

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

Arianna Ornaghi*

July 2, 2019

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: crime rates are lower 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 suggestive evidence that the limitations to politicians' ability to influence police officers are instead important.

JEL codes: H83, 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, Mirko Draca, Daniel Fetter, John Firth, Ludovica Gasse, 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 seminar participants at the MIT Political Economy lunch, the Harvard Economic History lunch, Arizona State University, IIES, Bank of Italy, EIEF, University of Warwick, Barcelona GSE Summer Forum, and the PSE Conflict Workshop 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 ensure 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 et al., 2003](#); [Kaufmann et al., 1999](#)) that there is a high degree of 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 through reduced turnover and improved recruitment – 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 in the late 1970s discussed whether dismissals for political reasons violated the First Amendment, the decision was in support of merit systems, but the dissenting opinion of Justice Stewart clearly supported a return to 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 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, pro-

mote and dismiss officers was taken from the mayor and given to a semi-independent civil service commission. Hiring and promotion decisions had to follow the results of formal examinations and dismissals were only permitted for just cause.

The first cities to establish merit systems, Albany, Utica and Yonkers (NY), did so in 1884. 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.

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, defined as crimes per 100,000 people. 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 crime rate was 45% lower in munic-

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

²Author's calculations based on data from [Ostrom et al. \(1977\)](#).

ipalities just above the threshold relative to municipalities just below. The result is not explained by pre-existing differences: there was no discontinuity in the outcome before the introduction of merit systems. The effect was gradual over time, and stabilized after four to five years. Interestingly, the effect was driven by a decline in the property crime rate as opposed to the violent crime rate: it appears that police officers under spoils and merit systems prioritized different types of crimes. Finally, to support the interpretation that merit system mandates improved performance, I show that the clearance rate for violent crimes, defined as crimes cleared by arrest over number of crimes, also experienced an increase, suggesting that police officers under merit systems might have become more effective in their investigations.

I test whether the results depend on the choice of sample, specification and estimation technique. The effect of merit systems on the crime rate is not driven by any of the choices made in the estimation and is also robust to running a series of placebo regressions with randomly assigned thresholds. In addition, I argue that it is improbable that the results are driven by other state-specific policies changing at the same threshold as the results are robust to dropping one state at the time. 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 itself. Protections granted by the mandate were enforceable in court from the moment in which the official census counts were published, which means that limitations to political influence over police officers were in place from when the mandate became binding, independently on whether a civil service commission was instituted in the municipality or not. This makes studying the effect of the mandate meaningful in this setting. Nonetheless, to make the results more credible, I additionally show evidence that the mandates did induce municipalities to fully adopt the reform. Unfortunately no systematic data on merit system adoption exists for the 1970s, but I was able to collect the information from the Municipal Codes of a small sample of municipalities. This allows me to provide suggestive evidence that indeed, the probability of having a full-fledged merit system increased after 1970 in places subject to the mandate. In addition, I show formally that this was the case for the only period in which data on adoption does exist, 1900-1940. Over this time period, merit system mandates increased the probability of introducing a civil service commission by 43 percentage points.

I explore three possible channels that might explain how merit systems had a positive effect on performance: increases in the resources available to police departments, changes in police officers' characteristics, and reduced political influence. 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. In particular, 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 two pieces of evidence supporting this claim. First, I show that states in which the police chief was also under a merit system, and thus arguably where political influence was more strongly curtailed, experienced a larger decline in crime rates. Second, I exploit the fact that at the end of the 1970s two Supreme Court decisions extended protections from dismissals for political reasons to all municipal employees to provide indirect evidence on the role played by this provision. After 1980, treated municipalities still had to create independent civil service commissions, but there was no discontinuity in whether employees were protected from being fired for political reasons. I find that merit system mandates had no effect on crime 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.

This paper contributes to the growing literature on the organization of bureaucracies by providing well-identified causal evidence on the effect of police departments' professionalization on a policy relevant performance measure: crime rates. The main finding that merit system had a positive effect on performance is in line with correlational evidence that professionalized bureaucracies tend to be more effective, such as the cross-country comparisons by [Evans and Rauch \(1999\)](#) and [Rauch and Evans \(2000\)](#). The closest work is that of [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. I add to this study by substantially improving identification and by investigating an outcome that is a direct measure of bureaucratic performance. With respect to the recent literature that focuses on the role played by selection of bureaucrats on performance (e.g. [Dal Bo et al. \(2013\)](#), [Ashraf et al. \(2018\)](#), [Deserranno \(2018\)](#), [Weaver \(2018\)](#), [Voth and Xu \(2019\)](#)), I show that professionalizing the bureaucracy may have a positive effect even when the composition of the bureaucracy itself is not affected. The proposed explanation that merit systems had a positive effect by shielding the bureaucracy from political influence is instead in line with recent evidence documenting the negative effects of political influence on the bureaucracy, including [Akhtari et al. \(2018\)](#), [Colonnelli et al. \(2018\)](#), [Xu \(Forthcoming\)](#). 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 et al., 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 and their robustness are presented in section 5, and evidence on reform adoption in section 6. Section 7 discusses potential mechanisms. Section 7 concludes. Additional tables and details are available in a separate online appendix.³

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.⁴ 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 loyalists ([Freedman, 1994](#)). At the height of the spoils system, wholesale replacement of federal employees was the norm ([United States Civil Service Commission, 1973](#)), 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

³The online appendix is available at the following [link](#).

⁴The 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.

Civil Service Commission and introduced competitive examinations 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 chief 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, and in many smaller cities and towns merit systems were actually 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 merit system mandates by performing a state-by-state search of Historical Statutes and State Session Laws, using secondary sources to locate relevant information (see Online Appendix B for details on how the search was conducted).⁵ As Figure I shows, there are eight states with mandates based on population thresholds, although only six of them effectively contribute to the estimation (Appendix Table II).⁶ While there were differences in the details of the legislation across states, the fundamental features of the reform were the same. When

⁵Sheriff's departments were not subject to the same legislation: the entirety of the paper focuses on municipal police departments only.

⁶Because 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

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 the results of formal 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. The provisions covered all police officers of lower ranks, but were sometimes extended to the police chief. Finally, civil service commissions 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.⁸

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.⁹

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

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.

⁷Police 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.

⁸In 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.

⁹*Elrod v. Burns*, 1976, 427 U.S. 347. *Branti v. Finkel*, 1980, 445 U.S. 518.

grade of patrolman, highlighting a limited role for career incentives. They engaged in patrolling, traffic control and early criminal investigations, with limited specialization, and relied on external support for more complex tasks (Falcone et al., 2002).¹⁰

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 three property crimes (burglary, larceny-theft, and motor vehicle theft) and four violent crimes (murder and negligent manslaughter, rape, robbery, and aggravated assault).¹¹ I use UCRs to study crime rates, defined as crimes per 100,000 people, and clearance rates, defined as number of crimes cleared by arrest over total number of crimes.¹² 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. Limited data on adoption are available for the main period of interest, but I was able to collect the year of introduction of the reform for a small number of municipalities in the sample from these cities' Municipal Codes. In addition, 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.¹³

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.¹⁴ I study total expenditures per 1,000 people and total

¹⁰Author's calculations based on a 1974 survey conducted by Elinor Ostrom (Ostrom et al., 1977). The survey provides information on all police departments in a random sample of standard metropolitan areas.

¹¹I 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.

¹²For 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. A crime is considered cleared by arrest if at least one person has been arrested, charged, and turned over for prosecution (FBI website). 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, although most arrests happen close to the date of the incident. Given that I find large effects on crimes, normalizing by volume is important. Clearance rates have been used as proxy for performance in the economics of crime literature, for example in McCrory (2007).

¹³Previous studies using these data include Tolbert and Zucker (1983) and Rauch (1995).

¹⁴The 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).

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.¹⁵ 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.^{16,17}

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:

$$y_{mt} = \beta \mathbb{1}(dist_m \geq 0) + f(dist_m) + \delta_{st} + \varepsilon_{mt} \text{ for } m \in RS \quad (1)$$

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); $\mathbb{1}(dist_m \geq 0)$ 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

¹⁵The 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.

¹⁶Using 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.

¹⁷I 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.

the procedure suggested by [Calonico et al. \(2014\)](#). 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. It ends in the year of the following census.¹⁸ 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.

To understand how the effect of the mandate changes over time, I estimate the following RD event study specification:

$$y_{mt} = \sum_{\sigma \in \{-5, +10\}} \beta_{\sigma} \mathbb{1}(dist_m \geq 0) \mathbb{1}(t - \tilde{c} = \sigma) + f_t(dist_m) + \delta_{st} + \varepsilon_{mt} \text{ for } m \in RS \quad (2)$$

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); $\mathbb{1}(dist_m \geq 0)$ 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 Panel A](#) presents the [McCrary \(2008\)](#) test for the 1970 census experiment: the density of the running variables shows no discontinuity at the threshold. Supporting the idea that there was no sorting, the figure additionally shows the McCrary test for all census experiments in which treatment was assigned, ranging from 1910 to 2000. Out of the ten census experiments, the McCrary test only barely fails for 1980, in line with statistical error.

