Civil Service Reforms: Evidence from U.S. Police Departments

Arianna Ornaghi*

May 19, 2018

Abstract

Does reducing politicians' control over public employees' hiring and firing improve bureaucratic performance? I answer this question exploiting population-based mandates for U.S. municipal police department merit systems in a regression discontinuity design. Merit system mandates improve performance: the property crime rate is lower and the violent crime clearance rate is higher in departments operating under a merit system than in departments under a spoils system. Changes in resources or police officers' characteristics do not drive the effect, but I provide indirect evidence that the limitations to politicians' ability to influence police officers through discretionary firings are instead important.

JEL codes: D73, M51

*I am extremely grateful to Daron Acemoglu, Claudia Goldin, and Ben Olken for their invaluable advice and guidance throughout this project. I also thank Enrico Cantoni, Daniel Fetter, John Firth, Ludovica Gazze, Daniel Gross, Sara Heller, Nick Hagerty, Greg Howard, Peter Hull, Donghee Jo, Gabriel Kreindler, Matt Lowe, Rachael Meager, Manisha Padi, Bryan Perry, Otis Reid, Frank Schilbach, Mahvish Shaukat, Cory Smith, Marco Tabellini and participants in the MIT Political Economy lunch and Harvard Economic History lunch for their comments and suggestions. This research was conducted while the author was Special Sworn Status researcher of the U.S. Census Bureau at the Center for Economic Studies. Research results and conclusions expressed are those of the author and do not necessarily reflect the views of the Census Bureau. This paper has been screened to insure that no confidential data are revealed. Correspondence: Arianna Ornaghi, Department of Economics, University of Warwick, Social Sciences Building, Coventry CV4 7AL, UK. Email: A.Ornaghi@warwick.ac.uk.

1 Introduction

Bureaucracies are a key component of state capacity. As policy implementers, they translate policy choice into outcomes and affect a state's ability to provide public goods. We know both from direct experiments (e.g. Chong et al., 2014) and expert surveys (e.g. La Porta et al., 1999; Hyden, Court, and Mease, 2003; Kaufmann, Kraay, and Zoido, 1999) that there is a high degree of cross-country variation in bureaucratic performance. Why are some bureaucracies effective while others fail? According to a long tradition in the social sciences, the first order answer to this question is whether or not politicians control the hiring and firing of public employees. There is no consensus, however, on the effect that politicians' control has on performance.

Historically, the entire American public administration was characterized by a spoils system in which politicians were free to hire and fire bureaucrats as they saw fit. In 1829, President Andrew Jackson justified the system on grounds of increased responsiveness: "More is lost by the long continuance of men in office than is generally to be gained by their experience" (as quoted in White, 1954, p. 347). By the end of the 19th century, however, the opposite view – that merit systems insulating bureaucrats from politics were necessary to give public employees long term incentives and foster expertise – had become more prominent. Reforms professionalizing the bureaucracy were first introduced at the federal level in the 1880s and soon started diffusing at lower levels of government. Nevertheless, the debate on whether politicians' control improved performance was by no means closed. When the Supreme Court was called upon in the late 1970s to discuss whether dismissals for political reasons violated the First Amendment, the decision was in support of merit systems, but the dissenting opinion of Justice Stewart once again endorsed spoils systems: "Patronage serves the public interest by facilitating the implementing of policies endorsed by the electorate."

Whether merit systems improve performance depends on the trade-off between expertise and responsiveness, and it is ultimately an empirical question. Evaluating the trade-off, however, has proven to be difficult. When bureaucratic organizations are defined at the country level, their effect is confounded by other country-specific factors. When within-country variation exists, endogenous adoption complicates the identification of causal effects. In addition, finding direct measures of bureaucratic performance can be challenging. The principal contribution of this paper is to provide well-identified causal evidence of the effect of bureaucracy professionalization on a credible set of performance measures.

The setting is that of municipal police departments in the United States. In particular, I contrast the performance of police departments operating under a spoils system with that of departments in which a merit system was exogenously introduced. Under a spoils system, politicians were free to hire and fire police officers as they saw fit. Under a merit system, the authority to appoint, promote and dismiss officers was taken from the mayor and given to a semi-independent civil service commission. Hiring and promotion decisions had to follow merit-based criteria and dismissals were only permitted for just cause.

The first cities to establish merit systems, Albany, Utica and Yonkers (NY), did so in 1884, just a year after the Pendleton Act introduced meritocratic hiring for part of the federal bureaucracy. However, it took a long time for the reform to diffuse at the local level, especially as far as smaller municipalities were concerned. As late as in the mid-1970s, only 20% of police departments in cities with fewer than 10,000 inhabitants had a merit system to hire their police officers.[1,2]

There is a high degree of variation in how merit systems were introduced at the local level. This paper focuses on states with population-based mandates for police department merit systems. The mandates operated in the following way. When the state legislation was first passed, all municipalities with population above the threshold in the latest available census were mandated to introduce a merit system. At the following censuses, previously untreated municipalities that had grown above the lower limit also became subject to the mandate and were required to introduce a merit system for their police department. Municipalities below the threshold were allowed to introduce a merit system at any time.

[1]Merit systems covered all employees in the largest cities but were restricted to members of police and fire departments in the vast majority of municipalities.

[2]Author's calculations based on data from Ostrom, Parks, and Whitaker (1977).

Whenever a population census was taken, treatment was assigned to all previously untreated municipalities above the cutoff. As a result, each census defines a separate experiment in which the effect of the mandate can be estimated using a standard cross-sectional RD design comparing municipalities just above and just below the threshold. For the causal effect of the mandate to be identified, municipalities just above and just below the threshold must be comparable. I validate the assumption by showing that the density of the running variable is smooth at the discontinuity and that municipality characteristics are balanced at baseline.

My main objective is to study how the introduction of merit systems affected the performance of police departments. I proxy for police performance using crime rates (crimes per 100,000 people) and clearance rates (crimes cleared by arrest over total crimes). The data are from the Uniform Crime Reports (UCRs) published by the Federal Bureau of Investigation. UCRs are available at the individual department level only starting from 1960. At the end of the 1970s, two U.S. Supreme Court decisions extended protections from political dismissals to all public employees regardless of municipality size, substantially changing what it meant to be under a merit system as opposed to a spoils system. The main analysis focuses on the 1960 to 1980 period, and exploits variation in treatment status from the 1970 census experiment.

My evidence indicates that merit system mandates improved police performance. In the first ten years after a municipality became subject to the mandate, the property crime rate was 46% lower and the violent crime clearance rate was 12% higher in municipalities just above the threshold relative to municipalities just below. The results are not explained by pre-existing differences: there is no discontinuity in the outcomes before the introduction of merit systems. Studying the effect as a function of years since treatment shows that it took two to three years for merit systems to first affect the property crime rate, but that after the first adjustment period, the effect was constant.

I test whether the results depend on the choice of sample, specification and estimation technique. The effect of merit systems on the property crime rate is not driven by any of the choices made in the estimation. The effect of merit systems on the violent crime clearance rate, however, is less robust. In addition, I argue that it is improbable that the results are driven by other state-specific policies changing at the same threshold. Finally, I discuss in detail why the results are unlikely to be an artifact of differential reporting.

The results discussed thus far show the effect of the mandate, as no data on adoption of full-fledged merit systems exist for the 1970s. Using pre-1940 data, I show that merit system mandates significantly increased the probability that a municipality had a full-fledged merit system. The effect is smaller than one, not only because municipalities below the cutoff could introduce a board, but also because municipalities above the cutoff could face delays. However, the protections granted by the mandate were enforceable in court from the moment in which the official census counts were published, which means that despite the fact that compliance was imperfect, a partial treatment was in place even before the creation of a civil service commission.

Having established that merit systems have a positive effect on performance, I turn to the question of what explains this effect. I explore three possible channels: increases in the resources available to police departments, changes in police officers' characteristics, and reduced political influence through protections against discretionary dismissals. First, I show that merit systems did not influence the resources available to police departments. There is no discontinuity at the threshold in expenditures or employment, which suggests that departments operating under a merit system used similar inputs as departments operating under a spoils system.

Second, I find scant evidence that merit systems selected and retained officers with different characteristics. I study the demographic composition of the departments using a novel dataset with individual-level information on police officers that I constructed from the full count microdata from the population censuses 1960 to 1980. I show that there were no differences at the threshold in the age, educational attainment or veteran status of police officers, suggesting that improved performance is unlikely to be explained by merit system departments having "better" police officers.

Given that the effect on performance cannot be explained by merit systems increasing resources or attracting police officers with different characteristics, the last channel, increased protection from political influence, is likely to be important. I provide indirect evidence of this by exploiting the fact that at the end of the 1970s, two Supreme Court decisions extended protections from dismissals for political reasons to all non-policymaking municipal employees, independent of whether they were part of a merit system. Municipalities treated after 1980 still had to create independent civil service commissions, but there was no discontinuity in whether employees were protected from political dismissals. I find that merit system mandates had no effect on crime or clearance rates after 1980, consistent with the hypothesis that the protections from discretionary firings that limited political control over police officers were important to explain the result.

The finding that merit systems have a positive effect on performance is in line with previous evidence that professionalized bureaucracies tend to be more effective, such as the cross-country comparisons in Evans and Rauch (1999) and Rauch and Evans (2000). The closest existing work is Rauch (1995), who found using a differences-in-differences design that the introduction of U.S. municipal merit systems increased infrastructure investment and city growth rates before 1940. With respect to this study, I make two contributions. First, given that the timing of municipal reforms is likely to be endogenous to local conditions, the regression discontinuity design is important to claim causality.[3] Second, studying police departments has the advantage of providing direct measures of performance, as opposed to further downstream outcomes that are only indirectly affected by the actions of the bureaucracy. The paper also provides complementary evidence to the growing number of papers studying the performance effects of specific features of bureaucratic organizations (e.g. Ashraf, Bandiera, and Lee, 2016; Iyer and Mani, 2012; Rasul and Rogger, 2016; Xu, 2017). I contribute to this literature by showing how these features, that potentially introduce different trade-offs, interact in determining performance when they operate together, as is typical in modern bureaucracies. In addition, the paper adds to existing work on the effect of U.S. federal and state merit systems on political outcomes (e.g. Folke, Hirano, and Snyder,

[3]For example, cities might introduce merit system in response to particularly bad crime spells, which would bias towards finding positive effects that are instead explained by mean reversion.

2011; Johnson and Libecap, 1994; Ujhelyi, 2014). Finally, the paper relates to studies looking at determinants of police performance by providing evidence of the role played by police organization (e.g. Chalfin and McCrary, Forthcoming; Evans and Owens, 2007; Levitt, 1997; Mas, 2006).

The remainder of the paper is organized as follows. Section 2 presents the background, section 3 presents the data, and section 4 discusses the empirical strategy. The main results are presented in section 5 and potential mechanisms are presented in section 6. Section 7 concludes. Additional tables and details are available in a separate online appendix.[4]

2 Background

Historical background

The Wickersham Commission reports, published in 1931, offer a dismal picture of the state of American policing at the beginning of the 20th century.[5] Police departments across the nation were described as tainted by corruption and incapable of controlling crime. The main culprit was identified to be excessive political influence in policing, which made the tenure of executive chiefs and officers alike too short and the selection of personnel with adequate qualifications impossible. In the words of J. Edgar Hoover (1938): "the real "Public Enemy Number One" against law and order is corrupt politics." To overcome these issues, the solution proposed was police professionalization through the establishment of effective merit systems.

The police was just one of the many public organizations under political control. In fact, starting from the Jackson Presidency, the entire American bureaucracy was under a full-fledged spoils system, where newly elected presidents would substitute office holders nominated in previous administrations for party loyals (Freedman, 1994). At the height of the spoils system, wholesale replacement of federal employees was the norm (United States Civil Service Commission, 1973),

[4]The online appendix is available at the following link.

[5]The National Commission on Law Observance and Enforcement, also known as the Wickersham Commission, was created by President Hoover in 1929 with the objective of studying the state of crime and policing and identifying possible solutions.

with replacement rates as high as 50% even for postmasters in charge of smaller offices (Fowler, 1943).

By the mid 19th century, however, the discussion of whether the spoils system was the best way to organize the bureaucracy had begun. The proponents of professionalization saw it as a response to widespread inefficiencies; those opposing reform were afraid of losing not only political power, but also the support of an aligned bureaucracy. The first civil service reform aimed at professionalizing public employees, the Pendleton Act, was adopted in 1883. The act created a bipartisan Civil Service Commission under the control of the President and introduced meritocratic hiring for around 10% of federal employees. Protection from partisan dismissals was established by the end of the 1890s (Lewis, 2010). Expansion was swift: by 1920, 80% of federal employees were covered by a merit system. Contemporaneous testimonies of postmasters and custom collectors report improvements in the functioning of their agencies following the reform (U.S. Civil Service Commission, 1884), and the consensus is that there was a positive effect on performance (Johnson and Libecap, 1994 and Carpenter, 2005).

Albany, Utica and Yonkers (NY) were the first cities to adopt a merit system in 1884. Adoption picked up again during the Progressive Era, when reformers identified professionalization as the remedy for the inefficiency of city hall. The diffusion of the reform, however, was slower than at the federal level, and by 1920, fewer than 40% of cities with more than 25,000 inhabitants had a merit system.

Police departments were one of the principal agencies involved in municipal merit systems, especially in many smaller cities and towns where merit systems were restricted to police and fire departments. Originally an offshoot of the Progressive movement (Fogelson, 1977, p. 44), the professionalization of the force was at the center of police reform long after the original impetus had subsided. In 1954, O. W. Wilson was still supporting the ideal: "sound personnel management operates on the merit principle that to the best-qualified goes the job - not to the victor belong the spoils."

Merit system mandates

There was wide variation in the legislative basis of municipal merit systems. In the majority of the cases, the reform was adopted independently by municipalities through ordinance or referendum. This makes studying the effect of merit systems challenging: because introducing the reform was a political decision taken by those who had to gain (or lose) from it, the timing was likely endogenous. In some cases, however, merit systems were introduced by higher levels of government: this paper focuses on states in which the legislature mandated merit systems for police departments of municipalities above certain population thresholds.

I collected information on state legislation related to police merit systems from primary and secondary sources (see Online Appendix B for details). As Figure I shows, there are eight states with mandates based on population thresholds.[6] While there were differences in the details of the legislation across states, the fundamental features of the reform were the same. When a merit system was introduced in a police department, the authority over hiring, promotions and dismissals was removed from the mayor and given to a semi-independent civil service commission. Hiring and promotion decisions, not regulated under a spoils system, had to be based on merit following competitive examinations. Police officers, who could be dismissed by the mayor at will under a spoils system, could only be fired for just cause and had access to a formal grievance procedure administered by the commission.[7]

When a merit system was introduced, already employed officers were grandfathered in. Moreover, the provisions covered all police officers of lower ranks, but were sometimes extended to the police chief.[8] Finally, civil service commissions

[6]Because Wisconsin had two different cutoffs based on whether a municipality was incorporated as a village or as a city, I consider Wisconsin villages and Wisconsin cities separately. When the legislation excludes municipalities under specific forms of government (for example, municipalities under a city manager form of government before 1933 in Wisconsin), I omit them from the analysis.

[7]Police unions may also make it hard for an administration to fire police officers. To the extent that there is to my knowledge no reason why the probability of being unionized should not be smooth across the discontinuity, this should not impact my results.

