainty in both the synthetic control weights and potential model misspecification (Abadie, 2021).

### 8.1.2 Results

Table 2 presents the synthetic difference-in-differences estimates comparing New Jersey to 48 control states. The estimated treatment effect is small and negative, a decline of about 1.4 percentage points in clearance rates (the estimate of $-0.0139$ in Table 2 is in proportion units). However, this estimate is not statistically significant. The 95 percent confidence interval is wide, ranging from a decline of about 13.7 percentage points to an increase of about 10.9 percentage points. The pre-treatment period spans 1996–2010 (16 years) and the post-treatment period covers 2011–2018 (8 years).

The placebo tests provide important context for interpreting these results. The time placebo test examines whether I would observe similar divergences if the policy had been implemented at different points in time. The low p-value of 0.0002 indicates that finding a divergence this large at a randomly chosen pre-treatment date would be highly unlikely. This suggests that something meaningful changed around 2011 when the policy was actually implemented, supporting the credibility of the timing.

The space placebo test tells a different story. This test applies the same method to states that never had the policy to see how often similar changes happen by random chance. The high p-value of 0.82 means that about 82 percent of control states showed changes as large as New Jersey's during this same time period. This suggests New Jersey's change is not unusual compared to normal state-to-state differences that have nothing to do with the policy.

These two tests give mixed results. The timing looks right that New Jersey's change happened exactly when the policy started. But the size of the change is not special since it's within the normal range of ups and downs observed across different states. This suggests the policy might have had some effect, but the estimate is not precise enough to distinguish a real policy effect from normal variation.

# 9 Appendix B: The Contract Outcome Analysis

## 9.1 Design Details

*Multiple testing.* For the family of twenty-seven contract outcomes I report Holm step-down adjusted $p$-values (Holm, 1979). The Romano and Wolf (2005) joint-bootstrap step-down, which can recover power by exploiting dependence across outcomes, is a natural extension. Pooling all twenty-seven outcomes into one family is deliberately conservative and costs power, because the step-down threshold tightens with the number of tests. Partitioning the outcomes into substantive families, namely education and tuition, time off and leave, allowances, and salary-guide structure, and correcting within each family raises power while retaining family-wise control at the family level. The two conclusions in Section 5 do not depend on how the family is drawn. Because a within-family correction is weakly less stringent than the pooled one, it cannot overturn the entry-wage compression, which is significant at $p < 0.001$; and because the non-salary estimates lie within a standard error of zero before any correction, no coarser or finer grouping can render them significant.

*Subgroup comparisons.* For the civil-service heterogeneity I re-estimate the Callaway–Sant'Anna ATT separately on each subgroup. The cross-subgroup test for ATT equality assumes independence between subgroup estimates and is computed as

$$z = \frac{\hat{\tau}^{\mathrm{NCS}} - \hat{\tau}^{\mathrm{CS}}}{\sqrt{\widehat{\mathrm{Var}}(\hat{\tau}^{\mathrm{NCS}}) + \widehat{\mathrm{Var}}(\hat{\tau}^{\mathrm{CS}})}},$$

where $\hat{\tau}^{\mathrm{NCS}}$ and $\hat{\tau}^{\mathrm{CS}}$ are the subgroup ATTs. The independence assumption is conservative because both subgroups share the same calendar window and are exposed to the same state-level shocks; a positive shock-driven correlation would shrink the true variance of the difference and make the test more conservative than reported.

*Descriptive composition.* Table 12 reports pre-cap (1996–2010) and post-cap (2011–2024) means by treatment status. The unconditional comparison is dominated by composition

shifts in which agencies code each provision in which years and does not represent the within-agency response that the difference-in-differences design isolates.

## 9.2 Personnel Evidence: Promotion and Workforce Composition

The analyses in this subsection use the 2021–2025 stacked release of the New Jersey Office of the Attorney General Law Enforcement Diversity Data. Because that data does not extend to the pre-cap period, neither analysis can be cast as the staggered difference-in-differences design used in the main text: there is no pre-cap baseline against which to form a within-agency counterfactual. The estimates below are post-cap cross-sectional comparisons of treated and never-treated agencies with year fixed effects,

$$Y_{it} = \alpha + \beta \,\text{Treated}_i + \lambda_t + \varepsilon_{it}, \qquad t \in \{2021, \ldots, 2025\},$$

weighted by the agency's total officer headcount and with standard errors clustered at the agency level. They are descriptive associations, not causal estimates.