¹⁸I 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.

[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., 2018](#)). I provide evidence that it is unlikely that other policies are driving my results in the robustness checks section.

5 Effects of Merit System Mandates 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.¹⁹ Variation in treatment status is from the 1970 census experiment.²⁰

The analysis estimates the effect of merit system mandates. Estimating the effect of the mandates themselves is meaningful in this setting. Protections against hiring and dismissals for political reasons could be challenged in court as soon as the municipality fell under the mandate: mandates effectively restricted political influence even without the creation of a separate commission.²¹ Nonetheless, if we define treatment as adoption of a full-fledged merit system, I estimate intention to treat effects, as there exists limited data on merit system adoption for the 1970s.

¹⁹[Appendix Table II](#) reports descriptive statistics. 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.

²⁰The 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.

²¹Wayne, Nebraska offers an example for this. Even if the municipality had not complied with the mandate by 1973, it had to reinstate a chief of police who had been unfairly dismissed by the mayor after a district court ruled he was entitled to civil service protections even without a commission.

I begin by showing descriptively how the 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 increase more slowly after 1970.

[Figure III panel A](#) shows the visual equivalent of the RD estimates for the crime rate. The panel on the left shows the falsification RD graph estimated on the sample of pre-treatment years (1960 to 1969), while the one on the right shows the main RD graph of interest, estimated on post-treatment years (1970 to 1979).²² Outcomes are defined as log crime rates.²³ 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. State-month fixed effects are partialled out. The RD graphs show that there was no difference in the crime rate at the discontinuity in the pre-treatment sample. However, after the mandate became effective, municipalities just above the threshold had a lower crime rate than those just below.²⁴

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). Panel A shows that there was no difference in the crime rate in the pre-period, but municipalities above the threshold had a lower 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

²²Preliminary 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. The results are unchanged if the post-treatment period is considered as starting in 1971.

²³The log transformation drops observations with 0 crimes, but I show in [Appendix Table V](#) that this does not make a difference: using crime rates, inverse hyperbolic sine of crime rates, crime counts and log of crime counts gives the same results. This is not surprising to the extent that in fewer than 2% of the observations of municipalities within a 1,250 bandwidth in the post period is the monthly crime rate equal to zero. The negative coefficient in the pre-period for the inverse hyperbolic sine specification can potentially be explained by the relatively low data quality in the very first few years of the UCR program. Even extensively cleaning the data, understanding whether a zero observation is a true zero or a missing datum remains a challenging task ([Maltz, 2006](#)). Considering all zero (total crime) observations as missing in the pre-period reduces the size of the coefficient. This is less of an issue in the post-treatment period, where zero observations are significantly fewer and the quality of the data appears to have improved. In addition, the negative coefficient could be explained by anticipation effects (see below for details): excluding years in the pre-treatment period when anticipation is likely also reduced the size of the coefficient.

²⁴Given 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](#): both municipalities above and below the threshold see higher crime rates over the period, but the increase is slower for places above the threshold.

magnitude of the effect is large: looking at the estimates for places within a 1,000 bandwidth from the threshold, the coefficient shows a 45% reduction in crime rates for treated places in the first ten years after the reform was introduced. The median municipality in the control group within a 1,000 bandwidth from the threshold has 10 total crimes per month, which means that municipalities under merit system mandates experienced 4.5 fewer total crimes per month. To put this in perspective, this is equivalent to the department hiring between 2 to 3 additional police officers according to the crime-police elasticities for municipalities of similar size reported in [Mello \(2018\)](#).

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, suggesting that indeed the result must be explained by changed police behavior. Overall, it appears that merit system mandates improved police performance.

Heterogeneous effects over time. To understand how the effect of the mandate changed over time, [Figure III panel B](#) shows the β_σ coefficients from the event study specification together with 95% confidence intervals.²⁵ The graph on the left shows that the effect is gradual over time and is statistically significant starting two to three years after treatment was assigned in 1970. None of the coefficients in the pre-period is statistically significant, but the point estimates are negative, and start becoming larger in magnitude two to three years before treatment. This is potentially concerning, as it may point to pre-existing differences in crime rates. However, it is important to note that the negative coefficients could be explained by early adoption of merit systems, which would not invalidate the design.

I provide evidence that this is indeed the case by estimating the event study separately for states in which early adoption was more or less likely. 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. Reassuringly, the graph on the right shows the presence of 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

²⁵Different 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.

is gradual at first, but remains constant in magnitude after a few years.²⁶

Heterogeneous effects by type of crime. To explore whether police officers under spoils and merit systems prioritize different types of crime, [Table I Panel B](#) estimates the main regression separately for the property and the violent crime rate (see [Appendix Figure II](#) for the respective RD and event study graphs). The effect appears to be entirely driven by the property crime rate: places above the threshold had a similar property crime rate as places below the threshold in the pre-period, but lower property rates after the mandate became effective. Merit systems had no effect on the violent crime rate: there is no discontinuity at the threshold, and the coefficient for being subject to a merit system mandate is never significantly different than zero.²⁷ This can be potentially rationalized by police officers under spoils systems over-prioritizing violent crimes, maybe because they are more relevant for politicians or more likely to be in the news, and merit system mandates inducing police officers to reallocate effort towards activities with positive marginal returns. Given that the decrease in the property crime rate comes at no cost for the violent crime rate, I interpret this result as again supporting evidence that merit system mandates improved police performance.

Reporting bias. For the interpretation of increased police performance to hold, it must be the case that the decline in 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 VIII](#) 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 et al. \(2010\)](#), an offense can be downgraded to a non-index crime or it can be reported as unfounded. The fact

²⁶A 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.

²⁷Violent crimes are a rare event. In the post-treatment period, only 37% of observations correspond to months with a positive violent crime rate, as opposed to 98% of observations corresponding to months with a positive property crime rate. As a result, we might be interested in understanding whether merit systems affect the extensive margin of violent crimes, even if the main analysis shows no effect on the violent crime rate conditional on it being positive. [Appendix Table VII](#) shows the effect of merit system mandates on the probability of reporting at least one violent crime in a month. The coefficient in the post-treatment period is negative, but only statistically significant for smaller thresholds. I interpret this result as confirming that most of the effect was indeed through changes in the property crime rate.

that I find similar effects across crime types is reassuring as not all crimes can be downgraded as easily.²⁸ 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 IX](#) 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.

Effects on clearance rates. 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.²⁹ Given that probability of arrest varies starkly by type of crime - clearance rates for property crimes are around 20% whereas clearance rates for violent crimes are much higher, at around 60% - and merit systems affected crime composition, I estimate effects separately by type of crime. [Table I Panel C](#) shows no difference in the property crime clearance rate, either pre- or post-treatment. There was no difference in the violent crime clearance rate in the pre-treatment sample, but in the post-period the coefficient is positive and statistically significant at the 5% level: the violent crime clearance rate is 0.10 points higher in police departments in municipalities just above the threshold than in those just below (see [Appendix Figure III](#) for the respective RD and event study graphs). Most property crimes never get cleared in the first place, as it is less likely that the victim can identify the suspect, so there is limited chance of improved investigations of property crimes to lead to an arrest: it is not surprising that increased effort by police officers would be reflected in the violent crime clearance rate only. Overall, it appears that the decline in crime was driven by a combination of deterrence, i.e. an increase of expected cost of crime due to police actions, and incapacitation, i.e. a decrease in the number of potential offenders because of increased arrest rates.

Robustness Checks

In this section, I show that my results are robust to a number of potential concerns. [Figure IV](#) shows the coefficient for the dummy for being above the threshold, together with 95% confidence intervals, for different samples, specifications, and estimation techniques. The relevant comparison is whether each coefficient is different than the one estimated using the baseline specification reported at the top of each graph.³⁰

²⁸In 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.

²⁹I interpret clearance rates as a proxy of performance to the extent that they should correlate with police officers' investigative efforts. Unfortunately, it is not possible to link the arrest data to the ultimate outcome in the judicial system for this time period.

³⁰[Appendix Figure IV](#) shows the same graphs separately by type of crime and for clearance rates. The equivalent tables are [Appendix Table Xa](#), [Appendix Table Xb](#) and [Appendix Table Xc](#). 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.

Robustness to data cleaning. To begin with, [Figure IV Panel A](#) focuses on data cleaning. First, it shows that the main result is robust to dropping outliers in the crime data identified using a procedure similar to [Evans and Owens \(2007\)](#), [Chalfin and McCrary \(Forthcoming\)](#) and [Mello \(2018\)](#) (for more details, see Online Appendix C). The result is not driven by especially high - or low - crime observations, that could possibly just be data errors. Second, it shows that using smoothed UCR population as opposed to linearly interpolated population from the census to define crime rates also does not make a difference.