[8]In Arizona, Louisiana and West Virginia, the police chief was not under a merit system. In Illinois, the commission nominated the chief by default, but the provision could be changed by ordinance. In Iowa, the chief did not receive protections but could be nominated only from an eligibility list. Whereas this is a potentially interesting dimension of heterogeneity, my sample is

were usually nominated by the mayor or by the governing body of the city, but overlapping terms and requirements on members' political affiliations decreased the risk of capture.[9]

The years of introduction of the reform at the state level range from 1907 to 1969. When the state legislation was first passed, all municipalities above the population threshold according to the latest available census had to introduce a merit system for their police department. In all subsequent censuses, municipalities that had grown above the cutoff also became subject to the mandate and had to introduce a merit system. In approximately half of the states, the mandate was explicitly based on the federal population census, whereas in the remaining ones any official municipal, state or federal census could also be used. Only a few states had penalties for non-compliance, but the protections given to police officers became binding the moment that the official counts from the census were released, and could be challenged in court. Finally, municipalities below the threshold were allowed to introduce a merit system through ordinance or referendum at any time. At the end of the 1970s, two U.S. Supreme Court decisions, Elrod v. Burns (1976) and Branti v. Finkel (1980), made dismissals for political reasons illegal for all non-policymaking municipal employees on grounds of violation of the First Amendment, substantially limiting political influence even in municipalities not under a merit system.[10]

The thresholds are between 4,000 and 15,000: the legislation focused on police departments in small municipalities. Small town police departments (e.g. departments in municipalities below 10,000 people) employed around one civilian and six full-time sworn officers, four of whom had grade of patrolman, highlighting a limited role for career incentives. They engaged in patrolling, traffic control and early criminal investigations, but relied on external support for more complex tasks.[11]

---

too small to push the analysis in this direction.

[9] In five out of nine cases (Arizona, Illinois, West Virginia, Wisconsin cities and Wisconsin villages), the commission was bipartisan, and in two additional states (Iowa and Louisiana), members were required to be non-political. In Montana and Nebraska, members were only required to be citizens of good standing supporting the merit system principle for public administration.

[10] Elrod v. Burns, 1976, 427 U.S. 347. Branti v. Finkel, 1980, 445 U.S. 518.

[11] Author's calculations based on a 1974 survey conducted by Elinor Ostrom (Ostrom, Parks, and Whitaker, 1977). The survey provides information on all police departments in a random sample of standard metropolitan areas.

3 Data

Crime. The crime data are from the Uniform Crime Reports (UCRs) published by the Federal Bureau of Investigation. UCRs are compiled from returns voluntarily submitted to the FBI by police departments, and are available for individual agencies starting from 1960. They report monthly counts of offenses known to the police and cleared by arrest for seven crimes (burglary, larceny-theft, motor vehicle theft, murder and negligent manslaughter, rape, robbery, and assault).[12]

I use UCRs to define two sets of outcomes. First, I look at monthly property (burglary, larceny and vehicle theft) and violent crime (robbery, assault, rape, and murder) rates, defined as crimes per 100,000 people.[13] Second, I look at monthly property and violent crime clearance rates, defined as the number of crimes cleared by arrest over total crimes.[14] Appendix Table II presents the descriptive statistics.

Reform adoption. I predict the year in which a municipality became subject to the mandate using population counts digitized from the official publications of the Census Bureau and information on state merit system laws. No information on actual adoption is available for the main period of interest, but I use three surveys conducted by the Civil Service Assembly of the United States in 1937, 1940 and 1943 to provide evidence on reform adoption for an earlier period.[15]

[12] I clean the data for missing values following the indications reported by Maltz (2006), but I do not use his data imputation procedure. I show that the results are robust to additional data cleaning aimed at identifying outliers in the robustness checks. Online Appendix C reports more details.

[13] For intercensal years, I linearly interpolate municipal population from the official publications of the Census Bureau. I prefer this to using municipal population reported in UCRs themselves as visual inspection of the data suggests that the variable presents a high degree measurement error, but I show that this does not impact the main results as a robustness check.

[14] The FBI website states: "for a crime to be cleared by arrest it must be the case that at least one person has been: (1) arrested; (2) charged with the commission of the offense; (3) turned over to the court for prosecution." There is no perfect correspondence between the crimes that are reported as being cleared in a certain month and the offenses taking place in that month. I ignore the issue when defining the outcome as I find a large effect on crimes, which suggests that in order to use clearance rates to proxy for performance, normalizing by volume is important. In addition, to avoid results being driven by outlier months in which the number of crimes cleared by arrest is higher than the number of crimes and support the interpretation of the outcome as fraction of crimes cleared by arrest, I windsorize the outcome at 1. Clearance rates have been defined in this way and used as proxy for performance in the economics of crime literature, for example in McCrary (2007).

[15] Previous studies using these data include Tolbert and Zucker (1983) and Rauch (1995).

Expenditures and employment. Data on expenditures and employment for police departments are from the Annual Survey of State and Local Government Finances and the Census of Governments published by the Census Bureau.[16] I study total expenditures per 1,000 people and total employment per 1,000 people.

Police officer characteristics. I construct a dataset of police departments' demographic characteristics starting from the restricted access full count microdata of the 1960 to 1980 Decennial Censuses.[17] I identify police officers using reported occupation, industry and class of worker, and I assign each police officer to the department of the municipality in which they were enumerated.[18,19]

4 Empirical strategy

The empirical strategy to identify the impact of merit systems exploits population-based mandates in a regression discontinuity design. The key feature of the setting is that each population census defines an experiment in which treatment is assigned to all previously untreated ("at risk") municipalities. The causal effect of the reform can be estimated using the following specification:

$$y_{mt} = \beta \mathbb{1}(dist_m \geqslant 0) + f(dist_m) + \delta_{st} + \varepsilon_{mt} \; for \; m \in RS \tag{1}$$

y_{mt} is outcome y for municipality m and month (or year) t; $dist_m$ is the population

[16]The data on expenditures are available at the municipality level starting from 1970, and the data on employment from 1972. Both datasets cover the universe of municipalities in 1972 and 1977 (from the Census of Government) and a sample of local governments in all other years (from the Annual Survey of State and Local Government Finances).

[17]The microdata are available for every individual who participated in the census, but starting in 1960 work-related questions were only asked in long form schedules, which means that I am effectively using a sample covering 15% to 25% of the U.S. population depending on the year.

[18]Using information on place of work to match police officers to departments is unfeasible because of the coding of the data. For the individuals for which I can identify place of work municipality, I can check whether the assumption is correct. I find that 73% of these police officers work for the department of the municipality in which they reside.

[19]I validate the procedure comparing the number of police officers in the census with the number I should expect to find given the long form sampling frame and the number of police officers reported for each department in the Census of Government. The procedure appears to work quite well. In 1970, for 84% of departments the discrepancy is lower than two and for 59% it is lower than one. The error rates are 91% and 63% in 1980.

distance to the threshold (i.e. the running variable); $\mathbb{1}(dist_m \geqslant 0)$ is an indicator for being above the threshold; $f(dist_m)$ are a set of flexible functions of the running variable; δ_{st} are state-month (or year) fixed effects; and RS is the set of "at risk" municipalities, i.e. all municipalities in the last census before the introduction of the state legislation and previously untreated municipalities in each census experiment thereafter. β estimates the effect of having a mandated merit system and is the coefficient of interest. The fixed effects are not needed for identification but increase precision. Standard errors are clustered at the municipality level to correct for the correlation induced by including the same municipality multiple times in the estimation.

I estimate the results using locally linear regression (Gelman and Imbens, 2016) and a uniform kernel, which is equivalent to estimating a linear regression on observations within the bandwidth separately on both sides on the discontinuity. I show results for three fixed bandwidths (750, 1,000, 1,250) and for an outcome- and sample-specific MSE-optimal bandwidth calculated using the procedure suggested by Calonico, Cattaneo, and Titiunik (2014). The optimal bandwidth is calculated separately for each outcome and sample after partialling out the fixed effects and allowing for clustering of the standard errors following Bartalotti and Brummet (2016).

The main effect is estimated pooling all post-treatment observations. The post-treatment period starts either in the year of introduction of the mandate at the state level or, for all the following census experiments, in the year of the population census itself.[20] It ends in the year of the following census.[21] As a falsification test, I check that there are no pre-existing discontinuities in the outcomes by estimating the same specification on pre-treatment observations.

[20]Preliminary counts for the population census were published between May and October, which makes the year when the census is taken a transition year. In the baseline estimation, I consider it a post-treatment year.

[21]I focus on the short-term effect of the mandate because the long-term effect would be confounded by the control municipalities growing above the threshold and being treated in following census experiments. I could estimate longer-term effects by comparing outcomes for places that were just above and just below the threshold in a certain census and below the threshold in the following one. However, given that most cities experience population growth, I do not have enough data to estimate such treatment effects.

The baseline specification estimates an average treatment effect of the pre- and post-treatment period, but we might be interested in understanding how the effect of the mandate changes over time. To do so, I estimate the following RD event study specification:

$$
\begin{aligned}
y_{mt} = \sum_{\sigma \in \{-5,+10\}} \beta_\sigma \mathbb{1}(dist_m \geqslant 0)\mathbb{1}(t - \tilde{c} = \sigma) + f_t(dist_m) + \delta_{st} \\
+ \varepsilon_{mt} \; for \; m \in RS
\end{aligned}
\tag{2}
$$

y_{mt} is outcome y for municipality m and month (or year) t; $dist_m$ is the population distance to the threshold (i.e. the running variable); $\mathbb{1}(dist_m \geqslant 0)$ is an indicator for being above the threshold; $\mathbb{1}(t - \tilde{c} = \sigma)$ is an indicator equal to 1 if σ years have elapsed since treatment ($\tilde{c}$ is treatment year for census experiment c); $f_t(dist_m)$ is a set of year specific flexible functions of the running variable; δ_{st} are state and month (or year) fixed effects; and RS is the set of "at risk" municipalities. β_σ estimates the effect of having a mandated merit system for σ years and is equivalent to the RD estimate from a cross-sectional RD that pools all observations measured σ years since treatment. The specification is estimated pooling both pre- and post-treatment observations. Standard errors are clustered at the municipality level.

The identification assumption is that all factors other than treatment vary continuously at the threshold. First, municipalities must not sort around the cutoff according to their characteristics. I validate the design by testing for discontinuities in the density of the running variable and in baseline covariates. Appendix Figure I presents McCrary (2008) tests for all census experiments in which treatment is assigned (1910 to 2000). Out of the ten census experiments, the McCrary test only barely fails for 1980, in line with statistical error. Most importantly, the McCrary test shows no discontinuity in the density of the running variable for the 1970 census experiment, which is the one used for the main analysis. Appendix Table III shows the results of a covariate balance test. The table reports the coefficient on the dummy for being above the threshold for three fixed bandwidths (750, 1,000, 1,250) and an outcome-specific MSE-optimal bandwidth. The outcomes are municipality characteristics measured in the population census in which treatment was assigned.

None of the coefficients for the 1970 census experiment is statistically different than zero: the places just below the threshold are a good control group for those just above. Reassuringly, even if the McCrary test fails for 1980, Appendix Table III shows covariate balance for the same census experiment.

Second, to estimate the causal effect of merit systems, it must also be the case that no other policies change at the same threshold, a particularly common issue for RD designs based on population cutoffs (Eggers et al., Forthcoming). I provide evidence that it is unlikely that other policies are driving my results in the robustness check section.

5 Results

Effects on performance

I study the effect of police professionalization on performance by estimating the impact of merit system mandates on crime and clearance rates. The analysis uses outcome data for the 1960 to 1980 period: crime data are available at the department level starting from 1960, and Supreme Court decisions extending protections from dismissals for political reasons to all municipal workers altered the content of the reform at the end of the 1970s. Variation in treatment status is from the 1970 census experiment.[22]

As there exists no data on merit system adoption for the 1970s, the analysis effectively estimates the effect of merit system mandates. If compliance to the mandate was imperfect, this means that I estimate intention to treat effects, where treatment is defined as adoption of a full-fledged merit system. However, it is important to note that protections against hiring and dismissals for political reasons could be challenged in court from the moment in which the mandates became effective.

[22]The 1970 census experiment is the only one for which outcome data are available for both the pre- and post-treatment period. The 1960 census experiment has outcome data for the post-treatment period only. As shown in Appendix Table IV, police departments in municipalities just above the threshold were more likely to submit data to the FBI in 1960. This is a potentially interesting outcome as it suggests that police departments under a merit system had better record keeping practices. However, it makes it impossible to interpret the results on crime rates, which is why I exclude the 1960 census experiment from the analysis.

As a result, a partial treatment effect was in place even without the institution of a full-fledged merit system, which makes estimating the effect of the mandate itself meaningful on its own.

I begin by showing descriptively how the total crime rate, defined as crimes per 100,000 people, changed over the period from 1965 to 1979 for municipalities that were under a merit system mandate and for municipalities that were not. Figure II shows the mean monthly crime rate by year separately for places above and below the threshold, together with 95% confidence intervals. Over the period of interest, places both above and below the threshold experienced a stark increase in crime rates, but while places below the threshold kept growing at a steady pace throughout the period, departments that fell under the merit system mandates saw crime rates increasing more slowly after 1970.

Figure III panel A shows the visual equivalent of the RD estimates separately for property and violent crime rates. I analyze property and violent crimes separately because they are likely to have different determinants, and thus be differentially affected by police actions.[23] The panels on the left show the falsification RD graphs estimated on the sample of pre-treatment years (1960 to 1969), while the ones on the right show the main RD graphs of interest, estimated on post-treatment years (1970 to 1979).[24] Outcomes are defined as log crime rates to make the coefficients comparable across experiments, especially given the large increase in crime rates over the period.[25] The dots show the average value of the outcome for different bins of the running variable. The line plots the fit from a locally linear regression estimated separately on each side of the discontinuity. Since the mean of the outcome may be different across experiments, I partial out state-month fixed effects.

[23] Given that the majority of crimes are against property, the effects on total crime tend to mirror the effects on property crimes.

[24] More precisely, the pre-treatment period for the violent crime rate is 1964-1969, as simple assault was not reported 1960-1963. The increase in sample size between the pre- and post-treatment period is explained by more agencies reporting data to the FBI. I do not restrict the analysis to a balanced sample because the estimation is based on within-month comparisons of places above and below the threshold, and I want to maximize all available data, but I show that restricting the estimation to a quasi-balanced sample does not make a difference in the robustness check section.

[25] This drops observations with 0 crimes. As shown in Appendix Table V, this does not make a difference: using crime rates expressed in levels, crime counts and log of crime counts gives the same results.

The RD graphs show that there was no difference in the property crime rate at the discontinuity in the pre-treatment sample. However, after the mandate became effective, municipalities just above the threshold had a lower property crime rate than those just below.[26] The regression estimates confirm the results. Table I shows the effect of having a mandated merit system for three fixed bandwidths and for a MSE-optimal bandwidth separately for the pre-treatment sample (columns 1 to 4) and for the post-treatment sample (columns 5 to 8). There was no difference in the property crime rate in the pre-period, but municipalities above the threshold had a lower property crime rate in the post-period with respect to those below. The coefficients are statistically significant at the 5% level, and the results are robust to different bandwidths. The magnitude of the effect is large: looking at the estimates for places within a 1,000 bandwidth from the threshold, the coefficient shows a 46% reduction in the property crime rate for treated places in the first ten years after the reform was introduced. This is equivalent to 4.6 fewer property crimes per month for a municipality of 5,000 inhabitants. Crime rates are noisy and standard errors are large: the 95% confidence interval is always negative but contains effects of very different magnitudes.