*Promotion.* I aggregate the Diversity Data to agency-year totals and compute four outcomes: the overall promotion rate (promoted / eligible for promotion), the Sergeant-promotion rate, the Sergeant share of all promotions, and the Sergeant share of the total workforce (Table 13). Treated agencies show no visible shift toward Sergeant promotion. The Sergeant-promotion rate coefficient is $-0.026$ pooled, the Sergeant share of all promotions $+0.019$, and the Sergeant share of the workforce $+0.005$, all within noise of zero. Treated non-civil-service agencies show a marginally lower overall promotion rate ($-0.205$). Whatever value moved toward supervisory wage levels, it was not accompanied by more promotion into those ranks. Two caveats limit this conclusion. The data begin in 2021, so any workforce adjustment between 2011 and 2020 cannot be observed, and the promotion measures capture annual flow rather than stock.

*Workforce composition.* Table 14 reports the same specification with the share of the

workforce in each demographic cell as the outcome. The coefficient compares levels across agencies within the post-cap window, so the design cannot distinguish a compositional response to the cap from a pre-existing difference between the two groups. The table reports twenty-four distinct tests, and after Holm correction across them no estimate survives at the five percent level. The closest is the share of officers aged 46 or older in civil-service agencies, which is 2.7 percentage points lower in treated agencies (Holm $p = 0.057$). Its direction matches the entry-wage-compression account, but the companion share-young estimate does not corroborate it. The racial-composition differences in non-civil-service agencies do not survive correction, and workforce size shows no detectable difference in any subset. I include the analysis because a compositional response is a natural implication of the substitution account and its absence from the data is worth recording, not because it establishes that response.

Table 11: Outcome Coverage. Number and percentage of non-missing agency-year observations for each outcome in the analysis panel (N = 10,103 agency-years).

| Outcome | N non-missing | Pct non-missing | Type |
|---|---|---|---|
| Tuition stipend, Associate | 4,164 | 41.2% | Dollar value |
| Tuition stipend, BA | 4,152 | 41.1% | Dollar value |
| Tuition stipend, MA | 2,177 | 21.5% | Dollar value |
| Tuition stipend, PhD | 187 | 1.9% | Dollar value |
| Tuition reimbursement value | 1,302 | 12.9% | Dollar value |
| Longevity at 4 yrs | 1,993 | 19.7% | Dollar value |
| Longevity at 8 yrs | 6,985 | 69.1% | Dollar value |
| Longevity at 12 yrs | 7,568 | 74.9% | Dollar value |
| Longevity at 16 yrs | 7,483 | 74.1% | Dollar value |
| Longevity at 20 yrs | 7,061 | 69.9% | Dollar value |
| Longevity at 24+ yrs | 5,811 | 57.5% | Dollar value |
| Uniform allowance | 7,784 | 77.0% | Dollar value |
| Holiday count | 9,079 | 89.9% | Dollar value |
| Sick leave | 9,588 | 94.9% | Dollar value |
| Vacation at 1 yr | 9,655 | 95.6% | Dollar value |
| Vacation at 5 yrs | 9,771 | 96.7% | Dollar value |
| Vacation at 10 yrs | 9,692 | 95.9% | Dollar value |
| Vacation at 15 yrs | 9,620 | 95.2% | Dollar value |
| Vacation at 20 yrs | 9,166 | 90.7% | Dollar value |
| Sergeant wage | 7,052 | 69.8% | Dollar value |
| Lieutenant wage | 4,429 | 43.8% | Dollar value |
| Captain wage | 2,750 | 27.2% | Dollar value |
| Starting salary | 9,881 | 97.8% | Dollar value |
| Has tuition incentive | 10,021 | 99.2% | Presence |
| Has tuition increment | 10,026 | 99.2% | Presence |
| Has tuition reimbursement | 10,013 | 99.1% | Presence |

Table 12: Descriptive Statistics: Pre- vs. Post-Cap Means by Treatment Status. Cells report mean (standard deviation). Treated = agency had a contract expiration during 2010–2017. Pre = 1996–2010; Post = 2011–2024. Real $ deflated to 2010 prices. Winsorized at 1/99 percentile.