Robustness to using yearly data. The core analysis uses monthly data to exploit all the variation that is available, which is especially important since I am dealing with a relatively small sample and I want to maximize power. However, using yearly crime data does not make a difference: the coefficient estimate is the very similar in magnitude, and still significant at the 5% level.

Robustness to population dynamics. 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 IV Panel A](#) shows that population dynamics do not explain my findings: the main result survives controlling for 1980 population and dropping places that experience a population growth rate from 1970 to 1980 above 20%.

Effect not driven by sample composition. Finally, I show that the main result is not due to changes in the composition of municipalities above and below the threshold. The result is unchanged if I restrict the analysis to a quasi-balanced sample of municipalities reporting crime data at least half of the months, corresponding to 85% of the municipalities and 88% of the observations). Taken together with the fact that the probability of submitting a report was not differential at the threshold, this is reassuring that the results are not entirely driven by a few very low-crime municipalities above the threshold starting reporting after 1970. But, one might still worry about the effect being driven by different municipalities reporting above and below the threshold in the pre- and post-treatment periods. I try to assuage this concern with two specifications. First, the result is robust to controlling for a host of baseline municipality characteristics. Second, I estimate a difference-in-difference specification that includes municipalities fixed effect, thus allowing me to estimate the effect off of changes in crime rates for the same municipality before and after the mandates become binding. The coefficient on the crime rate in this differences-in-differences specification is smaller in magnitude but still significant at the 10% level.³¹

Robustness to two-way clustering Clustering standard errors at the municipality and county-year

³¹This specification is similar to equation (1), but includes municipality fixed effects and allows the flexible controls of the running variable to vary by year. The coefficient reported is from an interaction of an indicator variable for being above the threshold and an indicator variable for being in the post-period. 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.

level to allow for errors to be correlated for places that are close to each other does not make a difference.

Robustness to overlapping legislation. 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 the results of state-by-state search for such policies.³² 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 IV Panel B](#) shows the result of this exercise. The magnitude of the coefficient is stable across samples. The stability of the coefficient 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. \(2018\)](#). 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.

Robustness to estimation. Finally, [Figure IV 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 main result is not affected, although the coefficient is larger in magnitude. Second, I estimate the main specification using locally quadratic regression and locally cubic regression with a uniform kernel. The result is robust to using a locally quadratic regression, but when a cubic polynomial is used the coefficient is not significant, albeit similar in magnitude. In addition, the results are robust to dropping the state-month fixed effects and allowing the running variable to vary flexibly both by census and by outcome year as in the event study specification.³³

Placebo thresholds. As an additional robustness check, I engage in a placebo exercise in which I randomly assign thresholds to states.³⁴ [Figure V](#) shows the distribution of the estimated coeffi-

³²Overlapping legislation was identified by searching for the threshold in historical State Statutes and State Session Laws. More details are available in Online Appendix C.

³³The discussion of the robustness checks has focused on the effect on the crime rate in the post-period, but robustness separately for property and violent crime and clearance rates are also reported in [Appendix Figure IV](#) and [Appendix Table Xa, Xb and Xc](#). Overall, the result on the property crime rate tracks that of the total crime rate, while the result on the violent crime clearance rate is less robust. The result that there are no pre-treatment differences is also robust to the different choices of sample, specifications and estimation, with one exception. 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 than the coefficient estimate for the post-treatment period under the same specification (Online Appendix Tables 6a and 6b). In all other specifications, in the few cases in which the coefficient is significant, it is always so at the 10% level and never for all thresholds.

³⁴In particular, I take 999 random permutations of the relevant thresholds for the 1970 census experiment (15,000,

cients resulting from the placebo regressions. Out of 999 regressions, only 23 display a treatment effect that is more extreme than the one of the baseline specification, which suggests that the result is not just a feature of population or crime dynamics around the relevant population cutoffs.³⁵

6 Merit System Adoption

Contemporaneous evidence. Unfortunately, no systematic data on adoption of merit systems exists for the 1970s. Municipal Codes do sometime report the date in which the reform was adopted, and I was able to collect information for 53 municipalities out of the 139 within a 1250 bandwidth from the threshold.³⁶ The small sample, especially above the threshold, does not allow me to estimate a proper first stage, but I use the data to provide suggestive evidence that mandates were indeed effective at inducing municipalities to adopt merit systems. [Figure V](#) shows rate of adoption by year separately for municipalities above and below the threshold. The figure allows me to make two important points. First, municipalities above the threshold did indeed adopt around the time they fell under the mandate. Given that these were small municipalities, it makes sense that once the department introduced the reform, both current and perspective police officers would have been aware of it, which strengthens the credibility of the main results. Second, the vast majority of municipalities below the threshold did not have a merit system by the end of the decade.³⁷ Given the effect on crime, it might be puzzling that not all municipalities introduced the reform. If politicians prioritize violent crimes, which are not affected by the reform, they might lack the incentive to adopt. Moreover, these reforms implied a costly administrative reorganization, which might further discourage adoption.³⁸

Historical evidence. In addition, thanks to the fact that some states introduced the mandates in 10,000, 8,000, 7,000, 5,500, 5,000, and 4,000) and assign thresholds to states according to these permutations, under the constraint that no state is assigned its real threshold. I keep the initial year of introduction of the reform constant, but I ignore instances in which the threshold was lowered (each state is assigned one threshold only). I then identify the correct risk set, running variable and treatment status to municipalities using the placebo threshold, and re-estimate the baseline specification.

³⁵Alternatively, only 18 regressions give a t-statistics higher than 1.96.

³⁶[Bostashvili and Ujhelyi \(2018\)](#) note that an important empirical challenge to study merit systems is to collect information on reform adoption for a large number of jurisdictions. To the best of my knowledge, the data is not reported in any state or national survey or publication, and the introduction of merit systems was not systematically reported in local newspapers. In addition, in a phone survey of a small sample of these municipalities, municipal officials would almost never be able to provide the information when not reported in the Municipal Code.

³⁷The figure shows a limited anticipation effect in the 1970 census experiment, which can be explained by municipalities for which I have adoption information mainly being from states with mandates based on the federal census. Splitting the sample of municipalities below the threshold into those that fell under the mandate in 1980 and those that did not, as in [Appendix Figure V](#), shows evidence of an anticipation effect for the 1980 census experiment, for which I have information for municipalities in all states.

³⁸This stance is clearly exemplified in the answer given by Eugene Smith, a candidate for the 1975 mayoral election of Elwood, Indiana, when asked whether he was planning to introduce a merit system for the fire and police departments: "Merit systems for police and firemen for the small number of employed on our forces would tend to cause more dissension rather than improved performance."

the first half of the 20th century, I provide evidence that the legislation was effective at inducing municipalities to adopt merit systems for the one period for which systematic data on adoption does exist, 1900-1940. 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.³⁹

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 percentage points 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 VI panel B](#) shows that the effect 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. Taken together, these two pieces of evidence suggest that merit system mandates were indeed effective at inducing municipalities to adopt.⁴⁰

7 Mechanisms

In this section, I explore three potential mechanisms that may explain why merit system mandates improved the performance of police departments: increased resources, changes in police officers' characteristics, and reduced political influence.⁴¹

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

³⁹When 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.

⁴⁰I 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. This is both because the majority of the sample is composed of municipalities from states in which the mandates are explicitly based on the federal population, and because 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 historical sample.

⁴¹I present these results in table form, but the equivalent RD graphs can be found in the Appendix.

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, as we would expect them to be reflected in payroll expenditures.⁴² 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.⁴³ I study whether selection was affected by testing for discontinuities in the demographic composition of police departments using 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.⁴⁴

[Table IV](#) shows that places with and without a merit system had 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. Moreover, 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

⁴²It 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.

⁴³Historically, the shift from a spoils to a merit system implied the introduction of formal testing procedures. By the 1970s, it is likely that both municipalities with and without a merit system had in place procedures to screen potential police officers ([Leonard, 1970](#)), but only police departments under merit systems were bound to the results of examinations. 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](#)), but the extent to which these exams are able to select and promote adequately police officers is still debated today.

⁴⁴Almost 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.

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. This interpretation is also consistent with the time pattern of the effect highlighted by the event study graphs in [Figure III Panel B](#): had the effect mainly been driven by changes in who police officers were, we would expect them to take a longer time to appear.

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 can provide two pieces of suggestive evidence supporting this mechanism.⁴⁵

Heterogeneous effects by whether chief is covered by merit system. If limitations to political influence are important, we should expect stronger effects if the chief of police is extended merit systems' protections as well as lower ranked officers, as is the case in half of the states in the sample.⁴⁶ I explore heterogeneous effects along this dimension by interacting the dummy for being above the threshold with a dummy for the states that have the provision, and show the respective coefficients in Figure V. The graph shows that indeed, the decline in crime rates was stronger in those states where the chief was also protected, although the coefficients are not statistically different from each other at conventional levels.