Both the RD graphs and the regression estimates show that merit systems had no effect on violent crime rates: there is no discontinuity at the threshold, and the coefficient for being subject to a merit system mandate is never significantly different than zero. It appears that merit systems affected police departments along dimensions that made them more effective at reducing property, but not violent, crimes. In addition, it is important to note that violent crimes are rare events, as evidenced by the large standard deviation, and I may not have enough power to detect an effect. However, given that violent crimes are generally considered to be more likely to be reported to the police, the effect may be seen as a red flag for differential reporting at the threshold, a possibility that I discuss in detail, and

[26]Given that I am partialling out state-month fixed effects and crime rates are significantly increasing over time, it is not possible to compare levels across the RD graph of the pre- and post-treatment period. The change over time in the outcome is more correctly inferred from Figure II (that looks very similar if restricted to the property crime rate): both municipalities above and below the threshold see higher property crime rates over the period, but the increase is slower for places above the threshold.

discard, below.

We may also be interested in understanding how the effect of the mandate changed over time. To do this, I estimate the event study specification (equation (2)) and show the β_σ coefficients together with 95% confidence intervals in Figure III panel B.[27] The graph for the property crime rate shows that the effect is gradual over time and is statistically significant starting five years after treatment is assigned in 1970. None of the coefficients in the pre-period is statistically significant, but the point estimates start being negative two to three years before treatment. This is potentially concerning, as it may point to pre-existing differences in crime rates before merit system mandates. However, since I am estimating intention to treat effects, a difference in the outcomes driven by early treatment, a "true" anticipation effect, would not invalidate the design.

I provide evidence that this is indeed the case by estimating the event study separately for states in which a "true" anticipation effect is more or less likely to appear. In particular, I exploit the fact that in four states in my sample (Illinois, Montana, Nebraska and West Virginia), the mandate was based on population measured in any official municipal, state or federal census. In these states, it is likely that the mandate became effective before the federal census was released, as the actual population of a municipality grew above the threshold and an official census was taken. On the contrary, there should be no anticipation where the mandate was explicitly based on the federal population census only.[28] Reassuringly, Appendix Figure II shows that there was an anticipation effect only in states where the mandate was based on any official census. When I focus on states where the mandate was strictly based on the federal census only, there is no difference in crime rates until 1972 (if anything, the coefficients are positive, although not significantly different than zero). The decline is gradual at first, but remains constant in magnitude in the following years. Whereas none of the coefficients in this event study is statistically

[27] Different from differences-in-differences event study specifications, there is no omitted category because the model never gets fully saturated and the omitted category is constituted by control municipalities in each experiment.

[28] This assumes that municipalities only adopted when they were mandated to do so, which seems reasonable to the extent that these reforms implied a costly reorganization and municipalities may not be able to precisely measure their population without a census being taken.

significant, the magnitudes are similar as in the full sample.[29]

I interpret these result as evidence that merit system mandates improved police performance. For this interpretation to hold, the effect cannot be driven by factors unrelated to police actions. To the extent that unobservables vary continuously at the threshold and there are no pre-treatment differences in the socio-economic composition of control and treated municipalities, the effect is unlikely to be explained by other external factors, which supports the interpretation that police performance improved.

Moreover, it must be the case that the decline in property crime rates represents a true decline in crime, and not just in crime statistics: there must be no differential reporting at the threshold. I provide three pieces of evidence that this is the case, related to different ways in which differential reporting may arise. First, citizens who experience a crime may not report it or, even if the crime is reported, the police may fail to create a record for it. Misreporting at this stage is less likely for crimes that involve insured goods such as burglaries and vehicle thefts, as insurance companies often would not honor theft claims without a police report. Appendix Table VII shows that merit systems had a negative effect both on the burglary and vehicle theft rate and on the larceny rate, although the coefficients on burglary and auto theft are not significant for all bandwidths. Second, after a record is created, it can be altered to distort crime incidents reported to the FBI. In particular, as discussed in Mosher, Miethe, and Hart (2010), an offense can be downgraded to a non-index crime or it can be reported as unfounded. The fact that I find similar effects across crime types is reassuring as not all crimes can be downgraded as easily.[30] Third, the department may fail to submit a report to the FBI as participation in the UCR program is voluntary. I can exclude the possibility since, as Appendix Table VIII

[29] A true anticipation effect is also consistent with the coefficient in the pre-treatment sample in Table I being negative for some bandwidths, although always smaller in magnitude than the effects I estimate in the post-treatment sample and never statistically significant. In fact, as shown in Appendix Table VI, estimating the main specification dropping pre-treatment years in which the anticipation effect is likely gives coefficients that are smaller in magnitude and, again, never statistically significant.

[30] In particular, larcenies below $50 are not an index crime, which makes them particularly susceptible to the issue. Unfortunately, counts of unfounded offenses are not reported before 1978 so I cannot test directly whether this dimension is affected.

shows, there is no discontinuity at the threshold in the probability of submitting crime data for any given month. Overall, it seems unlikely that the effects are driven by differential crime reporting.

Finally, to support the interpretation that merit system mandates did improve police performance, I explore what happened to a different set of outcomes that also proxy for police performance: clearance rates, defined as the number of crimes cleared by arrest over total crimes. Figure IV presents the RD graphs for property and violent crime clearance rates separately for the pre-treatment sample (graph to the left) and for the post-treatment sample (graph to the right). Even if there is no difference in the pre-period, the violent crime clearance rate is higher in places above the threshold with respect to places below after the mandate becomes effective. Table I confirms this. There is no difference in violent crime clearance rates in the pre-treatment sample, but, in the post-period, the coefficient is positive and statistically significant at the 5% level: police departments in municipalities just above the threshold are 12% more likely to clear a violent crime by arrest than those just below. At the same time, there is no difference in the property crime clearance rate, either pre- or post-treatment.[31] Finally, Figure IV panel B shows the event study graph for the violent crime clearance rate. The event study graph, although noisier, shows a similar time pattern in the treatment effect as the one for the property crime rate: a gradual increase in police performance starting two years after the introduction of the reform and a constant effect thereafter.[32]

In short, despite no pre-treatment differences, municipalities just above the threshold had lower property crime rates and higher violent crime clearance rates: merit system mandates had a positive effect on police performance.

[31] The fact that property crime rates decrease but we only see an increase in violent crime clearance rates is somewhat puzzling. A possible explanation is that as most property crimes never get cleared in the first place (clearance rates for property crimes are around 20% whereas clearance rates for violent crimes are much higher, at around 60%) so it may be reasonable for police officers exerting more effort under merit systems to focus on violent crimes investigations.

[32] Appendix Figure III shows the event study graphs separately by whether the mandate was explicitly based on the federal census only. States in which a "true" anticipation effect was likely present positive coefficient estimates before 1970, in line with the results for property crime rates, even though this anticipation effect is not strong enough to be reflected in the pooled estimates.

Robustness checks

In this section, I show that my results are robust to a number of potential concerns. Figure V shows the coefficient for the dummy for being above the threshold, together with 95% confidence intervals, for a number of robustness checks. The relevant comparison is whether each coefficient is different than the one estimated using the baseline specification reported at the top of each graph.[33]

To begin with, Figure V panel A shows that the results are robust to the data cleaning procedure. First, while I include simple assault in the violent crime definition because it is the most common type of violent crime in these small municipalities, the results are the same if I do not. Second, the results are also robust to following data cleaning procedures aimed at identifying outliers similar to those used in Evans and Owens (2007), Chalfin and McCrary (Forthcoming) and Mello (2018) (for more details, see Online Appendix C). Third, using smoothed UCR population as opposed to linearly interpolated population from the census to define crime rates also does not make a difference.

A potential concern is that places right above the threshold have different population dynamics with respect to places just below, and the estimates are picking up the fact that crime rates vary by population. Figure V panel A shows that population dynamics do not explain my findings: the main results survive controlling for 1980 population.

Finally, the results are also robust to using different sample restrictions and different specifications. First, I show that the results do not change if I restrict the analysis to a quasi-balanced sample of municipalities reporting crime data at least half of the time. Second, the results are robust to controlling for baseline municipality characteristics. Third, the results are also robust to estimating a differences-in-differences specification with city fixed effects: the coefficient on the property crime rate is smaller but still significant at the 10% level.[34] Fourth, clustering stan-

[33] The equivalent tables are Appendix Table IXa, Appendix Table IXb and Appendix Table IXc. I only show estimates for a 1,000 bandwidth for clarity, but the estimates for the full set of bandwidths can be found in the Online Appendix.

[34] This specification is similar to equation (1), but includes municipality fixed effects and allows

dard errors at the municipality and county-year level to allow for errors to be correlated for places that are close to each other does not make a difference.

For RD designs to recover causal effects of a certain policy, it must be the case that no other policies change at the same threshold. Appendix Table I shows that most of the states have at least one legislative provision that implies a policy discontinuity at the same cutoff, although most of them are not police related. However, no single provision is the same across states, which means that I can provide evidence that no other policy explains the effect by showing that the results are robust to dropping one state at a time. Were the effects driven by any of the other policy discontinuities, they should disappear once the state is dropped. Figure V panel B shows the result of this exercise. The magnitude of the coefficients is stable across samples, with the exception of the coefficient on the violent crime clearance rate that is almost double in magnitude when Illinois is dropped. The stability of the coefficients suggests that no other policy has a strong enough effect to bias the results or, in other words, collinear policies satisfy an "ignorability" assumption as defined in Eggers et al. (Forthcoming). Moreover, given that different states had different thresholds, this exercise also points towards the results not being driven by potential changes in population-based federal policies, such as eligibility for federal grants.

Finally, Figure V panel C shows that the specific estimation technique used does not matter for the results. First, I show robustness to using a triangular and an Epanechnikov kernels. The results are not affected, although the coefficient for the property crime rate is larger in magnitude. Second, I estimate the main specification using locally quadratic regression and locally cubic regression with a uniform kernel. The result on property crime rates is robust to using polynomials of different orders, but the result on violent crime clearance rates is not. In particular, although the magnitude and sign of the coefficient are similar, the coefficients for violent crime clearance rates are not significant in the post-treatment period. In addition, the results are robust to dropping the state-month fixed effects and allowing the run-

the flexible controls of the running variable to vary by year. It is estimated on the 1960 to 1979 period. I prefer equation (1) as my baseline specification because, as discussed in Lee and Lemieux (2010) and Hinnerich and Pettersson-Lidbom (2014), municipality fixed effects are not necessary for identification but introduce more restrictions.

ning variable to vary flexibly both by census and by outcome year as in the event study specification. Overall, the results appear to be robust to potential concerns.[35]

Merit system adoption

The results presented show the effect of merit system mandates, as no data on adoption of full-fledged merit systems exists for the 1970s. In this section, I exploit historical data on merit system adoption and the fact that some states introduced the mandates in the first half of the 20th century to show that the legislation was effective at inducing municipalities to adopt merit systems, at least before 1940.

I proxy for the presence of a full-fledged merit system using year of introduction of a civil service board, available from a census of civil service agencies. Table II shows the coefficient on the dummy for being above the threshold before and after treatment. Given that the outcome data are available until 1940, the first stage exploits variation in treatment status from the 1900, 1910, 1920 and 1930 census experiments.[36] There is no discontinuity at the threshold in the probability of having a civil service board before the mandate is introduced. In the post-period, however, places above the threshold are 33% to 43% more likely depending on the bandwidth to have a civil service board than the places below. The coefficients are statistically significant at the 5% level (at the 10% level in column 8). The effect is large but less than one, both because some places below the threshold introduced a civil service board and because some places above the threshold failed to. In fact, the event study graph shown in Appendix Figure IV panel B shows that the effect

[35]The discussion of the robustness checks has focused on the effect on the property crime rate and on the violent crime clearance rate in the post-treatment period. With few exceptions, the rest of the results - in particular, that there are no pre-treatment differences in property crime rates and violent crime clearance rates - are robust to the different choices of sample, specifications and estimation, with two exceptions. First, when the median household income is included among the baseline municipality characteristics, the pre-treatment coefficient estimate in the property crime rate analysis is negative and statistically significant, although smaller in magnitude then the coefficient estimate for the post-treatment period under the same specification (Online Appendix Tables 6a and 6b). Second, the coefficient is also negative and significant at the 10% level when the estimation uses a quadratic polynomial with a uniform kernel (Online Appendix Table 12).

[36]When I have data for multiple census experiments, I stack the year by municipality panels and estimate equation (1) including state-month/year-census experiment fixed effects and allowing the controls in the running variable to vary by census experiment.

of the mandate became larger over time, suggesting that there were some delays between when treatment was assigned and when a civil service board was created.[37] Overall, merit system mandates were effective at inducing municipalities to adopt, although there were delays in implementation.[38]

6 Mechanisms

In this section, I explore three potential mechanisms that may explain why merit system mandates improved the performance of police departments: increased resources available to police departments, changes in police officers' characteristics, and reduced political influence through protections from discretionary dismissals.[39]

Resources

I begin by ruling out that the effect can be explained by increases in the resources available to departments under a merit system. In particular, I test whether departments above the threshold had higher expenditures or employed more police officers by estimating equation (1) using data from the Annual Survey of Local Governments and the Census of Governments 1972-1979. Table III shows that places above and below the threshold had similar expenditure and employment rates. Departments operating under merit systems and under spoils systems used similar inputs and, most importantly, there was no adjustment in labor supply along the extensive margin. Significant changes in the labor supply of police officers along the intensive margin (for example, through overtime hours) are also unlikely,

[37] It is not surprising that the anticipation effect does not appear in the pre-1940 merit system adoption analysis. First, the majority of the sample is composed of municipalities from states in which the mandates are explicitly based on the federal population census. Second, the anticipation effect is not present when the mandate becomes effective based on the introduction of new statewide reforms, as is the case in many of the experiments included in the sample.

[38] I refrain from using these estimates to scale the effects discussed before because they are too small and underestimate adoption for the 1970 sample. First, whether the municipality has a civil service board is an imperfect measure of merit system adoption as it ignores the fact that protections granted to police department employees were valid and violations could be challenged in court from the moment in which an official population census was published. Second, the pre-1940 sample does not take into account the anticipation effects in reform adoption that were instead likely in the 1970s.

[39] I present these results in table form, but the equivalent RD graphs can be found in the Appendix.

as we would expect them to be reflected in payroll expenditures.[40] In short, merit systems had no effect on resources.

Police officers' characteristics

Merit system mandates may have a positive effect on performance by helping police departments attract and retain more productive officers. First, police officers in departments under a merit system may receive more training. According to the Olmstrom survey (1974) described in the background section, almost all police departments of municipalities with population below 10,000 people required training, but almost none provided training in house. To the extent that the departments would have covered these costs, the fact that expenditures did not change suggests that large adjustments along the training margin are unlikely.

Second, merit systems may affect selection: directly, by changing control over the final decision on who to hire, and indirectly, by changing the attributes of the job and thereby inducing different people to apply.[41] I study whether selection was affected by testing for discontinuities in the demographic composition of police departments. I measure the demographic composition of police departments starting from the microdata from the population census 1960 to 1980. In each census I focus on places that fell under the mandate ten years prior to allow for any effect to actually take place. I focus on outcomes that relate to the human capital of police officers: age, education and whether the police officer was a veteran.[42]

Table IV shows that places with and without a merit system appear to have police departments with comparable levels of human capital. There is no difference in the share of police officers with a high school degree or in average age. More-

[40] It is possible that police officers have the same labor supply but the fraction of time spent actively policing (for example the fraction of time spent patrolling) increases. This would not be picked up by payroll expenditures, but I interpret these adjustments as changes in effort.