| | Pre-cap (1996–2010) | | Post-cap (2011–2024) | |
|---|---|---|---|---|
| Outcome | Treated | Never-treated | Treated | Never-treated |
| Starting salary (real $) | 44,778 (7,601) | 43,539 (8,471) | 41,969 (7,011) | 40,613 (7,884) |
| Tuition stipend, Associate (real $) | 850 (506) | 838 (462) | 687 (408) | 616 (331) |
| Tuition stipend, BA (real $) | 1,377 (703) | 1,507 (801) | 1,126 (581) | 1,010 (564) |
| Tuition stipend, MA (real $) | 1,801 (910) | 1,780 (932) | 1,486 (798) | 1,381 (793) |
| Longevity at 8 yrs (real $) | 255 (560) | 246 (554) | 227 (527) | 303 (668) |
| Longevity at 16 yrs (real $) | 511 (1,127) | 433 (1,082) | 497 (1,122) | 564 (1,099) |
| Uniform allowance (real $) | 949 (299) | 991 (373) | 868 (315) | 746 (354) |
| Holiday count | 15.55 (2.60) | 15.77 (2.49) | 11.43 (2.09) | 10.40 (2.12) |
| Sick leave days | 28.16 (48.50) | 57.59 (99.27) | 20.49 (38.84) | 45.09 (88.06) |
| Has tuition incentive | 0.507 (0.500) | 0.384 (0.487) | 0.511 (0.500) | 0.485 (0.501) |
| Has tuition increment | 0.006 (0.078) | 0.000 (0.000) | 0.001 (0.037) | 0.000 (0.000) |
| Has tuition reimbursement | 0.451 (0.498) | 0.439 (0.497) | 0.453 (0.498) | 0.371 (0.484) |
| Non-salary basket (real $) | 3,431 (1,979) | 3,392 (1,729) | 3,171 (1,850) | 1,709 (553) |
| Basket / starting salary | 0.078 (0.047) | 0.076 (0.049) | 0.076 (0.045) | 0.043 (0.017) |

Table 13: Post-Cap Promotion Outcomes (2021–2025)

| Outcome | Subset | Treated coef. | SE | $p$ | $N$ |
|---|---|---|---|---|---|
| Promotion rate (promoted/eligible) | All | -0.0689 | (0.0945) | 0.466 | 1,461 |
| | Civil-service | +0.0502 | (0.0902) | 0.578 | 620 |
| | Non-civil-service | -0.2054* | (0.1243) | 0.098 | 831 |
| Sergeant-promotion rate (to-Sgt/eligible) | All | -0.0259 | (0.0441) | 0.557 | 1,461 |
| | Civil-service | +0.0250 | (0.0373) | 0.502 | 620 |
| | Non-civil-service | -0.0864 | (0.0655) | 0.187 | 831 |
| Sergeant share of all promotions | All | +0.0186 | (0.0386) | 0.630 | 1,186 |
| | Civil-service | -0.0241 | (0.0451) | 0.594 | 524 |
| | Non-civil-service | +0.0460 | (0.0471) | 0.328 | 654 |
| Sergeant share of workforce (crude) | All | +0.0048 | (0.0033) | 0.145 | 2,214 |
| | Civil-service | +0.0060 | (0.0048) | 0.207 | 847 |
| | Non-civil-service | +0.0023 | (0.0018) | 0.201 | 1,347 |

*Notes:* Weighted OLS with year fixed effects, weighted by agency size (*Number of Officers*). Standard errors clustered at the agency level. Sample is agency-year observations from 2021–2025 in the NJOAG Law Enforcement Diversity Data, restricted to agencies matching the contract panel. *Treated* indicates agencies whose police contract expired during the 2010–2017 cap window. This is a *post-cap equilibrium* comparison, not a difference-in-differences, because the Diversity Data does not extend to pre-cap years. Civil-service and non-civil-service $N$ sum to less than *All* because some agencies are not matched to the civil-service flag. Stars use unadjusted $p$-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 14: Workforce Composition Spillover (2021–2025 Post-Cap Equilibrium)