Effect of merit systems post-1980. 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.⁴⁷

⁴⁵Ideally, I would like to test directly how local politics influence police activity in the pre-period, and whether this changes after merit systems are introduced. Unfortunately, no data on 1960s and 1970s local elections is available for such small municipalities.

⁴⁶In Arizona, Iowa, Louisiana and West Virginia, the police chief did not receive protections. In Illinois, the chief received protection by default, but the provision could be changed by ordinance. Collecting data on whether the chief was indeed covered from the current Municipal Codes revealed that almost all municipalities waived the protections to the chief by ordinance. In all other states the chief was under a full merit system.

⁴⁷It is important to note that the analysis presented in this section hinges on the assumption that no other reform

Table V shows the effect of merit systems on performance for the 1980 census experiment. There is no discontinuity at the threshold in the crime rate: merit systems appear to have no effect when they do not imply a discontinuity in protections from dismissals for political reasons.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.

Discussion. 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 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, and recent work on Ghanaian civil servants by Rasul et al. (2019). In addition, by limiting dismissals, merit systems may decrease turnover and reduced disruption may have a positive effect on performance.⁴⁹ Finally, merit systems may also change the organizational culture of the department.

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.

⁴⁸The coefficient is negative for the MSE-optimal bandwidth, but visual inspection of the corresponding RD graph reported in Appendix Figure IX suggests that this is driven by places right below the threshold having an especially high property crime rate.

⁴⁹I 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, such as police officers who lived in another state, were in the armed forces or attended college five years before each census was taken. Appendix Table XIV shows no effect on turnover. Unfortunately, this is a rather poor proxy for turnover, which makes it hard to draw conclusive statements along this dimension. 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.

8 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, crime rates were 46% lower 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, 2018](#)).

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, [Vanden Eynde et al., 2017](#)), 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

- Akhtari, Mitra, Moreira, Diana and Trucco, Laura Carolina.** 2018, Political Turnover, Bureaucratic Turnover, and the Quality of Public Services. Working Paper.
- Ashraf, Nava, Bandiera, Oriana and Lee, Scott S.** 2018, Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services. Working paper.
- Banfield, Edward C. and Wilson, James Q.** 1963, *City Politics*, Harvard University Press and The M.I.T. Press.
- Bartalotti, Otavio and Brummet, Quentin.** 2016, Regression Discontinuity Designs with Clustered Data: Mean Square Error and Bandwidth Choice, in **Matias D. Cattaneo and Juan C. Escanciano.**, eds, 'Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)', Emerald Group Publishing.
- Bostashvili, David and Ujhelyi, Gergely.** 2018, Political Budget Cycles and the Civil Service: Evidence from Highway Spending in US States. Working Paper.
- Calonico, Sebastian, Cattaneo, Matias D. and Titiunik, Rocio.** 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 McCrary, Justin.** Forthcoming. 'Are US Cities Under-Policed? Theory and Evidence', *Review of Economics and Statistics* .
- Chong, Alberto, La Porta, Rafael, Lopez-de Silanes, Florencio and Shleifer, Andrei.** 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'.
- Civil Service Assembly of the United States and Canada.** 1940, 'Civil Service Agencies in the United States: A 1940 Census'.
- Civil Service Assembly of the United States and Canada.** 1943, 'Civil Service Agencies in the United States: A 1943 Supplement'.
- Colonnelli, Emanuele, Prem, Mounu and Teso, Edoardo.** 2018, Patronage and Selection in Public Sector Organizations. Working Paper.
- Dal Bo, Ernesto, Finan, Frederico and Rossi, Martin A.** 2013. 'Strengthening State Capabilities: the Role of Financial Incentives in the Call to Public Service', *Quarterly Journal of Economics* 128(3), 1169–1218.
- Deserranno, Erika.** 2018. 'Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda', *American Economic Journal: Applied Economics* .
- Eggers, Andrew C., Freier, Ronny, Grembi, Veronica and Nannicini, Tommaso.** 2018. 'Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions', *American Journal of Political Science* .
- Evans, Peter and Rauch, James E.** 1999. 'Bureaucracy and Growth: A Cross-national Analysis

of the Effects of" Weberian" State structures on Economic Growth', *American Sociological Review* pp. 748–765.

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

Falcone, David N, Wells, L Edward and Weisheit, Ralph A. 2002. 'The Small-Town Police Department', *Policing: An International Journal of Police Strategies & Management* 25(2), 371–384.

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

Folke, Olle, Hirano, Shigeo and Snyder, James M. 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 Imbens, Guido. 2016, Why High-order Polynomials should not be used in Regression Discontinuity Designs. NBER Working Paper 19649.

Hinnerich, Björn Tyrefors and Pettersson-Lidbom, Per. 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, Court, Julius and Mease, Ken. 2003. 'The Bureaucracy and Governance in 16 Developing Countries'. Overseas Development Institute, World Governance Survey Discussion Paper 7.

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

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

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

Lee, David S. and Lemieux, Thomas. 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.
- McCrary, Justin.** 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. Working Paper.
- Mosher, Clayton J., Miethe, Terance D. and Hart, Timothy C.** 2010, *The Mismeasure of Crime*, Sage Publications.
- Oliveros, Virginia and Schuster, Christian.** 2016, Merit, Tenure, and Bureaucratic Behavior: Evidence from a Conjoint Experiment in the Dominican Republic. Working Paper.
- Ostrom, Elinor, Parks, Roger B and Whitaker, Gordon P.** 1977, *Policing Metropolitan America.*, Superintendent of Documents, U.S. Govt. Printing Office, Washington, D.C. 20402.
- Rasul, Imran and Rogger, Daniel.** 2018. ‘Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service’, *Economic Journal* 128(608), 413–446.
- Rasul, Imran, Rogger, Daniel and Williams, Martin J.** 2019, Management and Bureaucratic Effectiveness: Evidence from the Ghanaian Civil Service. Working Paper.
- 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 Evans, Peter B.** 2000. ‘Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries’, *Journal of Public Economics* 75(1), 49–71.
- Rawson, Zivile A.,** ed. 1980, *How to Pass Civil Service Examinations, Patrolman.*, Civil Service Publishing Corporation, Brooklyn.
- Ruggles, Steven, Genadek, Katie, Goeken, Ronald, Grover, Josiah and Sobek, Matthew.** 2015, ‘Integrated Public Use Microdata Series: Version 6.0. [Machine-readable database]’.
- Tolbert, Pamela and Zucker, Lynne.** 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*b*, ‘Annual Survey of State and Local Government Finances and Census of Governments’.
- U.S. Census Bureau.** 1972-1980*a*, ‘Annual Survey of State and Local Government Employment and Census of Governments’.
- U.S. Civil Service Commission.** 1884, ‘Annual Reports’.
- Vanden Eynde, Oliver, Kuhn, Patrick M and Moradi, Alexander.** 2017. ‘Trickle-Down ethnic politics: drunk and absent in the Kenya police force (1957-1970)’, *American Economic Journal: Economic Policy*.
- Voth, Joachim and Xu, Guo.** 2019, Patronage for Productivity: Selection and Performance in the

Age of Sail. Working Paper.

Weaver, Jeff. 2018, Jobs for Sale: Bribery and Misallocation in Hiring. Working Paper.

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.

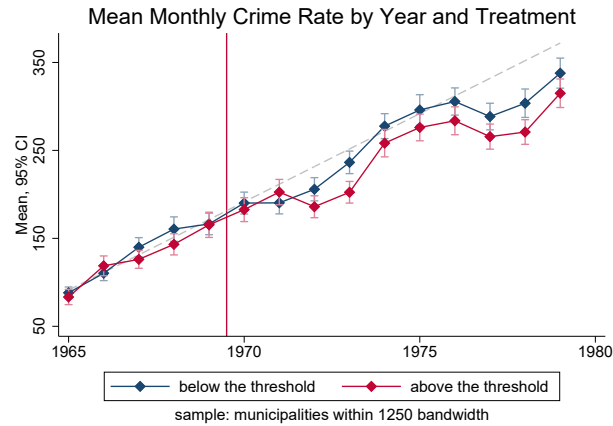
Xu, Guo. Forthcoming. 'The Costs of Patronage: Evidence from the British Empire', *American Economic Review* .

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 state, 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. The Online Appendix reports more details on the legislation in each state.