[41] Historically, the shift from a spoils to a merit system implied the introduction of formal testing procedures. By the 1970s, both municipalities with and without a merit system had in place procedures to screen potential police officers (Leonard, 1970). Selection tests comprised a medical examination, a physical test, and aptitude tests that usually included sections regarding police work, verbal and quantitative ability, and general knowledge (Rawson, 1980).

[42] Almost all police officers in my sample are white males: there is not enough variation to test whether merit systems had an effect on the racial or gender composition of police departments.

over, there is no difference in the share of individuals who were veterans, which is interesting to the extent that merit systems sometimes also introduced veteran preferences. Coefficients are generally small and are never significantly different than zero. The zero coefficients, however, are not precisely estimated, which means that I can only rule out large effects being explained by selection.

Overall, merit systems did not impact the observable characteristics of police officers. While it is still possible that the unobserved characteristics of police officers differed under the two systems, the fact that I find no clear break in any of these salient dimensions suggests a limited role for selection in explaining the performance improvement. Moreover, this interpretation is also consistent with the time pattern of the effect highlighted by the event study graphs in Figure III and Figure IV: had the effect mainly been driven by changes in who police officers were, we would expect them to take a longer time to appear.[43]

Finally, by limiting dismissals, merit systems may decrease turnover and as a result retain police officers with more experience. I can proxy for turnover using the 1970 and 1980 census data by identifying police officers who did not have the same job five years prior.[44] Appendix Table XIII shows no effect on turnover. The coefficients are not significant, and, if anything, positive: disruption does not explain why places under a spoil system had higher crime rates.[45]

Limitations to political influence

Given that the effect of police professionalization on performance cannot be explained by increased resources or changes in selection, the limitations to political influence introduced by merit systems are likely to be important. I provide indirect

[43] A previous version of the paper discussed the effect of merit system mandates on the demographic composition of police departments using the full count microdata for the 1910-1940 population censuses. Overall, I did not find evidence of selection being affected by merit systems in the pre-1940 period: merit systems did not impact the probability that foreigners were hired, they did not change the degree of ethnic patronage and they did not improve human capital.

[44] These are police officers who lived in another state, were in the armed forces or attended college five years before each census was taken.

[45] The same table also shows no discontinuity in average wage, which suggests that improved performance cannot be explained by police officers having stronger monetary incentives in merit system departments.

evidence of this by looking at what happens when protections from dismissals for political reasons are not part of the treatment.

At the end of the 1970s, a series of U.S. Supreme Court decisions made dismissals for political reasons illegal for all non-policymaking municipal employees. When municipalities grew above the threshold, they were still mandated to create independent civil service commissions, but there was no discontinuity in whether dismissals for political reasons could be used to influence police officers' behavior: they could not, neither in the treatment nor in the control group. As a result, I can study the effect of merit system mandates after 1980 to provide indirect evidence of the role of the provision in explaining the effect on performance.[46]

Table V shows the effect of merit systems on performance for the 1980 census experiment. There is no discontinuity at the threshold in crime or clearance rates: merit systems appear to have no effect when they do not imply a discontinuity in protections from dismissals for political reasons.[47,48] This is consistent with the hypothesis that the limitations to politicians' influence that came with merit systems were important to explain the effect on performance. What makes this result especially interesting is that the setting studied, small town police departments in the 1970s, does not appear to be characterized by high levels of patronage and corruption. It is unclear what the true extent of patronage was in this period. Overall, the excessive corruption that had characterized police employment under political machines was a thing of the past. Banfield and Wilson (1963) argue that "the more common practice among small cities without a civil service system is a rather informal but at the same time highly nonpolitical personnel system." However, they also

[46]It is important to note that the analysis presented in this section hinges on the assumption that no other reform interacting with merit systems took place at the end of the 1970s, and I cannot rule out that the null results in 1980 may be caused by other changes impacting policing during this decade.

[47]The coefficient for the property crime rate is negative for the MSE-optimal bandwidth, but visual inspection of the corresponding RD graph reported in Appendix Figure VII suggests that this is driven by places right below the threshold having an especially high property crime rate. The same pattern explains why the linear fit shown in the violent crime rate graph seems to suggest a negative effect, even if coefficient estimates are never significant.

[48]A previous version of the paper showed results pooling together the 1980, 1990 and 2000 census experiment. I prefer to focus on the 1980 census experiment only to focus on the census experiment closest to the main results, as so to minimize the influence of time variation in explaining the change. The main take-aways are unchanged.

reckon that many appointments were indeed political. Consistent with this interpretation, Freedman (1994) states: "there are probably thousands of small pockets of patronage lodged in the 80,000 plus units of local government in the United States." Still, even in this setting, merit systems implied a shift from an informal organizational system with power over hiring and firing in the hands of the political authority, to a professionalized bureaucracy in which this power was much more limited.

Taking this into consideration, how can we rationalize the effect of merit systems going through limitations to political influence? First, even in the absence of outright patronage, changing who is in charge of the police department can affect the ultimate incentive structure faced by police officers, which may impact effort allocation. Moreover, merit systems may affect police officers' motivation. While I cannot provide direct evidence for this hypothesis, the explanation that motivation is important to explain police officers' performance is consistent, for example, with previous work on police departments by Mas (2006), who showed that final offer arbitration decisions against the wage required by the police officers have a negative effect on performance. Finally, merit systems may also change the organizational culture of the department.

7 Conclusion

Merit systems reducing politicians' control over bureaucrats' hiring and firing foster expertise and create a long-term incentive structure, but come at the cost of decreased responsiveness to the executive and the electorate. Whether they improve performance is unclear a priori and must be ascertained empirically. I address the question by looking at the introduction of merit systems for U.S. municipal police departments in the 1970s. To address potential endogeneity concerns in reform adoption, I exploit statewide merit system mandates based on population thresholds to implement a regression discontinuity design. I find that merit systems increased performance. In the first ten years after the reform, the property crime rate was 46% lower and the violent crime clearance rate was 12% higher in municipalities just above the threshold with respect to municipalities just below.

Providing well-identified empirical evidence of the effect of merit systems on performance is the principal contribution of the paper. The finding that professionalizing a public organization improves performance is consistent with cross-country correlations (e.g. Evans and Rauch, 1999; Rauch and Evans, 2000), evidence from large U.S. cities (Rauch, 1995) and recent work on perceived determinants of bureaucrats' effectiveness (Oliveros and Schuster, 2016) and on management practices and public service delivery (Rasul and Rogger, 2016).

Looking at the mechanisms suggests that merit systems' positive effect on performance is likely explained by the fact that they reduce a politicians' ability to influence the incentive structure that police officers face on the job. Whereas it is no surprise that political influence may distort public employees' behavior (e.g., among others, Eynde, Moradi, and Kuhn, 2016), what makes this result especially interesting is the fact that it holds in what appears to be an informal but relatively low patronage setting. Understanding the mechanisms behind this particular result is a fascinating question that I hope to address in future research.

References

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee. 2016. "Do-gooders and Go-getters : Career Incentives, Selection, and Performance in Public Service Delivery." Working paper.

Banfield, Edward C. and James Q. Wilson. 1963. *City Politics*. Harvard University Press and The M.I.T. Press.

Bartalotti, Otavio and Quentin Brummet. 2016. "Regression Discontinuity Designs with Clustered Data: Mean Square Error and Bandwidth Choice." In *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, edited by Matias D. Cattaneo and Juan C. Escanciano. Emerald Group Publishing.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6):2295–2326.

Carpenter, Daniel. 2005. "The Evolution of National Bureaucracy in the United States."

Chalfin, Aaron and Justin McCrary. Forthcoming. "Are US Cities Under-Policed? Theory and Evidence." *Review of Economics and Statistics* .

Chong, Alberto, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2014. "Letter Grading Government Efficiency." *Journal of the European Economic Association* 12 (2):277–299.

Civil Service Assembly of the United States and Canada. 1937. "Civil Service Agencies in the United States: A 1937 Census."

———. 1940. "Civil Service Agencies in the United States: A 1940 Census."

———. 1943. "Civil Service Agencies in the United States: A 1943 Supplement."

Eggers, Andrew C., Ronny Freier, Veronica Grembi, and Tommaso Nannicini. Forthcoming. "Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions." *American Journal of Political Science* .

Evans, Peter and James E. Rauch. 1999. "Bureaucracy and Growth: A Cross-national Analysis of the Effects of" Weberian" State structures on Economic Growth." *American Sociological Review* :748–765.

Evans, William N. and Emily G. Owens. 2007. "COPS and Crime." *Journal of Public Economics* 91 (1):181–201.

Eynde, Oliver Vanden, Alexander Moradi, and Patrick M. Kuhn. 2016. "Trickle-Down Ethnic Politics: Drunk and Absent in the Kenya Police Force (1957-1970)." Centre for the Study of African Economies, University of Oxford.

Fogelson, Robert M. 1977. *Big-City Police*. Harvard University Press Cambridge, MA.

Folke, Olle, Shigeo Hirano, and James M. Snyder. 2011. "Patronage and Elections in U.S. States." *American Political Science Review* 105 (03):567–585.

Fowler, Dorothy Ganfield. 1943. *The Cabinet Politician: The Postmasters General, 1829-1909*. Columbia University Press.

Freedman, Anne E. 1994. *Patronage: an American Tradition*. Wadsworth Publishing Company.

Gelman, Andrew and Guido Imbens. 2016. "Why High-order Polynomials should not be used in Regression Discontinuity Designs." NBER Working Paper 19649.

Hinnerich, Björn Tyrefors and Per Pettersson-Lidbom. 2014. "Democracy, Redistribution, and Political Participation: Evidence From Sweden 1919-1938." *Econometrica* 82 (3):961–993.

Hoover, J Edgar. 1938. "Lawlessness - A National Menace." *American Journal of Medical Jurisprudence* 1:242–246.

Hyden, Goran, Julius Court, and Ken Mease. 2003. "The Bureaucracy and Governance in 16 Developing Countries." Overseas Development Institute, World Governance Survey Discussion Paper 7.

Iyer, Lakshmi and Anandi Mani. 2012. "Traveling Agents: Political Change and Bureaucratic Turnover in India." *Review of Economics and Statistics* 94 (3):723–739.

Johnson, Ronald N. and Gary D. Libecap. 1994. *The Federal Civil Service System and the Problem of Bureaucracy*. University of Chicago Press.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido. 1999. "Governance Matters." World Bank Policy Research Working Paper 2196.

La Porta, Rafael, Florencio Lopez-de Silanes, Andrei Shleifer, and Robert Vishny. 1999. "The Quality of Government." *Journal of Law, Economics, and Organization* 15 (1):222–279.

Lee, David S. and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (June):281–355.

Leonard, Vivian A. 1970. *Police Personnel Administration*. Charles C. Thomas Publisher Ltd.

Levitt, Steven D. 1997. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." *American Economic Review* 87 (3):270–290.

Lewis, David E. 2010. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance*. Princeton University Press.

Maltz, Michael D. 2006. *Analysis of Missingness in UCR Crime Data*. Criminal Justice Research Center, Ohio State University.

Mas, Alexandre. 2006. "Pay, Reference Points and Police Performance." *Quarterly Journal of Economics* 121 (3):783–821.

McCrary, Justin. 2007. "The Effect of Court-ordered Hiring Quotas on the Composition and Quality of Police." *The American Economic Review* 97 (1):318–353.

———. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2):698–714.

Mello, Steven. 2018. "More COPS, Less Crime."

Mosher, Clayton J., Terance D. Miethe, and Timothy C. Hart. 2010. *The Mismeasure of Crime*. Sage Publications.

Oliveros, Virginia and Christian Schuster. 2016. "Merit, Tenure, and Bureaucratic Behavior: Evidence from a Conjoint Experiment in the Dominican Republic." Working Paper.

Ostrom, Elinor, Roger B Parks, and Gordon P Whitaker. 1977. *Policing Metropolitan America.* Superintendent of Documents, U.S. Govt. Printing Office, Washington, D.C. 20402.

Rasul, Imran and Daniel Rogger. 2016. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." CEPR Discussion Paper No. DP11078.

Rauch, James E. 1995. “Bureaucracy, Infrastructure, and Economic Growth: Evidence from U.S. Cities During the Progressive Era.” *American Economic Review* 85 (4):968–979.

Rauch, James E. and Peter B. Evans. 2000. “Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries.” *Journal of Public Economics* 75 (1):49–71.

Rawson, Zivile A., editor. 1980. *How to Pass Civil Service Examinations, Patrolman.* Civil Service Publishing Corporation, Brooklyn.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. “Integrated Public Use Microdata Series: Version 6.0. [Machine-readable database].”

Tolbert, Pamela and Lynne Zucker. 1983. “Institutional Sources of in the Formal Change Structure of Organizations: The Diffusion of Civil Service Reform, 1880 - 1935.” *Administrative Science Quarterly* 28 (1):22–39.

Ujhelyi, Gergely. 2014. “Civil Service Rules and Policy Choices: Evidence from US State Governments.” *American Economic Journal: Economic Policy* 6 (2):338–380.

United States Civil Service Commission. 1973. *Biography of an Ideal: A History of the Federal Civil Service*. Office of Public Affairs, U.S. Civil Service Commission.

U.S. Census Bureau. 1970-1980. “Annual Survey of State and Local Government Finances and Census of Governments.”

———. 1972-1980. “Annual Survey of State and Local Government Employment and Census of Governments.”

U.S. Civil Service Commission. 1884. “Annual Reports.”

White, Leonard Dupee. 1954. *The Jacksonians: A Study in Administrative History, 1829-1861*. Macmillan.

Wilson, Orlando W. 1954. “Toward a Better Merit System.” *The Annals of the American Academy of Political and Social Science* 291 (1):87–96.

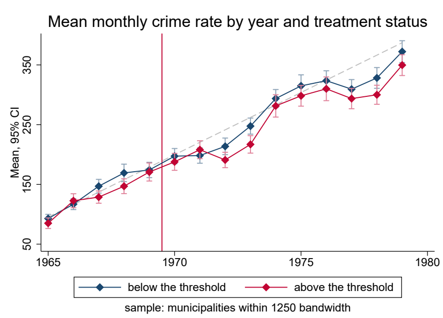
Xu, Guo. 2017. “The Costs of Patronage: Evidence from the British Empire.” Working paper.

Figure I: Population-based merit system mandates for police departments

state	year	threshold
Arizona	1969	15,000
Illinois	1949 & 1951 & 1957	15,000 & 13,000 & 5,000
Iowa	1917	8,000
Louisiana	1944 & 1964	13,000 & 7,000
Montana	1907 & 1947 & 1975	10,000 & 5,000 & 0
Nebraska	1957	5,000
West Virginia	1937 & 1969	5,000 & 10,000
Wisconsin (cities)	1917	4,000
Wisconsin (villages)	1941	5,500

Notes: this table summarizes legislation mandating merit systems by state. For each states, it reports the year in which a population-based mandate was introduced and the corresponding threshold. When multiple years are reported, the threshold was modified over time. In 1975 Montana expanded the mandate to all municipalities.