| Outcome | Subset | Treated coef. | SE | $p$ | Holm $p$ | $N$ |
|---|---|---|---|---|---|---|
| Log workforce size | All | +0.3622 | (0.5641) | 0.521 | 1.000 | 2,214 |
| | Civil-service | +0.7707 | (0.6743) | 0.253 | 1.000 | 847 |
| | Non-civil-service | -0.5329 | (0.4434) | 0.229 | 1.000 | 1,347 |
| Share White | All | +0.0935 | (0.0668) | 0.161 | 1.000 | 2,214 |
| | Civil-service | +0.0935 | (0.1007) | 0.353 | 1.000 | 847 |
| | Non-civil-service | +0.1518 | (0.0713) | 0.033 | 0.733 | 1,347 |
| Share non-White | All | -0.0935 | (0.0668) | 0.161 | 1.000 | 2,214 |
| | Civil-service | -0.0935 | (0.1007) | 0.353 | 1.000 | 847 |
| | Non-civil-service | -0.1518 | (0.0713) | 0.033 | 0.733 | 1,347 |
| Share Black | All | -0.0527 | (0.0401) | 0.188 | 1.000 | 2,214 |
| | Civil-service | -0.0474 | (0.0741) | 0.523 | 1.000 | 847 |
| | Non-civil-service | -0.0743 | (0.0287) | 0.010 | 0.223 | 1,347 |
| Share Hispanic | All | -0.0455 | (0.0522) | 0.383 | 1.000 | 2,214 |
| | Civil-service | -0.0386 | (0.0845) | 0.648 | 1.000 | 847 |
| | Non-civil-service | -0.0942 | (0.0468) | 0.044 | 0.924 | 1,347 |
| Share Asian | All | +0.0007 | (0.0043) | 0.872 | 1.000 | 2,214 |
| | Civil-service | -0.0021 | (0.0077) | 0.785 | 1.000 | 847 |
| | Non-civil-service | +0.0038 | (0.0035) | 0.278 | 1.000 | 1,347 |
| Share Female | All | -0.0133 | (0.0162) | 0.412 | 1.000 | 2,214 |
| | Civil-service | -0.0124 | (0.0270) | 0.645 | 1.000 | 847 |
| | Non-civil-service | -0.0208 | (0.0171) | 0.223 | 1.000 | 1,347 |
| Share Young ($\leq 35$) | All | -0.0324 | (0.0388) | 0.404 | 1.000 | 2,214 |
| | Civil-service | +0.0132 | (0.0175) | 0.449 | 1.000 | 847 |
| | Non-civil-service | -0.0733 | (0.0625) | 0.241 | 1.000 | 1,347 |
| Share Senior ($\geq 46$) | All | +0.0094 | (0.0227) | 0.679 | 1.000 | 2,214 |
| | Civil-service | -0.0266* | (0.0088) | 0.002 | 0.057 | 847 |
| | Non-civil-service | +0.0456 | (0.0354) | 0.198 | 1.000 | 1,347 |

*Notes:* Weighted OLS with year fixed effects, weighted by agency size (*Number of Officers*). Standard errors clustered at the agency level. Sample is agency-year observations from 2021–2025 in the NJOAG Law Enforcement Diversity Data, restricted to agencies matching the contract panel. *Treated* indicates agencies whose police contract expired during the 2010–2017 cap window. This is a post-cap equilibrium comparison, because the Diversity Data does not extend to pre-cap years. Civil-service and non-civil-service $N$ sum to less than *All* because some agencies are not matched to the civil-service flag. *Holm p* applies the Holm step-down correction across the twenty-four distinct tests in this family; the share non-White is the complement of the share White and inherits its adjusted value rather than entering the family separately. Stars reflect Holm-adjusted $p$-values.

# References


Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, *59*(2), 391–425.

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2), 470–492.

Akerlof, G. A., & Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *Quarterly Journal of Economics*, 105(2), 255–283.

Anderson, A., & Krause, L.A. (1987). Interest arbitration: The alternative to the strike. *Fordham Law Review*, *56*, 153.

Anzia, S. F., & Moe, T. M. (2015). Public sector unions and government spending. *Economics & Politics*, 27(1), 1–43.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic Difference-in-Differences. *American Economic Review*, 111(12), 4088–4118.

Ashenfelter, O., & Hyslop, D. (2001). Measuring the effect of arbitration on wage levels: The case of police officers. *Industrial and Labor Relations Review*, 54(2), 316–331.

Bewley, T. F. (1999). *Why wages don't fall during a recession*. Harvard University Press.

Biasi, B. (2021). The labor market for teachers under different pay schemes. *American Economic Journal: Economic Policy*, 13(3), 63–102.

Bollman, K., Gomez, A., Ross, M. B., & Sloan, C. W. (2024). *More with less: The impact of mandatory overtime on police wellness and productivity*. Working paper.

Borusyak, K., Jaravel, X., & Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Breza, E., Kaur, S., & Shamdasani, Y. (2018). The morale effects of pay inequality. *Quarterly Journal of Economics*, 133(2), 611–663.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102(6), 2981–3003.