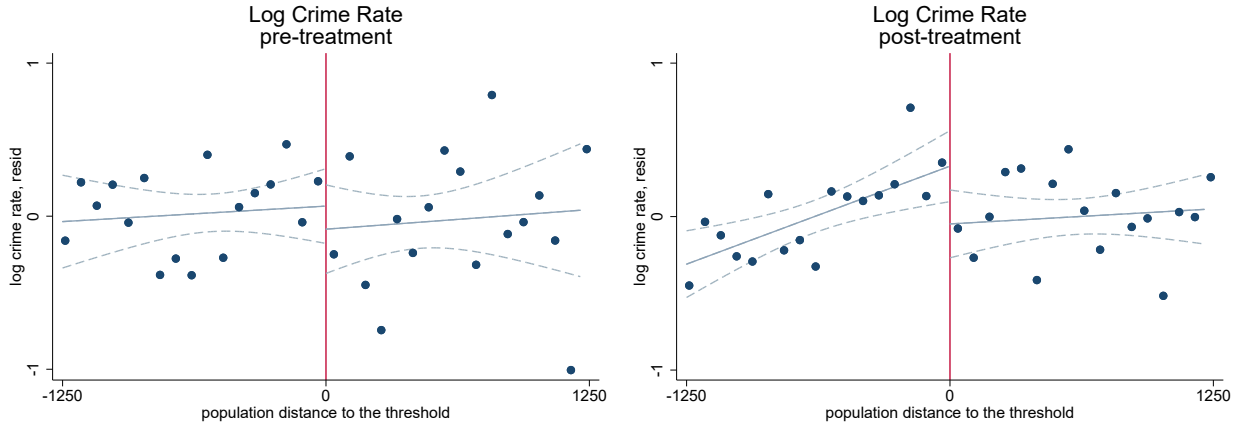
Figure II: Crime Rates Grow at Slower Pace in Municipalities under Mandates



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.

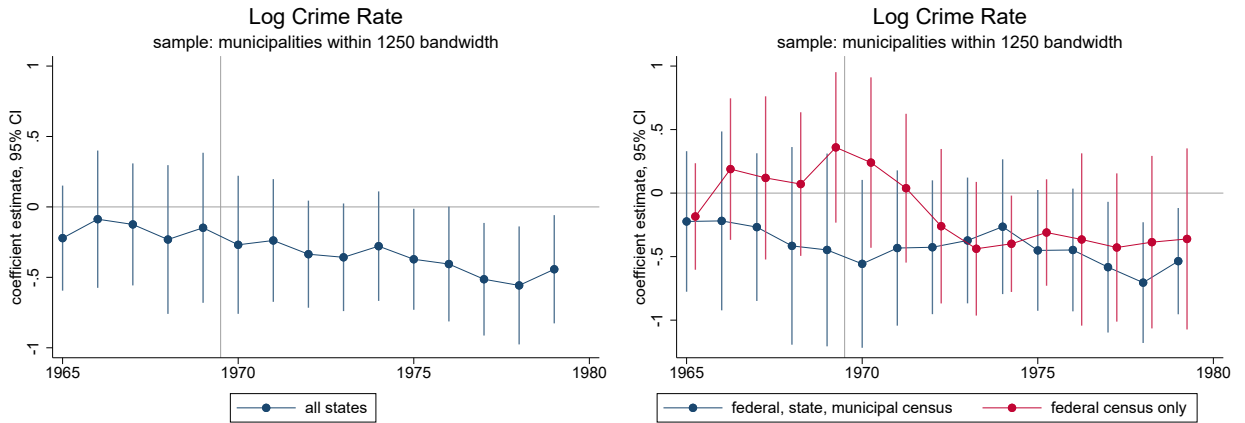
Figure III: Merit Systems Departments Have Lower Crime Rates

Panel A: RD Graphs



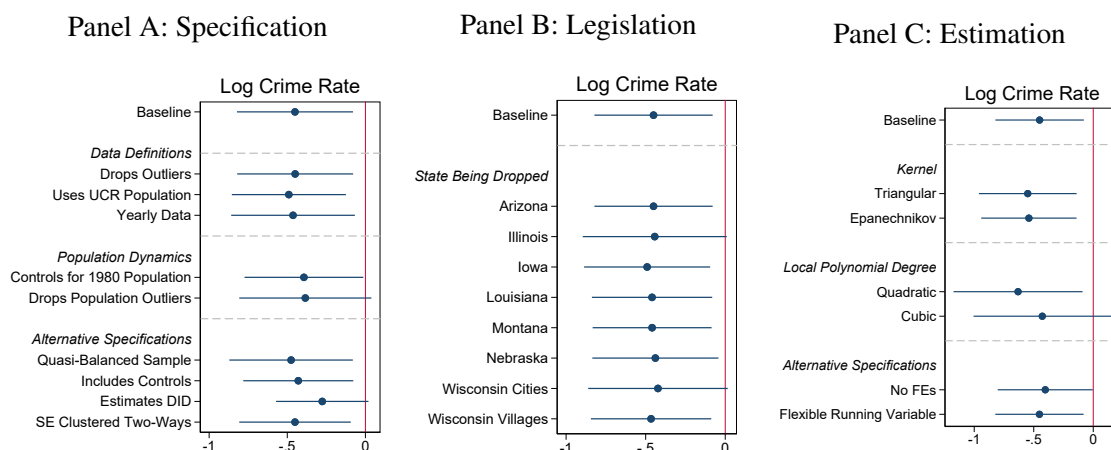
Notes: the graphs show the effect of merit system mandates on monthly crime rates for pre-treatment years (1960 to 1969, on the left) and post-treatment years (1970 to 1979, on the right). Merit systems are mandated for municipalities above the threshold in 1970. Crime rates are crimes per 100,000 people. 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



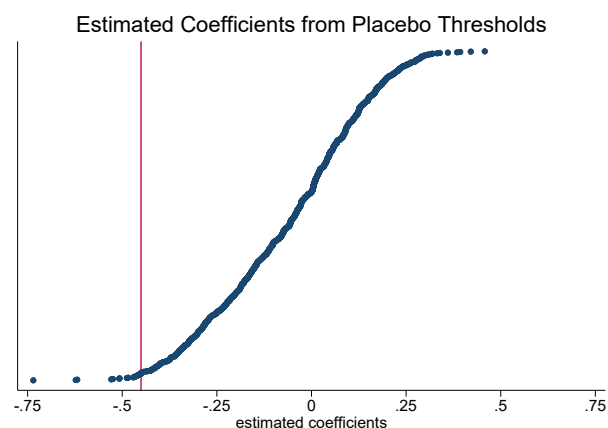
Notes: the graphs show the effect of merit system mandates estimated using the event study specification (equation (2)) on monthly crime rates 1965 to 1979. The graph on the left shows the estimates for the full sample of states. The graph on the right shows separate estimates for states with and without mandates explicitly based on federal population census. Merit systems are mandated for municipalities above the threshold in 1970. 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: Robustness Checks



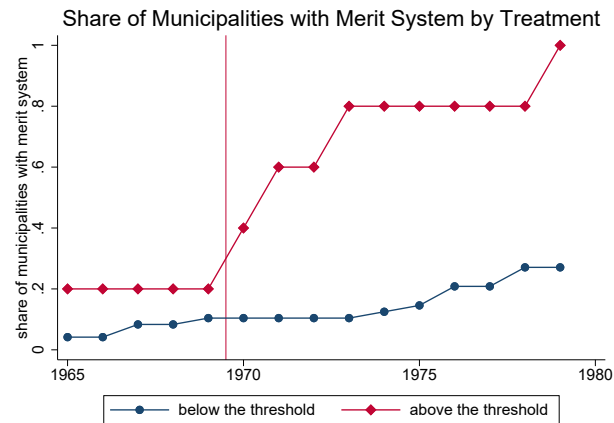
Notes: the graphs show the robustness of the main results. Panel A shows robustness to different ways of defining the outcomes, population dynamics and alternative specifications. Panel B shows that the results are robust to dropping one state at a time. Panel C shows that the results are robust to using different estimation techniques. The graphs report RD estimates on total crime rates, 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 unless otherwise specified. Standard errors are clustered at the municipality level, and state-month fixed effects are included unless otherwise specified.

Figure V: Placebo Thresholds



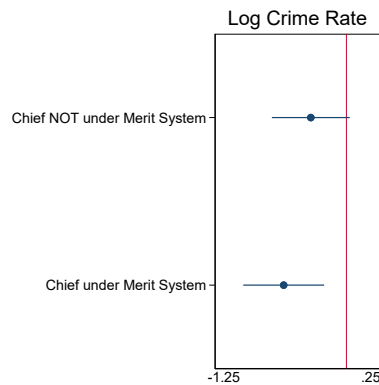
Notes: the graph shows the cumulative distribution of the coefficient for being above the thresholds from 999 placebo regressions in which states are assigned one of the other states' threshold. The sample is restricted to municipalities within a 1000 distance from the threshold. Out of 999 regressions, only 23 regressions display a treatment effect more extreme than the one of the baseline specification (24 regressions estimate a larger coefficient in absolute value). The vertical line shows the coefficient estimated with the baseline specification.

Figure VI: Merit System Mandates Increase Probability of Adoption



Notes: the graph shows the share of municipalities under a merit system by year separately for municipalities above the threshold in the 1970 census experiment (treated in 1970) and below the threshold in the 1970 census experiment (treated after 1970), for the 1965-1979 period. The sample is restricted to municipalities within a 1250 distance from the threshold. Information on merit system adoption is available for 5 municipalities above the threshold and 48 municipalities below the threshold.