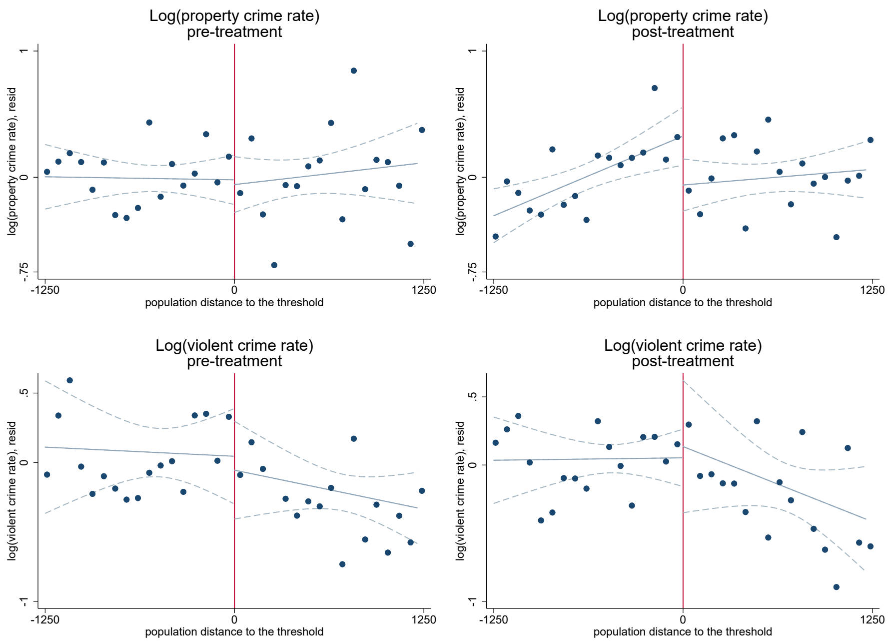
Figure II: Crime rates grow at slower pace in municipalities under merit system mandates



Notes: the graph shows the mean monthly total crime rate by year separately for municipalities above and below the threshold 1965-1979, together with 95% confidence intervals for the mean. The sample is restricted to municipalities within a 1250 distance from the threshold. The dashed line shows the predicted crime rate, using the property crime growth of the pre-treatment period. Merit systems are mandated for municipalities above the threshold in 1970.

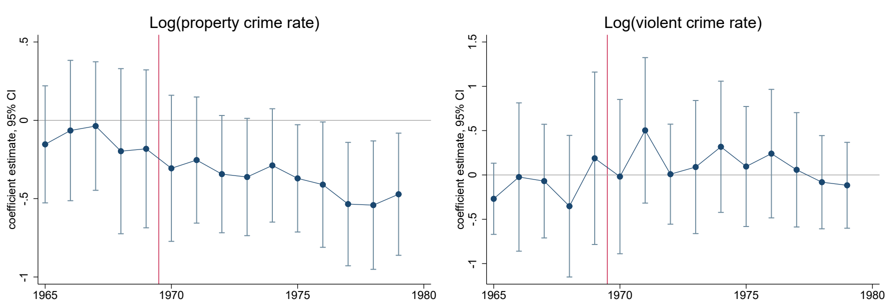
Figure III: Merit systems departments have lower property crime rates

Panel A: RD graphs



Notes: the graphs show the effect of merit system mandates on monthly property and violent crime rates for pre-treatment years (1960 to 1969, on the left) and post-treatment years (1970 to 1979, on the right). Merit systems are mandated for municipalities above the threshold in 1970. Crime rates are crimes per 100,000 people. The points show the average value of the outcome within a 75 population distance bin; the line plots a linear fit estimated separately on each side of the discontinuity and prediction intervals that allow for clustering at the municipality level. State-month fixed effects are partialled out.

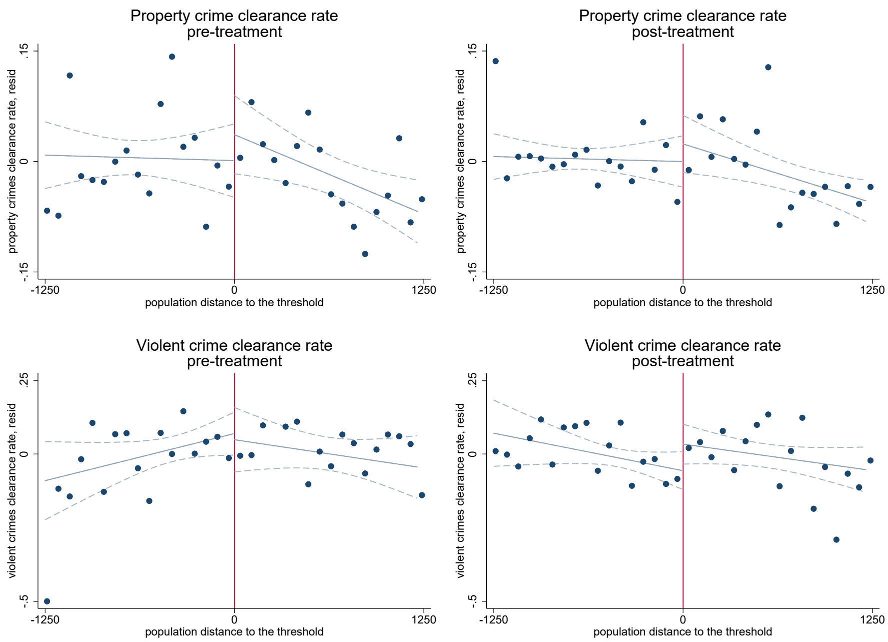
Panel B: Event study graphs



Notes: the graphs show the effect of merit system mandates estimated using the event study specification (equation (2)) on monthly property and violent crime rates for the full sample of states 1965 to 1979. The sample exploits variation in treatment status from the 1970 census experiment. Crime rates are crimes per 100,000 people. The points are the point estimates β_σ from the event study specification with 95% confidence intervals. The coefficients are estimated using locally linear regression and a uniform kernel for a 1250 bandwidth. Standard errors are clustered at the municipality level.

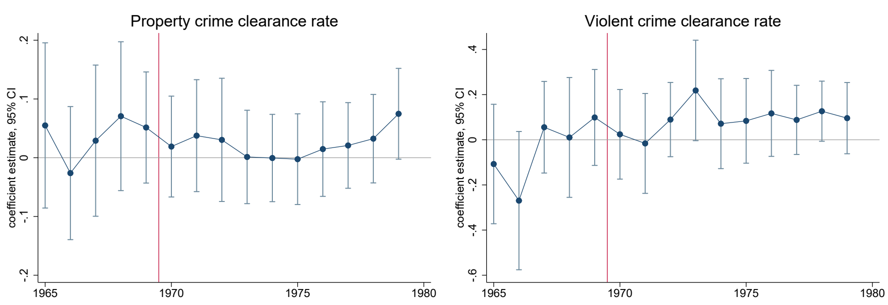
Figure IV: Merit systems departments have higher violent crime clearance rates

Panel A: RD graphs



Notes: the graphs show the effect of merit system mandates on monthly property and violent crime clearance rates for pre-treatment years (1960 to 1969, on the left) and post-treatment years (1970 to 1979, on the right). Merit systems are mandated for municipalities above the threshold in 1970. Clearance rates are number of crimes cleared by arrest over total number of crimes. The points show the average value of the outcome within a 75 population distance bin; the line plots a linear fit estimated separately on each side of the discontinuity and prediction intervals that allow for clustering at the municipality level. State-month fixed effects are partialled out.

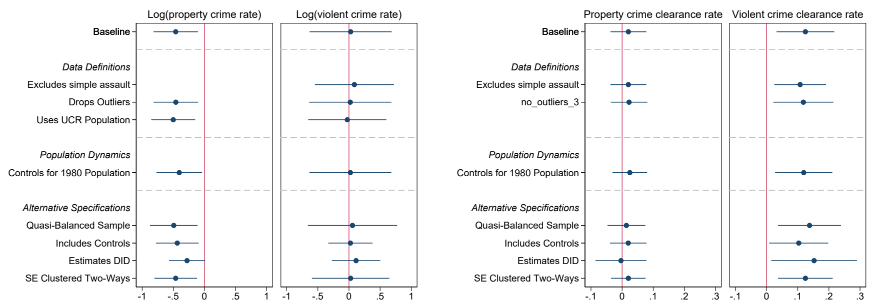
Panel B: Event study graphs



Notes: the graphs show the effect of merit system mandates estimated using the event study specification (equation (2)) on monthly property and violent crime clearance rates for the full sample of states 1965 to 1979. The sample exploits variation in treatment status from the 1970 census experiment. Clearance rates are number of crimes cleared by arrest over total number of crimes. The points are the point estimates β_σ from the event study specification with 95% confidence intervals. The coefficients are estimated using locally linear regression and a uniform kernel for a 1250 bandwidth. Standard errors are clustered at the municipality level.

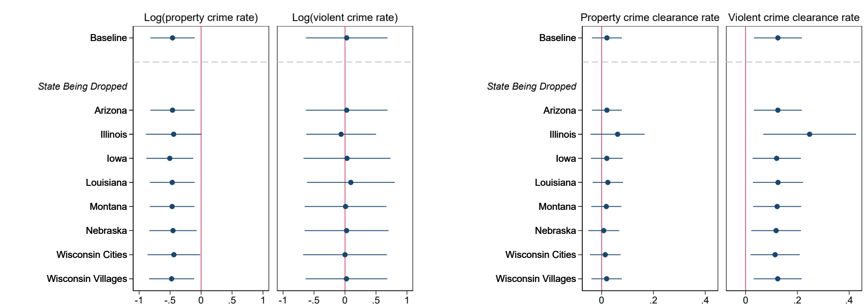
Figure V: Robustness checks

Panel A: Robustness to data cleaning, population dynamics and specification



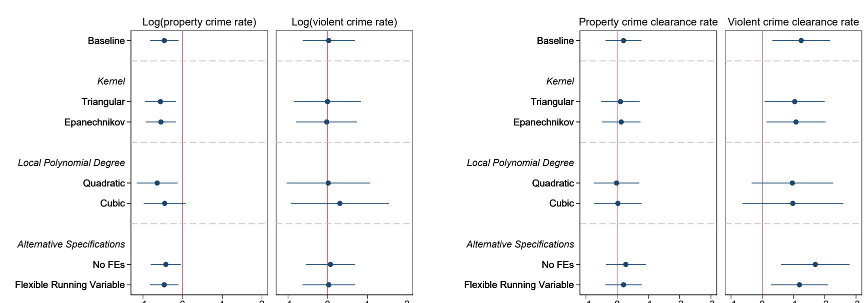
Notes: the graphs show that the main results are robust to different ways of defining the outcomes, population dynamics and alternative specifications. The graphs report RD estimates on crime rates (on the left) and clearance rates (on the right), together with 95% confidence intervals, for the sample of post-treatment years (1970 to 1979). Variation in treatment status is from the 1970 census experiment. All coefficients are estimated using locally linear regression and a uniform kernel for a 1000 bandwidth. Standard errors are clustered at the municipality level, and state-month fixed effects are always included.

Panel B: Robustness to overlapping legislation



Notes: the graphs show that the results are robust to dropping one state at a time. Outcomes, samples and estimation are as described in panel A.

Panel C: Robustness to estimation



Notes: the graphs show that the results are robust to using different estimation techniques. Outcomes, samples and estimation are as described in panel A.

Table I: Effect of merit system mandates on crime and clearance rates

Sample	pre-treatment				post-treatment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(property crime rate)	-0.293	-0.179	-0.034	-0.098	-0.587***	-0.461**	-0.394**	-0.620***
	(0.189)	(0.162)	(0.157)	(0.232)	(0.213)	(0.180)	(0.160)	(0.230)
Clusters	80	101	123	59	89	113	137	73
Observations	5715	7302	8790	4113	8891	11215	13589	7387
Bandwidth	750	1000	1250	583	750	1000	1250	632
Log(violent crime rate)	-0.254	-0.300	-0.106	-0.256	-0.030	0.027	0.091	-0.053
	(0.350)	(0.291)	(0.271)	(0.356)	(0.429)	(0.333)	(0.296)	(0.378)
Clusters	67	88	108	55	89	113	137	102
Observations	1059	1325	1624	892	4402	5540	6542	5048
Bandwidth	750	1000	1250	660	750	1000	1250	858
Property crime clearance rate	0.036	0.031	0.034	0.037	0.013	0.020	0.023	0.005
	(0.041)	(0.039)	(0.038)	(0.041)	(0.034)	(0.029)	(0.026)	(0.036)
Clusters	80	101	122	56	89	113	137	82
Observations	4329	5570	6648	2989	8891	11215	13589	8179
Bandwidth	750	1000	1250	556	750	1000	1250	703
Violent crime clearance rate	-0.024	-0.031	-0.030	-0.012	0.123**	0.125***	0.098**	0.126**
	(0.077)	(0.069)	(0.067)	(0.077)	(0.052)	(0.047)	(0.048)	(0.055)
Clusters	67	88	108	38	89	113	137	79
Observations	1059	1325	1624	658	4402	5540	6542	3971
Bandwidth	750	1000	1250	493	750	1000	1250	680

Notes: The table shows the effect of merit system mandates on police performance. It presents RD estimates on monthly crime rates and clearance rates for pre-treatment years (1960 to 1969, columns 1 to 4) and post-treatment years (1970 to 1979, columns 5 to 8). Variation in treatment status is from the 1970 census experiment. Crime rates are crimes per 100,000 people and clearance rates are number of crimes cleared by arrest over total number of crimes. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-month fixed effects are included in all columns.

Table II: Effect of merit system mandates on pre-1940 reform adoption

Sample	pre-treatment				post-treatment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Civil service board	0.185	0.096	0.183	0.190	0.334**	0.430**	0.437**	0.337*
	(0.151)	(0.159)	(0.138)	(0.183)	(0.168)	(0.177)	(0.171)	(0.198)
Observations	42	52	61	39	42	52	61	37
Clusters	646	863	1060	595	572	747	902	481
Bandwidth	750	1000	1250	713	750	1000	1250	651

Notes: The table shows the effect of mandates on the adoption of civil service boards in the pre-1940 sample. It presents RD estimates on an indicator variable for whether a municipality has a civil service board for the sample of pre-treatment years (columns 1 to 4) and post-treatment years (columns 5 to 8). Pre-treatment years span from the year of the previous census to the year before treatment is assigned. Post-treatment years span from the year in which treatment is assigned to the year before the following census. Variation in treatment status is from the 1900, 1910, 1920 and 1930 census experiments. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-year-census experiment fixed effects are included in all columns.

Table III: Effect of merit system mandates on expenditures and employment

Sample	post-treatment			
	(1)	(2)	(3)	(4)
Log(expenditures per 1,000 people)	-0.030	0.131	-0.034	0.020
	(0.208)	(0.186)	(0.163)	(0.202)
Clusters	89	113	137	95
Observations	492	632	753	531
Bandwidth	750	1000	1250	805
Log(employment per 1,000 people)	-0.112	-0.018	-0.092	-0.028
	(0.231)	(0.204)	(0.169)	(0.212)
Clusters	88	112	136	107
Observations	372	483	572	460
Bandwidth	750	1000	1250	940

Notes: The tables shows the effect of the merit system mandate on the resources available to the police department. The table presents RD estimates on yearly expenditures and employment for the sample of post-treatment years (columns 1 to 4). Post-treatment years are 1970 to 1979 for expenditures and 1972 to 1979 for employment. Variation in treatment status is from the 1970 census experiment. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-year fixed effects are included in all columns.