Clemens, J., & Cutler, D. M. (2014). Who pays for public employee health costs? *Journal of Health Economics*, 38, 65–76.

Crawford, V. P., & Meng, J. (2011). New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review*, 101(5), 1912–1932.

Cunningham, J. P., Feir, D., & Gillezeau, R. (2021). Collective bargaining rights, policing, and civilian deaths. *Journal of Public Economics*, 200, 104502.

Dharmapala, D., McAdams, R. H., & Rappaport, J. (2022). Collective bargaining rights and police misconduct: Evidence from Florida. *The Journal of Law, Economics, and Organization*, 38(1), 1–41.

Farber, H. S. (1980). An analysis of final-offer arbitration. *Journal of Conflict Resolution*, 24(4), 683–705.

Fehr, E., Goette, L., & Zehnder, C. (2009). A behavioral account of the labor market: The role of fairness concerns. *Annual Review of Economics*, 1(1), 355–384.

Ferman, B., & Pinto, C. (2020). Revisiting the synthetic control estimator. *Econometrics Journal*, *23*(3), 413–438.

Feuille, P., & Delaney, J. T. (1986). Collective bargaining, interest arbitration, and police salaries. *Industrial and Labor Relations Review*, 39(2), 228–240.

Frandsen, B. R. (2021). The surprising impacts of unionization: Evidence from matched employer-employee data. *Journal of Labor Economics*, 39(4), 861–894.

Freeman, R. B. (1986). Unionism comes to the public sector. *Journal of Economic Literature*, 24(1), 41–86.

Freeman, R. B., & Medoff, J. L. (1984). *What do unions do?* Basic Books.

Gittleman, M., & Pierce, B. (2012). Compensation for state and local government workers. *Journal of Economic Perspectives*, 26(1), 217–242.

Gneezy, U., & List, J. A. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica*, 74(5), 1365–1384.

Goncalves, F. (2021). *Do Police Unions Increase Misconduct?* Working paper, University of California, Los Angeles.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.

Gruber, J. (1994). The incidence of mandated maternity benefits. *American Economic Review*, 84(3), 622–641.

Hirsch, B. T. (2004). What do unions do for economic performance? *Journal of Labor Research*, 25(3), 415–455.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.

Huq, A. Z., & McAdams, R. H. (2016). Litigating the blue wall of silence: How to challenge the police privilege to delay investigation. *University of Chicago Legal Forum*, *2016*, 213.

Kaplan, J. (2021). Jacob Kaplan's Concatenated Files: Uniform Crime Reporting Program Data: Offenses Known and Clearances by Arrest (Return A), 1960-2020. Ann

Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E100707V17

Kaur, S. (2019). Nominal wage rigidity in village labor markets. *American Economic Review*, 109(10), 3585–3616.

Koszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121(4), 1133–1165.

Kube, S., Maréchal, M. A., & Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. *American Economic Review*, 102(4), 1644–1662.

Mas, A. (2006). Pay, reference points, and police performance. *The Quarterly Journal of Economics*, 121(3), 783–821.

Mas, A. (2008). Labour unrest and the quality of production: Evidence from the construction equipment resale market. *Review of Economic Studies*, 75(1), 229–258.

New Jersey Office of the Attorney General. (2024). *Law Enforcement Diversity Data: documentation and methodology notes*. New Jersey Office of the Attorney General public dashboard. https://www.njoag.gov/policerecruiting/ (accessed June 15, 2026).

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5), 2555–2591.

Romano, J. P., & Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469), 94–108.

Rosen, S. (1986). The theory of equalizing differences. In O. Ashenfelter & R. Layard (Eds.), *Handbook of Labor Economics* (Vol. 1, pp. 641–692). Elsevier.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2), 2218–2244.

Rushin, S. (2016). Police union contracts. *Duke Law Journal*, 66, 1191–1266.

Shapiro, C., & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review*, 74(3), 433–444.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199.

Valletta, R. G. (1993). Union effects on municipal employment and wages: A longitudinal approach. *Journal of Labor Economics*, 11(3), 545–574.

Woodbury, S. A. (1983). Substitution between wage and nonwage benefits. *American Economic Review*, 73(1), 166–182.

Zax, J. S., & Ichniowski, C. (1988). The effects of public sector unionism on pay, employment, department budgets, and municipal expenditures. In R. B. Freeman & C. Ichniowski (Eds.), *When public sector workers unionize* (pp. 323–364). University of Chicago Press.