Figure VII: Heterogeneous Effects by whether the Chief under a Merit System



Notes: the graph shows the coefficient estimate for being above the threshold separately for municipalities in states where the police chiefs were also extended merit systems' protections and those in which they were not, together with 95% confidence intervals.

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) |
| Panel A: Effect on Total Crimes | | | | | | | | |
| Log Crime Rate | -0.308 (0.198) | -0.194 (0.169) | -0.055 (0.163) | -0.226 (0.231) | -0.575*** (0.222) | -0.451** (0.187) | -0.383** (0.167) | -0.625*** (0.233) |
| Clusters | 80 | 101 | 123 | 62 | 89 | 113 | 137 | 76 |
| Observations | 5811 | 7413 | 8929 | 4437 | 8906 | 11242 | 13623 | 7730 |
| Bandwidth | 750 | 1000 | 1250 | 623 | 750 | 1000 | 1250 | 661 |
| Control Mean | 4.484 | 4.476 | 4.512 | 4.570 | 5.410 | 5.356 | 5.313 | 5.453 |
| Panel B: Heterogeneous Effects by Type of Crime | | | | | | | | |
| Log Property Crime Rate | -0.293 (0.189) | -0.179 (0.162) | -0.034 (0.157) | -0.098 (0.232) | -0.587*** (0.213) | -0.461** (0.180) | -0.394** (0.160) | -0.628*** (0.222) |
| Clusters | 80 | 101 | 123 | 59 | 89 | 113 | 137 | 77 |
| Observations | 5715 | 7302 | 8790 | 4113 | 8891 | 11215 | 13589 | 7822 |
| Bandwidth | 750 | 1000 | 1250 | 583 | 750 | 1000 | 1250 | 666 |
| Control Mean | 4.435 | 4.433 | 4.473 | 4.554 | 5.363 | 5.312 | 5.271 | 5.408 |
| Log Violent Crime Rate | 0.117 (0.334) | 0.028 (0.294) | 0.071 (0.263) | 0.127 (0.345) | 0.081 (0.395) | 0.087 (0.320) | 0.115 (0.293) | 0.093 (0.409) |
| Clusters | 65 | 88 | 109 | 59 | 89 | 113 | 137 | 79 |
| Observations | 950 | 1171 | 1454 | 858 | 3036 | 3780 | 4407 | 2750 |
| Bandwidth | 750 | 1000 | 1250 | 698 | 750 | 1000 | 1250 | 687 |
| Control Mean | 3.516 | 3.473 | 3.433 | 3.539 | 3.467 | 3.445 | 3.467 | 3.486 |
| Panel C: Effect on Clearance Rates | | | | | | | | |
| Property Crime Clearance Rate | 0.029 (0.043) | 0.025 (0.041) | 0.029 (0.040) | 0.040 (0.044) | 0.015 (0.036) | 0.028 (0.031) | 0.030 (0.028) | 0.005 (0.037) |
| Clusters | 80 | 101 | 122 | 55 | 89 | 113 | 137 | 76 |
| Observations | 4329 | 5570 | 6648 | 2940 | 8891 | 11215 | 13589 | 7719 |
| Bandwidth | 750 | 1000 | 1250 | 546 | 750 | 1000 | 1250 | 662 |
| Control Mean | 0.218 | 0.210 | 0.214 | 0.233 | 0.192 | 0.193 | 0.198 | 0.191 |
| Violent Crime Clearance Rate | -0.096 (0.098) | -0.078 (0.090) | -0.067 (0.082) | -0.143 (0.106) | 0.109** (0.047) | 0.106** (0.042) | 0.089** (0.042) | 0.055 (0.052) |
| Clusters | 62 | 84 | 105 | 46 | 89 | 113 | 137 | 59 |
| Observations | 830 | 1039 | 1280 | 637 | 3036 | 3780 | 4407 | 2127 |
| Bandwidth | 750 | 1000 | 1250 | 615 | 750 | 1000 | 1250 | 542 |
| Control Mean | 0.647 | 0.648 | 0.617 | 0.649 | 0.601 | 0.619 | 0.615 | 0.568 |

Notes: The table shows the effect of merit system mandates on police performance. It presents RD estimates 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. Panel A present estimates on total crime rates. Panel B presents estimates separately for property and violent crime rates. Panel C present estimates on clearance rates. 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. The control mean is the mean of the outcome variable for all municipalities below the threshold within the respective bandwidth.

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 (0.151) | 0.096 (0.159) | 0.183 (0.138) | 0.190 (0.183) | 0.334** (0.168) | 0.430** (0.177) | 0.437** (0.171) | 0.337* (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 |
| Control Mean | 0.018 | 0.013 | 0.010 | 0.020 | 0.039 | 0.043 | 0.034 | 0.047 |

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. The control mean is the mean of the outcome variable for all municipalities below the threshold within the respective bandwidth.

Table III: Effect of Merit System Mandates on Expenditures and Employment

| Sample | post-treatment | | | |
|---------------------------------|----------------|---------|---------|---------|
| | (1) | (2) | (3) | (4) |
| Log Expenditures per 1,000 | -0.030 | 0.131 | -0.034 | 0.020 |
| People | (0.208) | (0.186) | (0.163) | (0.202) |
| Clusters | 89 | 113 | 137 | 95 |
| Observations | 492 | 632 | 753 | 531 |
| Bandwidth | 750 | 1000 | 1250 | 805 |
| Control Mean | 3.049 | 3.076 | 3.067 | 3.055 |
| 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 |
| Control Mean | 0.909 | 0.906 | 0.895 | 0.907 |

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. The control mean is the mean of the outcome variable for all municipalities below the threshold within the respective bandwidth.

Table IV: Effect of Merit System Mandates on Demographic Composition of Police Departments

| Sample | post-treatment | | | |
|------------------------|-------------------|-------------------|------------------|-------------------|
| | (1) | (2) | (3) | (4) |
| Share with High School | -0.120 (0.105) | 0.028 (0.098) | 0.026 (0.082) | -0.106 (0.110) |
| Clusters | 136 | 176 | 221 | 127 |
| Observations | 179 | 236 | 302 | 159 |
| Bandwidth | 750 | 1000 | 1250 | 650 |
| Control Mean | 0.614 | 0.614 | 0.614 | 0.614 |
| Average Age | 3.345 (3.059) | -0.323 (2.685) | 0.203 (2.374) | 3.326 (3.376) |
| Clusters | 136 | 176 | 221 | 122 |
| Observations | 179 | 236 | 302 | 152 |
| Bandwidth | 750 | 1000 | 1250 | 622 |
| Control Mean | 37.730 | 37.730 | 37.730 | 37.730 |
| Share Veteran | -0.014 (0.112) | 0.001 (0.104) | 0.015 (0.091) | 0.007 (0.100) |
| Clusters | 136 | 176 | 221 | 189 |
| Observations | 179 | 236 | 302 | 254 |
| Bandwidth | 750 | 1000 | 1250 | 1092 |
| Control Mean | 0.499 | 0.499 | 0.499 | 0.499 |

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. The control mean is the mean of the outcome variable for all municipalities below the threshold within a 3000 bandwidth.

Table V: Effect of Merit System Mandates on Crime Post-1980

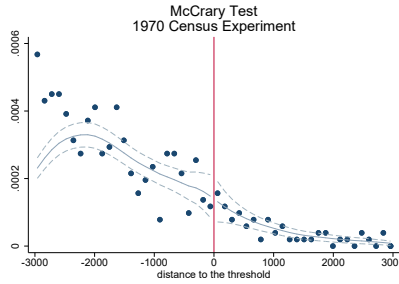
| Sample | post-treatment | | | |
|----------------|-------------------|-------------------|-------------------|----------------------|
| | (1) | (2) | (3) | (4) |
| Log Crime Rate | -0.150 (0.211) | -0.006 (0.193) | -0.023 (0.150) | -1.296*** (0.367) |
| Clusters | 74 | 102 | 127 | 22 |
| Observations | 8375 | 11487 | 14140 | 2475 |
| Bandwidth | 750 | 1000 | 1250 | 266 |
| Control Mean | 5.422 | 5.460 | 5.435 | 5.431 |

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 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. 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. The control mean is the mean of the outcome variable for all municipalities below the threshold within the respective bandwidth.