Table IV: Effect of merit system mandates on demographic composition of police departments

Sample	post-treatment			
	(1)	(2)	(3)	(4)
Share who finished high school	-0.120	0.028	0.026	-0.106
	(0.105)	(0.098)	(0.082)	(0.110)
Clusters	136	176	221	127
Observations	179	236	302	159
Bandwidth	750	1000	1250	650
Average age	3.345	-0.323	0.203	3.326
	(3.059)	(2.685)	(2.374)	(3.376)
Clusters	136	176	221	122
Observations	179	236	302	152
Bandwidth	750	1000	1250	622
Share veteran	-0.014	0.001	0.015	0.007
	(0.112)	(0.104)	(0.091)	(0.100)
Clusters	136	176	221	189
Observations	179	236	302	254
Bandwidth	750	1000	1250	1092

Notes: The table shows the effect of merit system mandates on demographic composition of police departments. It presents RD estimates on the share of police officers who have a high school degree, their average age and the share who have veteran status for the sample of post-treatment years (columns 1 to 4). Outcomes are measured in the 1960, 1970 and 1980 census, and variation in treatment assignment is from the 1950 to 1970 census experiments. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-census year fixed effects are included in all columns.

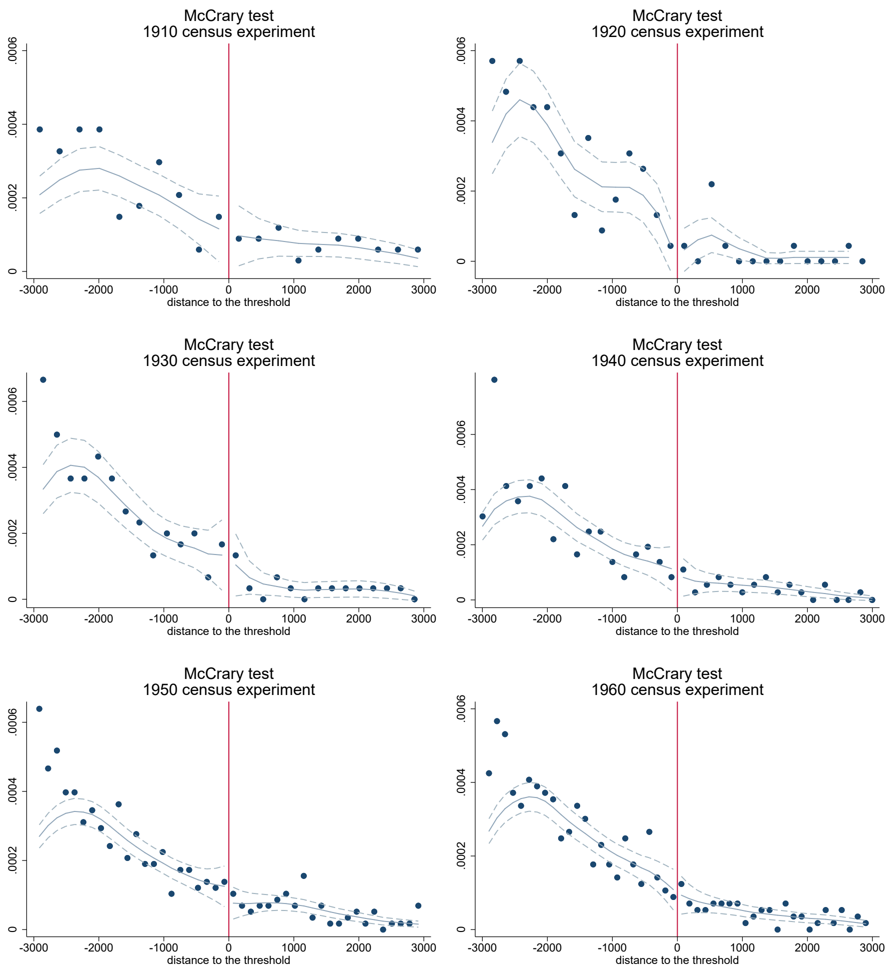
Table V: Effect of merit system mandates on crime and clearance rates post-1980

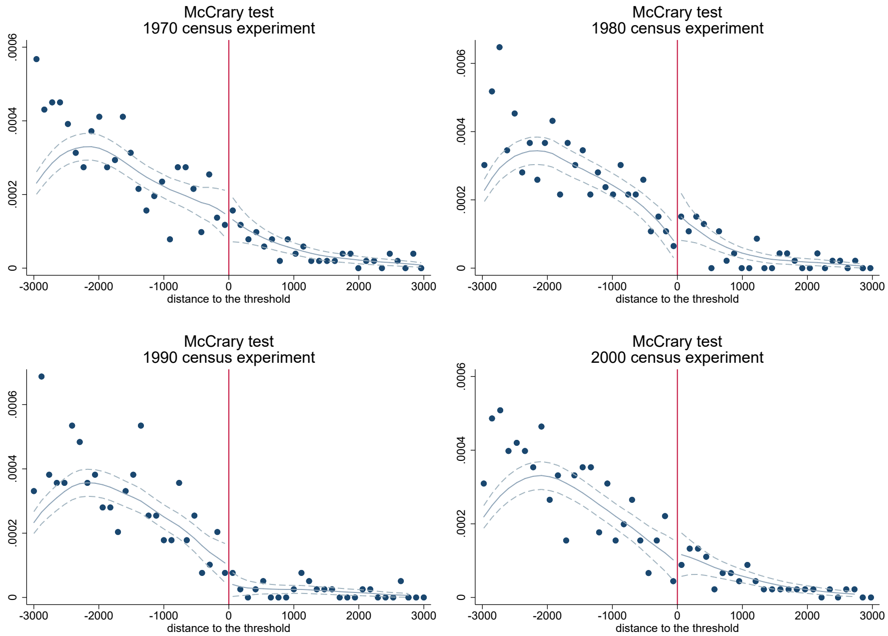
Sample	post-treatment			
	(1)	(2)	(3)	(4)
Log(property crime rate)	-0.167	-0.016	-0.027	-1.304***
	(0.210)	(0.192)	(0.150)	(0.357)
Clusters	74	102	127	22
Observations	8360	11464	14102	2470
Bandwidth	750	1000	1250	266
Log(violent crime rate)	-0.228	-0.061	-0.182	-0.346
	(0.184)	(0.154)	(0.143)	(0.319)
Clusters	74	102	127	24
Observations	6067	8330	10229	1830
Bandwidth	750	1000	1250	285
Property crime clearance rate	0.016	-0.021	-0.026	-0.023
	(0.029)	(0.026)	(0.024)	(0.042)
Clusters	74	102	127	32
Observations	8360	11464	14102	3617
Bandwidth	750	1000	1250	407
Violent crime clearance rate	-0.013	-0.007	-0.017	-0.039
	(0.104)	(0.082)	(0.064)	(0.139)
Clusters	74	102	127	45
Observations	6067	8330	10229	3648
Bandwidth	750	1000	1250	504

Notes: The table shows the effect of the merit system mandates on police performance when there is no discontinuity in whether police officers are protected from patronage dismissals. The table presents RD estimates on monthly crime rates and clearance rates for post-treatment years (1980 to 1989). Variation in treatment status is from the 1980 census experiment. Crime rates are crimes per 100,000 people and clearance rates are number of crimes cleared by arrest over total number of crimes. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-month fixed effects are included in all columns.

For Online Publication Only

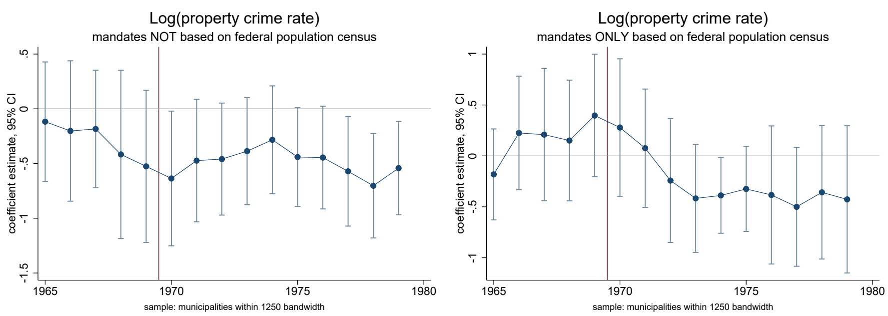
Appendix Figure I: McCrary tests 1910 to 2000





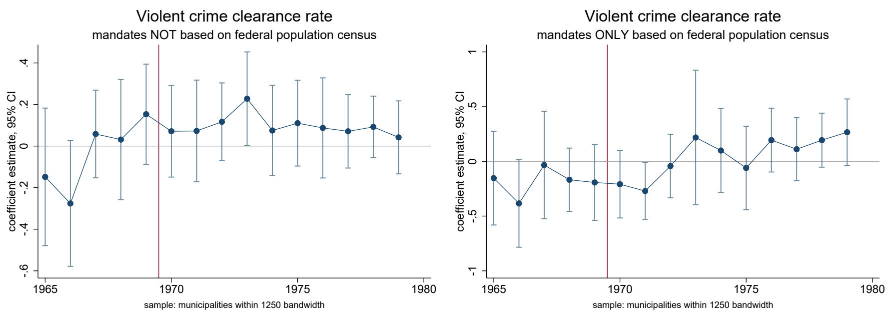
Notes: the graphs shows the McCrary test for the 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990 and 2000 census experiments.

Appendix Figure II: Merit systems lower property crime rates, event study graphs separately for states with and without mandates based on federal population census



Notes: the graph shows the effect of merit system mandates estimated using the event study specification (equation (2)) on property crime rates separately for states with and without mandates explicitly based on federal population census (panel (b)). Crime rates are crimes per 100,000 people. The sample exploits variation in treatment status from the 1970 census experiments. The sample includes both pre-treatment and post-treatment years and spans 1965 to 1979. The points are the point estimates β_σ from the event study specification with 95% confidence intervals. The coefficients are estimated using locally linear regression and a uniform kernel for a 1250 bandwidth. Standard errors are clustered at the municipality level.

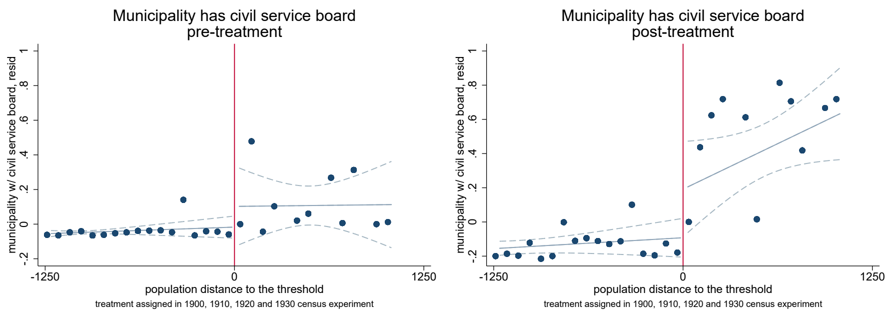
Appendix Figure III: Merit systems increase violent crime clearance rates, event study graphs separately for states with and without mandates based on federal population census



Notes: the graph shows the effect of merit system mandates estimated using the event study specification (equation (2)) on violent crime clearance rates separately for states with and without mandates explicitly based on federal population census (panel (b)). Crime rates are crimes per 100,000 people. The sample exploits variation in treatment status from the 1970 census experiments. The sample includes both pre-treatment and post-treatment years and spans 1965 to 1979. The points are the point estimates β_σ from the event study specification with 95% confidence intervals. The coefficients are estimated using locally linear regression and a uniform kernel for a 1250 bandwidth. Standard errors are clustered at the municipality level.

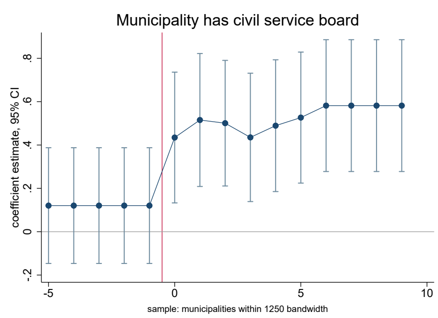
Appendix Figure IV: Merit system mandates increase reform adoption pre-1940

Panel A: RD graphs



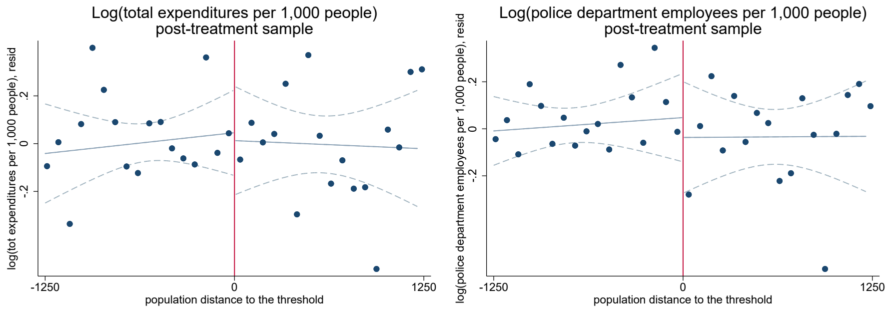
Notes: the graphs show the effect of merit system mandates on pre-1940 reform adoption for the sample of pre-treatment years (on the left) and post-treatment years (on the right). Merit systems are mandated for places above the threshold. The sample exploits variation in treatment status from the 1900, 1910, 1920 and 1930 census experiments. Pre-treatment years span from the year of the previous census to the year in which treatment is assigned. Post-treatment years span from the year in which treatment is assigned to the year before the following census. The points show the average value of the outcome within a 75 population distance bin; the line plots a linear fit estimated separately on each side of the discontinuity and prediction intervals that allow for clustering at the municipality level. State-year-census experiments fixed effects are partialled out.

Panel B: Event study graphs



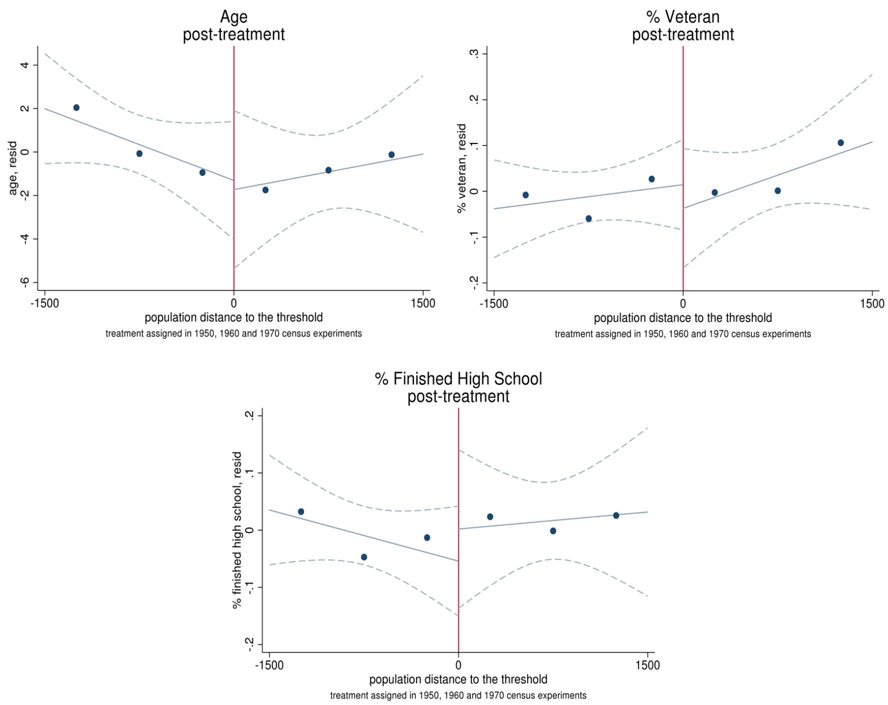
Notes: the graph shows the effect of merit system mandates on pre-1940 reform adoption estimated using the event study specification (equation (2)). The sample exploits variation in treatment status from the 1900, 1910, 1920 and 1930 census experiments. The sample includes both pre-treatment and post-treatment years. Pre-treatment years span from the year of the previous census to the year in which treatment is assigned. Post-treatment years span from the year in which treatment is assigned to the year before the following census. The points are the point estimates β_σ from the event study specification with 95% confidence intervals. The coefficients are estimated using locally linear regression and a uniform kernel for a 1250 bandwidth. Standard errors are clustered at the municipality level.

Appendix Figure V: Merit systems do not affect expenditures or employment



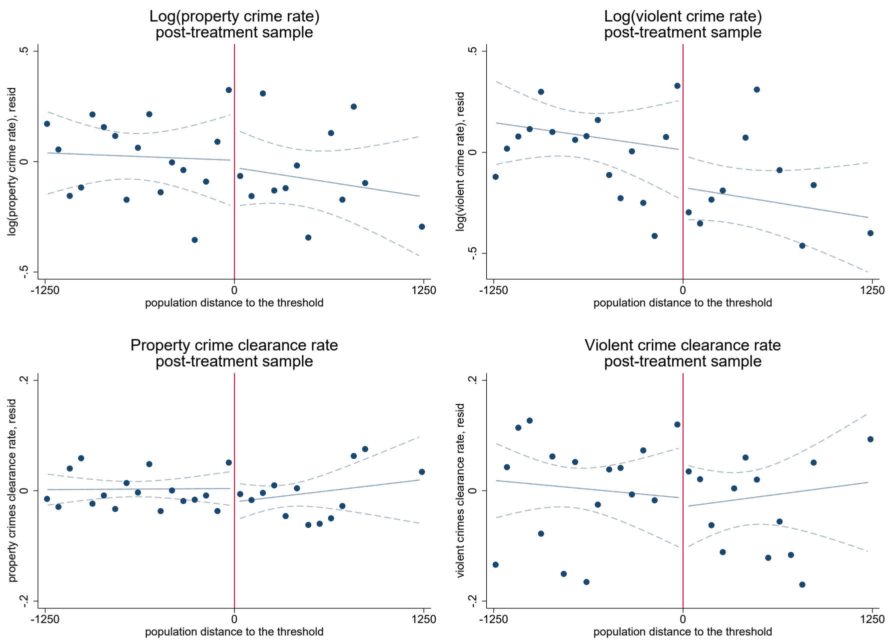
Notes: the graphs show the effect of merit system mandates on expenditures and employment for post-treatment years. Merit systems are mandated for municipalities above the threshold in 1970. Post-treatment years are 1970 to 1979 for expenditures and 1972 to 1979 for employment. The points show the average value of the outcome within a 75 population distance bin; the line plots a linear fit estimated separately on each side of the discontinuity and prediction intervals that allow for clustering at the municipality level. State-year fixed effects are partialled out.

Appendix Figure VI: Merit systems do not affect the demographic composition of police departments



Notes: the graphs show the effect of merit system mandates on the demographic composition of police departments (average age, share with veteran status, and share with high school degree). Merit systems are mandated for places above the threshold in 1950, 1960 and 1970. Outcomes are measured in the 1960, 1970 and 1980 census. The points show the average value of the outcome within a 500 population distance bin; the line plots a linear fit estimated separately on each side of the discontinuity and prediction intervals that allow for clustering at the municipality level. RD graphs are coarser to avoid disclosure. State-year fixed effects are partialled out.

Appendix Figure VII: Merit systems do not affect crime or clearance rates post-1980, RD graphs



Notes: the graphs show the post-1980 effect of merit system mandates on crime rates and clearance. Crime rates are crimes per 100,000 people. Clearance rates are number of crimes cleared by arrest over total number of crimes. Merit systems are mandated for places above the threshold. The sample exploits variation in treatment status from the 1980 census experiment. Post-treatment years span from the year of the census experiment to the year before the following census. The points show the average value of the outcome within a 75 population distance bin; the line plots a linear fit estimated separately on each side of the discontinuity and prediction intervals that allow for clustering at the municipality level. State-month fixed effects are partialled out.

Appendix Table I: Legislative provisions implying policy discontinuities at the same threshold

state	overlap with municipality classification	overlap with police legislation	details
Arizona	no	no	Other legislation: procedure to publish notice of bonds emission.
Illinois	no	yes	Police legislation: mimum salary. Other legislation: community nurses, parks, strong mayor form of government, arbitration procedure for firemen, pension fund for city employees (overlaps only for 2 years).
Iowa	no	no	Other legislation: appropriation of special funds on part of county to fund construction in certain cities.
Louisiana	no	no	-
Montana	yes	no	-
Nebraska	yes	yes	Police legislation: possibility to introduce pension funds for policemen. Other legislation: way of setting up a new charter.
West Virginia	yes	yes	Police legislation: pension and relief fund for policemen and firemen (after 1969 only). Other legislation: number of councilmen, incorporation procedure, bonds.
Wisconsin (cities)	no	no	-
Wisconsin (villages)	no	no	-

Appendix Table II: Descriptive statistics

Statistics	N	Mean	Sd
Number of municipalities in 1970 experiment	139		
Number of municipalities treated in the 1970 experiment	40		
Pre-treatment sample			
Property crime rate	7741	97.410	123.539
Violent crime rate	4790	14.904	41.021
Property crime clearance rate	4528	0.207	0.316
Violent crime clearance rate	1304	0.675	0.417
Post-treatment sample			
Property crime rate	9947	255.559	238.156
Violent crime rate	9947	29.811	52.020
Property crime clearance rate	9470	0.192	0.235
Violent crime clearance rate	4507	0.662	0.398

Notes: the table reports summary statistics (number of observations, mean and standard deviation) for property and violent crime and clearance rates for the sample of pre-treatment year (1960-1969) and post-treatment years (1970-1979). Crime rates are crimes per 100,000 people and clearance rates are number of crimes cleared by arrest over total number of crimes.

Appendix Table III: Covariate balance test

Census year	1970				1980			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Population growth	0.047	0.201	-0.042	0.180	-0.315	-0.095	-0.231	-0.314
	(0.356)	(0.390)	(0.295)	(0.395)	(0.311)	(0.249)	(0.216)	(0.383)
Observations	90	114	138	68	75	104	132	58
Bandwidth	750	1000	1250	602	750	1000	1250	636
Male	-0.004	-0.001	-0.002	-0.005	0.002	0.004	0.002	-0.001
	(0.007)	(0.006)	(0.005)	(0.007)	(0.010)	(0.008)	(0.007)	(0.009)
Observations	90	114	138	95	77	106	134	85
Bandwidth	750	1000	1250	794	750	1000	1250	832
Non-white	0.003	0.001	-0.001	0.004	-0.015	-0.004	-0.056*	-0.064***
	(0.035)	(0.031)	(0.026)	(0.037)	(0.022)	(0.023)	(0.031)	(0.020)
Observations	90	114	138	86	77	106	134	37
Bandwidth	750	1000	1250	725	750	1000	1250	412
Male 15 to 30	0.000	-0.007	-0.003	-0.010	0.009	0.009	0.006	0.007
	(0.023)	(0.020)	(0.017)	(0.026)	(0.010)	(0.008)	(0.007)	(0.010)
Observations	90	114	138	59	77	106	134	83
Bandwidth	750	1000	1250	537	750	1000	1250	817
Finished college	0.052	0.049	0.029	0.053	-0.038	-0.012	-0.025	-0.035
	(0.053)	(0.044)	(0.039)	(0.053)	(0.045)	(0.041)	(0.031)	(0.052)
Observations	90	114	138	88	77	106	134	63
Bandwidth	750	1000	1250	731	750	1000	1250	641
Unemployed	0.010	0.008	0.006	0.011	0.014	0.000	-0.001	0.015
	(0.013)	(0.011)	(0.009)	(0.013)	(0.018)	(0.015)	(0.013)	(0.022)
Observations	90	114	138	83	77	106	134	52
Bandwidth	750	1000	1250	705	750	1000	1250	548
Below poverty line	0.038	0.032	0.037*	0.038	-0.009	-0.015	-0.017	-0.009
	(0.025)	(0.022)	(0.020)	(0.025)	(0.019)	(0.016)	(0.015)	(0.016)
Observations	90	114	138	90	77	106	134	98
Bandwidth	750	1000	1250	746	750	1000	1250	922
Median hh income	1.009	1.322	0.567	1.447	0.011	2.319	0.341	-0.695
	(1.442)	(1.217)	(1.092)	(1.800)	(2.406)	(2.640)	(1.885)	(3.479)
Observations	90	114	138	62	77	106	134	48
Bandwidth	750	1000	1250	563	750	1000	1250	491

Notes: *** p<0.01, ** p<0.05, * p<0.1. The tables shows the results of a covariate balance test for the 1970 census experiments (columns 1 to 4) and the 1980 census experiment (columns 5 to 8). The table presents RD estimates on municipality characteristics at baseline for the samples of places to which treatment is assigned in the respective census experiment. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. State fixed effects are included in all columns. Robust standard errors are shown in parentheses.

Appendix Table IV: Effect of merit system mandates on reporting for the 1960 census experiment

Sample	post-treatment			
	(1)	(2)	(3)	(4)
Monthly crime report missing	-0.040	-0.175**	-0.146**	-0.143
	(0.101)	(0.078)	(0.074)	(0.093)
Clusters	77	107	136	91
Observations	8760	12300	15600	10440
Bandwidth	750	1000	1250	840

Notes: *** p<0.01, ** p<0.05, * p<0.1. The table shows that police departments differentially reported data to the FBI at the threshold for the 1960 census experiment. The table presents RD estimates on a dummy equal to one if the department did not submit a report for the month for the sample of post-treatment years (1960 to 1969). Variation in treatment status is from the 1960 census experiment. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-month fixed effects are included in all columns.

Appendix Table V: Effect of merit system mandates on alternative definitions of the crime outcomes

Sample	pre-treatment				post-treatment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Property crime rate	-53.104*	-34.156	-14.993	-43.795	-191.402**	-148.522**	-122.177**	-148.522**
	(28.368)	(23.745)	(21.891)	(31.541)	(78.338)	(64.077)	(52.614)	(64.077)
Clusters	80	101	123	57	89	113	137	113
Observations	7096	9108	11119	4822	9106	11576	14128	11576
Bandwidth	750	1000	1250	569	750	1000	1250	1002
Violent crime rate	-16.978	-13.799	-8.349	-13.147	4.666	8.270	9.186	3.139
	(13.864)	(11.305)	(9.897)	(11.238)	(33.055)	(26.591)	(23.719)	(16.793)
Clusters	77	98	119	99	89	113	137	265
Observations	4394	5708	6914	5753	9106	11576	14128	24782
Bandwidth	750	1000	1250	1036	750	1000	1250	2270
Property crimes	-2.081	-1.209	-0.124	-1.763	-10.150**	-8.005**	-7.590**	-7.293**
	(1.342)	(1.101)	(1.039)	(1.422)	(4.082)	(3.515)	(3.042)	(3.100)
Clusters	80	101	123	55	89	113	137	133
Observations	7096	9108	11119	4651	9106	11576	14128	13678
Bandwidth	750	1000	1250	554	750	1000	1250	1182
Violent crimes	-0.738	-0.607	-0.357	-0.607	0.186	0.343	0.373	0.220
	(0.635)	(0.516)	(0.456)	(0.516)	(1.720)	(1.393)	(1.246)	(1.021)
Clusters	77	98	119	98	89	113	137	210
Observations	4394	5708	6914	5708	9106	11576	14128	20140
Bandwidth	750	1000	1250	1001	750	1000	1250	1885
Log(property crimes)	-0.302	-0.197	-0.033	-0.125	-0.626***	-0.530***	-0.490***	-0.696***
	(0.189)	(0.161)	(0.161)	(0.219)	(0.215)	(0.187)	(0.168)	(0.228)
Clusters	80	101	123	60	89	113	137	79
Observations	5715	7302	8790	4220	8891	11215	13589	7948
Bandwidth	750	1000	1250	594	750	1000	1250	690
Log(violent crimes)	-0.242	-0.283	-0.087	-0.258	-0.049	-0.027	0.005	-0.127
Clusters	(0.345)	(0.287)	(0.271)	(0.352)	(0.428)	(0.336)	(0.297)	(0.382)
Observations	67	88	108	56	89	113	137	104
Bandwidth	1059	1325	1624	925	4402	5540	6542	5110
Bandwidth	750	1000	1250	669	750	1000	1250	891

Notes: *** p<0.01, ** p<0.05, * p<0.1. The table shows that results are robust to different ways of defining the crime outcomes. It presents RD estimates on crime rates in levels, crime counts in levels, and crime counts in logs for pre-treatment years (1960 to 1969, columns 1 to 4) and post-treatment years (1970 to 1979, columns 5 to 8). Variation in treatment status is from the 1970 census experiment. Crime rates are crimes per 100,000 people. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-month fixed effects are included in all columns.

Appendix Table VI: Effect of merit system mandates on crime and clearance rates, restricted pre-treatment sample

Sample	pre-treatment			
	(1)	(2)	(3)	(4)
Log(property crime rate)	-0.149	-0.056	0.059	-0.232
	(0.178)	(0.148)	(0.145)	(0.193)
Clusters	76	96	118	55
Observations	4476	5738	6994	3024
Bandwidth	750	1000	1250	557
Log(violent crime rate)	-0.251	-0.307	-0.107	-0.308
	(0.252)	(0.214)	(0.209)	(0.319)
Clusters	60	78	95	33
Observations	577	745	946	335
Bandwidth	750	1000	1250	475
Property crime clearance rate	0.043	0.032	0.036	0.043
	(0.049)	(0.044)	(0.042)	(0.049)
Clusters	76	96	117	76
Observations	3090	4006	4852	3090
Bandwidth	750	1000	1250	752
Violent crime clearance rate	-0.193*	-0.152	-0.142	-0.171
	(0.108)	(0.096)	(0.095)	(0.124)
Clusters	60	78	95	38
Observations	577	745	946	385
Bandwidth	750	1000	1250	558

Notes: *** p<0.01, ** p<0.05, * p<0.1. The tables shows robustness to restricting the sample of pre-treatment years to a sample less likely to have an anticipation effect. It presents RD estimates on crime rates for a restricted sample of pre-treatment years: 1960 to 1969 for states with mandates based on the federal population census only and 1960 to 1967 for states with mandates based on federal, state or municipal census. Variation in treatment status is from the 1970 census experiment. Crime rates are crimes per 100,000 people. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-month fixed effects are included in all columns.