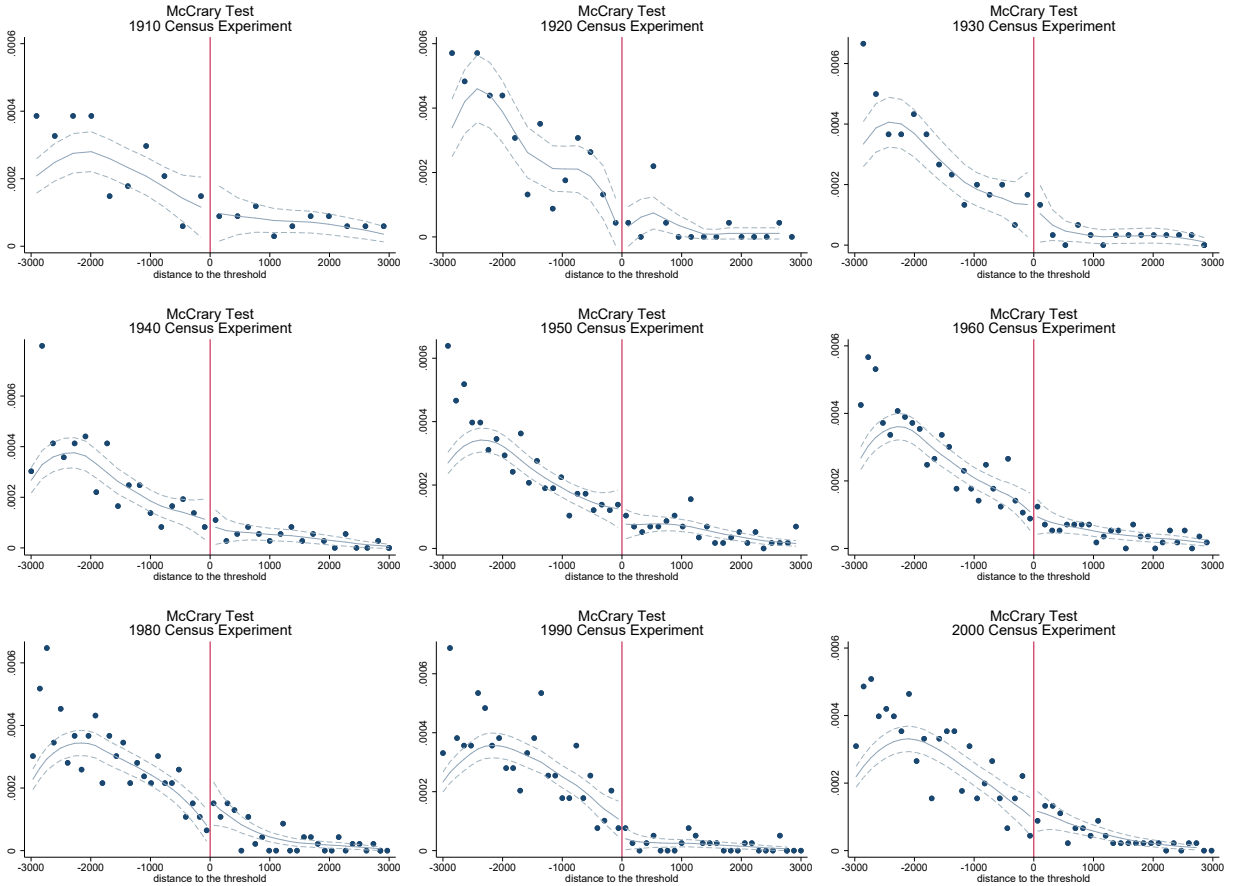
For Online Publication Only

Appendix Figure I: McCrary Tests 1910 to 2000

Panel A: 1970 Census Experiment



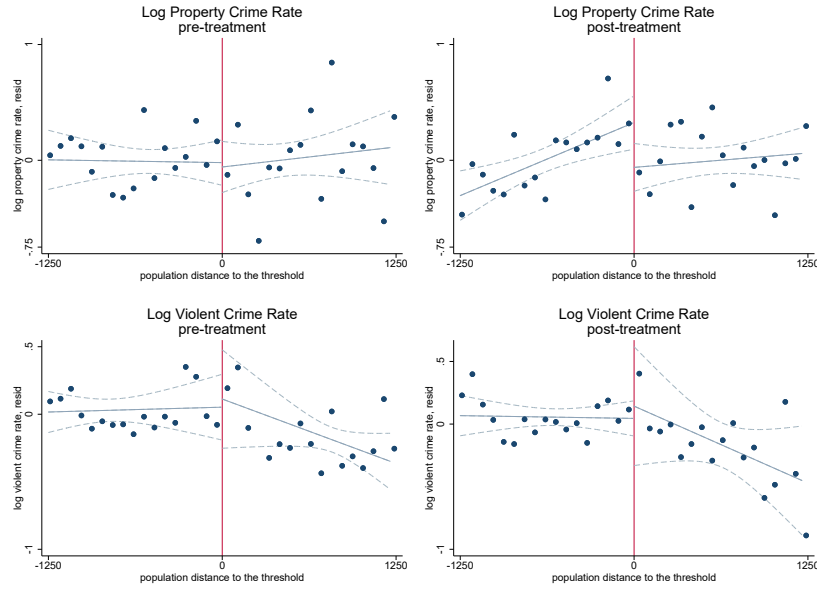
Panel B: All Other Experiments



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

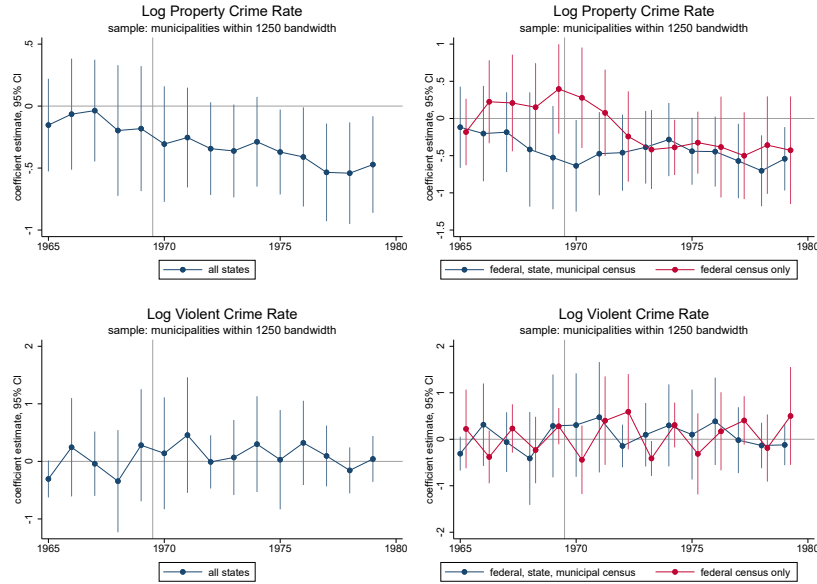
Appendix Figure II: Merit System Mandates Decrease the Property Crime Clearance Rates

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.

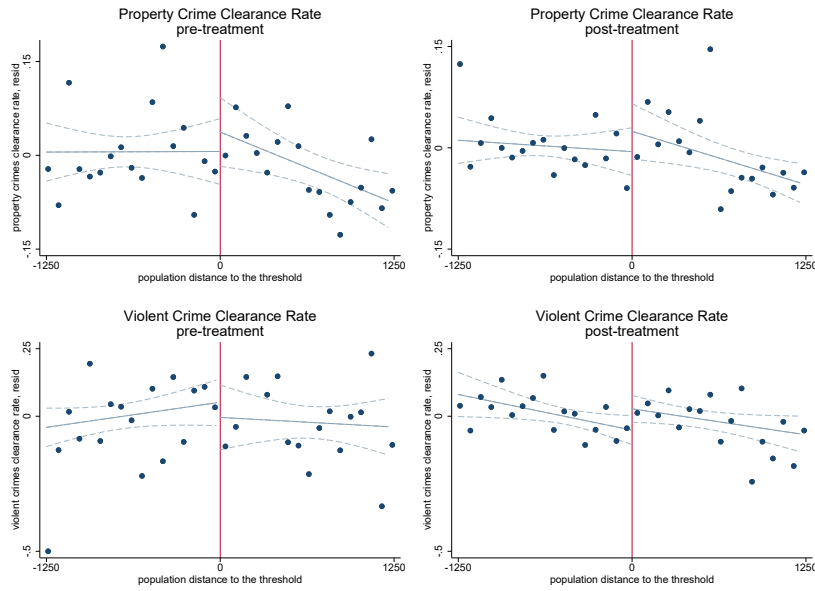
Panel B: Event Study Graphs



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 graphs on the left show the estimates for the full sample of states. The graph on the right show separate estimates for states with and without mandates explicitly based on federal population census. Merit systems are mandated for municipalities above the threshold in 1970. 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.