Appendix Table VII: Crime-by-crime effect of merit system mandates on property crime and clearance rates

Sample	pre-treatment				post-treatment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(burglary and vehicle theft rate)	-0.048	-0.025	0.027	0.053	-0.410*	-0.265	-0.220	-0.432**
	(0.168)	(0.140)	(0.126)	(0.206)	(0.218)	(0.181)	(0.158)	(0.206)
Clusters	80	101	123	42	89	113	137	95
Observations	3845	4984	6134	1907	7673	9615	11472	8167
Bandwidth	750	1000	1250	403	750	1000	1250	802
Log(larceny rate)	-0.189	-0.084	0.019	-0.328**	-0.570***	-0.457**	-0.380**	-0.627***
	(0.182)	(0.137)	(0.135)	(0.157)	(0.212)	(0.180)	(0.159)	(0.217)
Clusters	79	100	122	52	89	113	137	76
Observations	4837	6210	7444	3171	8640	10897	13148	7542
Bandwidth	750	1000	1250	538	750	1000	1250	644
Burglary and vehicle theft clearance rate	0.039	0.058	0.065*	0.039	0.055*	0.049**	0.061**	0.041
	(0.042)	(0.039)	(0.036)	(0.042)	(0.029)	(0.025)	(0.023)	(0.029)
Clusters	79	100	121	80	89	113	137	76
Observations	3030	3907	4749	3051	7673	9615	11472	6636
Bandwidth	750	1000	1250	758	750	1000	1250	655
Larceny clearance rate	0.036	0.024	0.031	0.009	-0.003	0.007	0.006	-0.007
	(0.051)	(0.048)	(0.046)	(0.049)	(0.041)	(0.034)	(0.031)	(0.044)
Clusters	78	99	120	46	89	113	137	65
Observations	3722	4825	5743	2204	8640	10897	13148	6459
Bandwidth	750	1000	1250	470	750	1000	1250	572

Notes: *** p<0.01, ** p<0.05, * p<0.1. The tables shows the reduced form effect of merit systems on crime rates by crime type. It presents RD estimates on burglary and larceny crime and clearance rates for pre-treatment years (1960 to 1969, columns 1 to 4) and post-treatment years (1970 to 1979, columns 5 to 8). Variation in treatment status is from the 1970 census experiment. Crime rates are crimes per 100,000 people and clearance rates are number of crimes cleared by arrest over total number of crimes. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-month fixed effects are included in all columns.

Appendix Table VIII: Effect of merit systems on reporting

Sample	pre-treatment				post-treatment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Monthly crime report missing	-0.019	-0.021	-0.029	-0.024	0.043	0.022	-0.001	0.031
	(0.134)	(0.116)	(0.105)	(0.115)	(0.055)	(0.042)	(0.040)	(0.045)
Clusters	90	114	138	115	90	114	138	103
Observations	10800	13680	16560	13800	10560	13380	16260	12120
Bandwidth	750	1000	1250	1031	750	1000	1250	858

Notes: *** p<0.01, ** p<0.05, * p<0.1. The table shows that police departments did not differentially report crime data to the police in the 1970 census experiment. It presents RD estimates on a dummy equal to one if the department did not submit a report for the month for pre-treatment years (1960 to 1969, columns 1 to 4) and post-treatment years (1970 to 1979, columns 5 to 8). Variation in treatment status is from the 1970 census experiment. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-month fixed effects are included in all columns.

Appendix Table IXa: Effect on crime and clearance rates, robustness to data cleaning, population dynamics and specification

Sample	post-treatment							
Specification	Excludes simple assault	Drops outliers	Uses UCR population	Controls for 1980 population	Quasi-Balanced Sample	Includes controls	Estimates DID	SE Clustered Two-way
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(property crime rate)	-	-0.459**	-0.501***	-0.404**	-0.492**	-0.436**	-0.279*	-0.461***
	-	(0.180)	(0.178)	(0.185)	(0.193)	(0.173)	(0.147)	(0.173)
Clusters	-	113	113	113	95	113	113	113
Observations	-	11205	11215	11215	9954	11215	18517	11215
Bandwidth	-	1000	1000	1000	1000	1000	1000	1000
Log(violent crime rate)	0.087	0.021	-0.027	0.023	0.056	0.023	0.114	0.027
	(0.320)	(0.334)	(0.317)	(0.332)	(0.361)	(0.179)	(0.197)	(0.314)
Clusters	113	113	113	113	95	113	112	113
Observations	3780	5528	5540	5540	4805	5540	6864	5540
Bandwidth	1000	1000	1000	1000	1000	1000	1000	1000
Property crime clearance rate	-	0.022	-	0.025	0.014	0.020	-0.004	0.020
	-	(0.030)	-	(0.028)	(0.031)	(0.030)	(0.041)	(0.028)
Clusters	-	113	-	113	95	113	113	113
Observations	-	11064	-	11215	9954	11215	16785	11215
Bandwidth	-	1000	-	1000	1000	1000	1000	1000
Violent crime clearance rate	0.108**	0.118**	-	0.119**	0.138***	0.103**	0.152**	0.125***
	(0.042)	(0.049)	-	(0.046)	(0.051)	(0.048)	(0.069)	(0.044)
Clusters	113	113	-	113	95	113	112	113
Observations	3780	5450	-	5540	4805	5540	6864	5540
Bandwidth	1000	1000	-	1000	1000	1000	1000	1000

Notes: The table shows that the main results are robust to different ways of defining the outcomes, controlling for population dynamics and alternative specifications. It presents RD estimates on crime rates and clearance rates for post-treatment years (1970 to 1979). Variation in treatment status is from the 1970 census experiment. In particular, the results are robust to: (1) excluding simple assault from the definition of violent crimes; (2) dropping outliers; (3) using UCR population to calculate crime rates; (4) controlling for 1980 population; (5) restricting the sample of municipalities reporting at least half of the times; (6) including baseline controls; (7) estimating a DID specification; (8) clustering standard errors at the municipality and county-year level. Crime rates are crimes per 100,000 people and clearance rates are number of crimes cleared by arrest over total number of crimes. All coefficients are estimated using locally linear regression and a uniform kernel for a 1000 bandwidth. Standard errors clustered at the municipality level are shown in parentheses in columns 1 to 7. State-month fixed effects are included in all columns.

Appendix Table IXb: Effect on crime and clearance rates, robustness to overlapping legislation

Sample	post-treatment							
State being excluded	AZ	IL	IA	LA	MT	NE	WI CITY	WI VILL
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(property crime rate)	-0.461**	-0.442**	-0.505***	-0.465**	-0.467**	-0.453**	-0.439**	-0.476***
	(0.180)	(0.225)	(0.191)	(0.182)	(0.182)	(0.192)	(0.214)	(0.184)
Clusters	113	60	101	103	108	105	91	110
Observations	11215	5957	9896	10552	10968	10303	8758	10856
Bandwidth	1000	1000	1000	1000	1000	1000	1000	1000
Log(violent crime rate)	0.027	-0.063	0.032	0.094	0.009	0.025	0.000	0.023
	(0.333)	(0.282)	(0.354)	(0.357)	(0.332)	(0.341)	(0.341)	(0.333)
Clusters	113	60	101	103	108	105	91	110
Observations	5540	1928	5028	5030	5474	5184	5096	5500
Bandwidth	1000	1000	1000	1000	1000	1000	1000	1000
Property crime clearance rate	0.020	0.061	0.020	0.024	0.018	0.008	0.014	0.019
	(0.029)	(0.052)	(0.031)	(0.030)	(0.029)	(0.030)	(0.030)	(0.030)
Clusters	113	60	101	103	108	105	91	110
Observations	11215	5957	9896	10552	10968	10303	8758	10856
Bandwidth	1000	1000	1000	1000	1000	1000	1000	1000
Violent crime clearance rate	0.125***	0.247***	0.120**	0.125**	0.122***	0.118**	0.114**	0.124***
	(0.047)	(0.090)	(0.047)	(0.049)	(0.047)	(0.048)	(0.048)	(0.047)
Clusters	113	60	101	103	108	105	91	110
Observations	5540	1928	5028	5030	5474	5184	5096	5500
Bandwidth	1000	1000	1000	1000	1000	1000	1000	1000

Notes: The table shows that the results are not driven by any single state and thus do not depend on other state-specific laws also changing at the same threshold. The table presents RD estimates on crime and clearance rates for post-treatment years (1970 to 1979), excluding one state at the time. Variation in treatment status is from the 1970 census experiment. Crime rates are crimes per 100,000 people and clearance rates are number of crimes cleared by arrest over total number of crimes. Arizona does not have any municipality in the risk set within the specified bandwidth from the threshold. West Virginia is not shown as there are no municipalities in the risk set within a 3,000 bandwidth from the threshold. The coefficients are estimated using locally linear regression and a uniform kernel for a 1000 bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-month fixed effects are included in all columns.

Appendix Table IXc: Effect on crime and clearance rates, robustness to the estimation

Sample	post-treatment					
Estimation	LLR, Triangular Kernel	LLR, Epanech-nikov Kernel	LQR, Uniform Kernel	LCR, Uniform Kernel	LLR, Uniform Kernel, no FEs	LLR, Uniform Kernel, more flexible running variable
	(1)	(2)	(3)	(4)	(5)	(6)
Log(property crime rate)	-0.558***	-0.548***	-0.640**	-0.451*	-0.422**	-0.462**
	(0.197)	(0.193)	(0.259)	(0.269)	(0.192)	(0.180)
Clusters	113	113	113	113	113	113
Observations	11215	11215	11215	11215	11216	11215
Bandwidth	1000	1000	1000	1000	1000	1000
Log(violent crime rate)	-0.004	-0.024	0.015	0.308	0.071	0.028
	(0.425)	(0.388)	(0.531)	(0.622)	(0.312)	(0.335)
Clusters	113	113	113	113	113	113
Observations	5540	5540	5540	5540	5664	5540
Bandwidth	1000	1000	1000	1000	1000	1000
Property crime clearance rate	0.010	0.013	-0.002	0.002	0.028	0.020
	(0.031)	(0.031)	(0.037)	(0.038)	(0.032)	(0.029)
Clusters	113	113	113	113	113	113
Observations	11215	11215	11215	11215	11216	11215
Bandwidth	1000	1000	1000	1000	1000	1000
Violent crime clearance rate	0.104**	0.108**	0.096	0.098	0.170***	0.119***
	(0.049)	(0.048)	(0.066)	(0.081)	(0.055)	(0.046)
Clusters	113	113	113	113	113	113
Observations	5540	5540	5540	5540	5664	5540
Bandwidth	1000	1000	1000	1000	1000	1000

Notes: *** p<0.01, ** p<0.05, * p<0.1. The table shows robustness to different choices made in the estimation. It presents RD estimates on crime and clearance rates for post-treatment years (1970 to 1979). In particular, column 1 and 2 are estimated using locally linear regression and a triangular kernel and an Epachnikov kernel respectively. They include state-month fixed effects. Column 3 is estimated using locally quadratic regression and a uniform kernel and includes state-month fixed effects. Column 4 is estimated using locally cubic regression and a uniform kernel and also includes state-month fixed effects. Column 5 is estimated using locally linear regression and a uniform kernel but does not include state-month fixed effects. Finally, column 6 is also estimated using locally linear regression and a uniform kernel but allows the running variable to vary by year. Crime rates are crimes per 100,000 people and clearance rates are number of crimes cleared by arrest over total number of crimes. All columns present estimates restricting to a 1000 bandwidth. Standard errors clustered at the municipality level are shown in parentheses.

Appendix Table X: Descriptive statistics for employment and expenditures

Statistics	N	Mean	Sd
Employment per 1,000 people	507	26.645	(25.577)
Police employees per 1,000 people	381	2.681	(1.235)

Notes: This table reports summary statistics (number of observations, mean and standard deviation) for exployment and police employees per 1,000 people for the baseline sample of post-treatment. Post-treatment years are 1970 to 1979 for expenditures and 1972 to 1979 for employment.

Appendix Table XI: Descriptive statistics for police officers 1960-1980

Census year	1960	1970	1980	pooled
	(1)	(2)	(3)	(4)
Panel (a): information on the sample				
Experiment year	1950	1960	1970	1950-1970
Municipalities	132	127	106	365
Police Officers	300	300	250	800
Panel (b): descriptive statistics				
Age	41.350	37.220	34.010	37.730
	(12.74)	(12.47)	(10.41)	(12.34)
Highest grade achieved	11.210	13.390	14.940	13.080
	(2.416)	(2.119)	(1.964)	(2.663)
Finished high school	0.414	0.712	0.738	0.614
	(.493)	(.454)	(.441)	(.487)
Finished two years of college		0.616	0.713	0.455
		(.487)	(.453)	(.498)
Veteran status	0.583	0.434	0.475	0.499
	(.494)	(.497)	(.5)	(.5)

Notes: This table reports descriptive statistics for policemen characteristcs. Each column reporting information for a specific census. The census year reported at the top of the column refers to when the outcomes are measured; variation in treatment status is from the census experiment ten year prior. Panel (a) reports the states in the sample, the number of municipalities, the number of police officers and the number of newly hired police officers. Panel (b) reports mean and standard deviation for the police officers in municipalities in the control groups and within a 3000 population bandwidth.

Appendix Table XII: Effect on demographic composition of police departments, individual level regressions

Sample	post-treatment			
	(1)	(2)	(3)	(4)
Finished high school	-0.088	0.009	-0.002	-0.069
	(0.097)	(0.093)	(0.076)	(0.101)
Clusters	136	176	221	134
Observations	400	500	650	400
Bandwidth	750	1000	1250	716
Age	1.362	-1.594	-0.993	-0.892
	(2.412)	(2.220)	(2.054)	(2.245)
Clusters	136	176	221	191
Observations	400	500	650	550
Bandwidth	750	1000	1250	1105
Veteran	-0.019	0.028	0.054	0.036
	(0.093)	(0.087)	(0.074)	(0.086)
Clusters	136	176	221	186
Observations	400	500	650	550
Bandwidth	750	1000	1250	1081

Notes: *** p<0.01, ** p<0.05, * p<0.1. The table shows the effect of merit systems on the demographic composition of police departments is robust to estimating regressions at the individual level. It presents RD estimates on a dummy for having a high school degree, age and a dummy for having veteran status for post-treatment years. Outcomes are measured in the 1960, 1970 and 1980 census, and variation in treatment assignment is from the 1950 to 1970 census. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-census year fixed effects are included in all columns. Observation numbers are rounded to avoid disclosure.

Appendix Table XIII: Effect of merit systems on turnover and wages

Sample	post-treatment			
	(1)	(2)	(3)	(4)
Fraction new hire	-0.000	0.098	0.088	-
	(0.127)	(0.123)	(0.109)	-
Clusters	119	159	191	-
Observations	99	129	155	-
Bandwidth	750	1000	1250	-
Average wage	0.395	0.628	0.651	0.579
	(0.636)	(0.529)	(0.463)	(0.608)
Clusters	179	236	302	197
Observations	136	176	221	149
Bandwidth	750	1000	1250	823

Notes: *** p<0.01, ** p<0.05, * p<0.1. The tables shows the effect of the merit system mandate on outcomes related to the organization of police departments. The table presents RD estimates on turnover and wages for post-treatment years. The outcomes are fraction of police officers who are certaintly new hires and average wage. Outcomes are measured in the 1960, 1970 and 1980 census, and variation in treatment assignment is from the 1950 to 1970 census. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-census year fixed effects are included in all columns.

Appendix Table XIV: Effect of merit system mandates on reporting post-1980

Sample	post-treatment			
	(1)	(2)	(3)	(4)
Monthly crime report missing	0.019	0.022	0.031	0.022
	(0.029)	(0.031)	(0.030)	(0.033)
Clusters	125	158	195	82
Observations	21120	29640	39000	12600
Bandwidth	750	1000	1250	506

Notes: *** p<0.01, ** p<0.05, * p<0.1. The table shows that police departments did not differentially report crime data to the police in the 1980 census experiment. It presents RD estimates of the effect of merit systems on a dummy equal to one if the department did not submit a report for the month for post-treatment years (1980 to 1989, columns 1 to 4). Variation in treatment status is from the 1980 census experiment. The coefficients are estimated using locally linear regression and a uniform kernel for four different bandwidths: 750, 1000, 1250 and an outcome and sample specific MSE-optimal bandwidth. Standard errors clustered at the municipality level are shown in parentheses. State-month fixed effects are included in all columns.