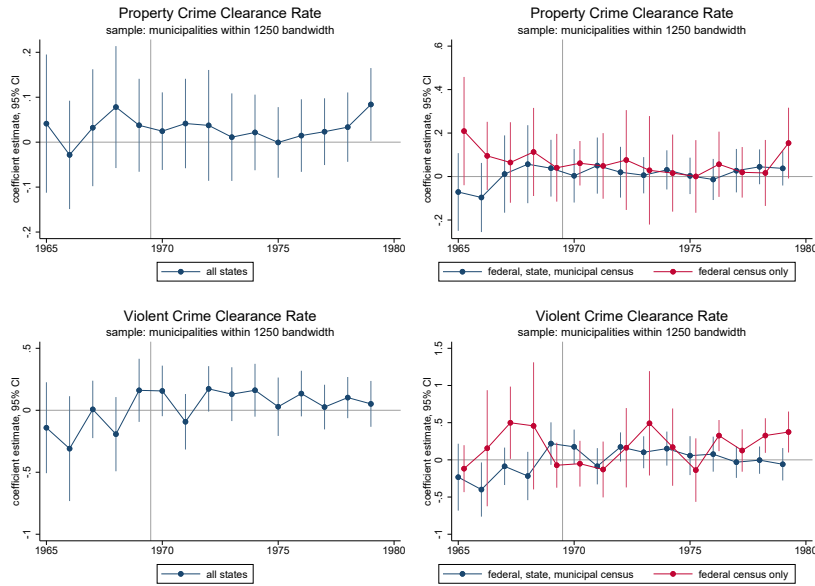
Appendix Figure III: Merit System Mandates Increase the Violent Crime Clearance Rates

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.

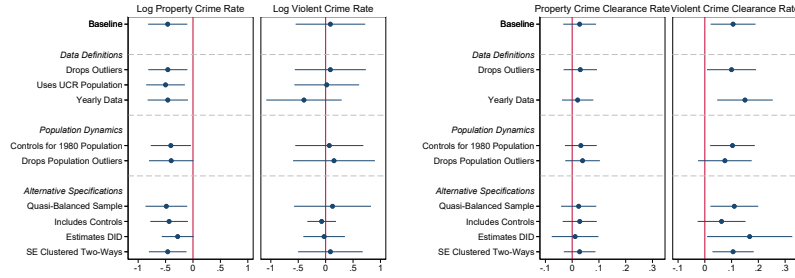
Panel B: Event Study Graphs



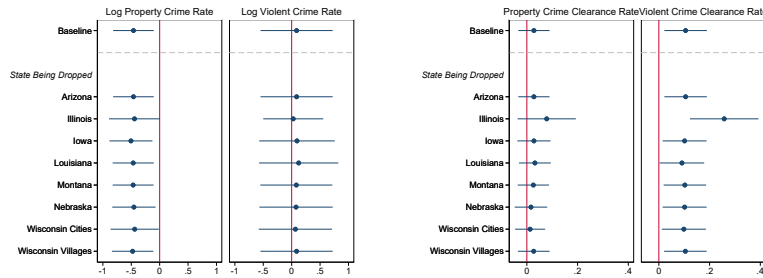
Notes: the graphs show the effect of merit system mandates estimated using the event study specification (equation (2)) on monthly property and violent crime clearance rates for the full sample of states 1965 to 1979. The graphs on the left show the estimates for the full sample of states. The graph on the right show separate estimates for states with and without mandates explicitly based on federal population census. 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 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: Robustness Checks for Property and Violent Crime and Clearance Rates

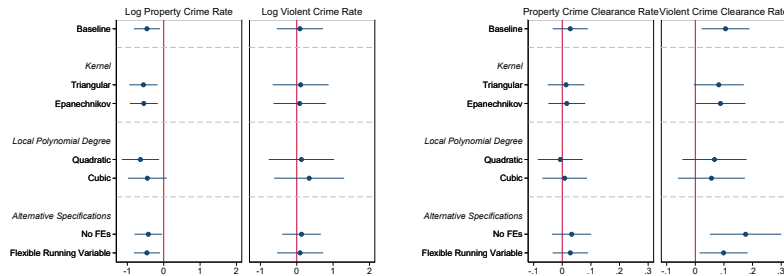
Panel A: Robustness to Data Cleaning, Population Dynamics and Specification



Panel B: Robustness to Overlapping Legislation

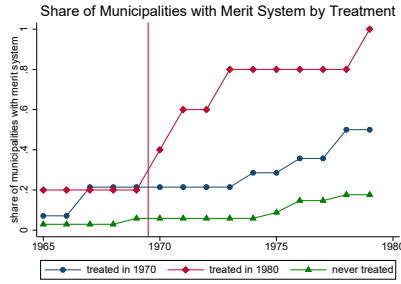


Panel C: Robustness to Estimation



Notes: the graphs show robustness of the main results. Panel A shows robustness to different ways of defining the outcomes, population dynamics and alternative specifications. Panel B shows that the results are robust to dropping one state at a time. Panel C shows that the results are robust to using different estimation techniques. 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.

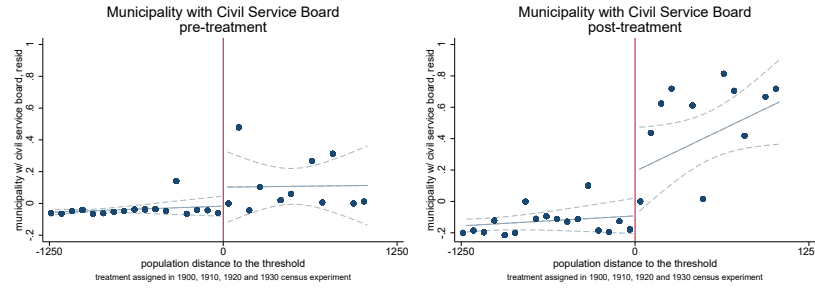
Appendix Figure V: Merit System Mandates Increased Probability of Adoption, by Group



Notes: the graph shows the share of municipalities under a merit system by year separately for municipalities above threshold in the 1970 census experiment (treated in 1970), below the threshold in the 1970 census experiment but above the threshold in the 1980 census experiment (treated in 1980), and below the threshold both in the 1970 and in the 1980 census experiment, for the 1965-1979 period. The sample is restricted to municipalities within a 1250 distance from the threshold. Information on merit system adoption is available for 5 municipalities above the threshold and 48 municipalities below the threshold.

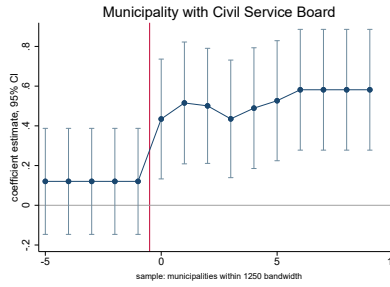
Appendix Figure VI: Merit System Mandates Increased Reform Adoption pre-1940

Panel A: RD graphs



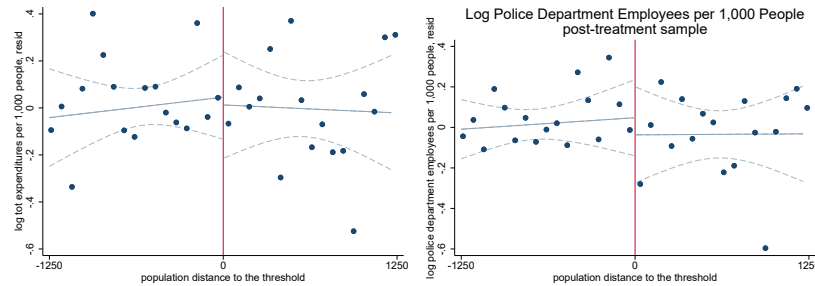
Notes: the graphs show the effect of merit system mandates on pre-1940 reform adoption for the sample of pre-treatment years (on the left) and post-treatment years (on the right). Merit systems are mandated for places above the threshold. The sample exploits variation in treatment status from the 1900, 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



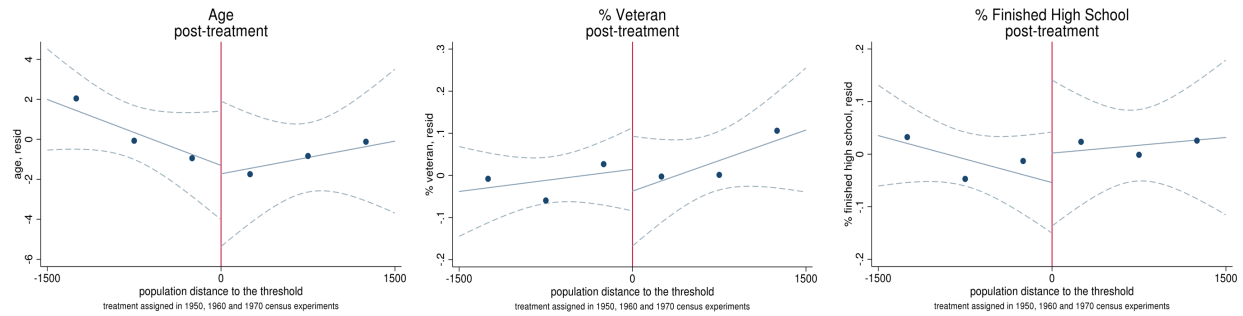
Notes: the graph shows the effect of merit system mandates on pre-1940 reform adoption estimated using the event study specification (equation (2)). The sample exploits variation in treatment status from the 1900, 1910, 1920 and 1930 census experiments. The sample includes both pre-treatment and post-treatment 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_{τ} 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 VII: Merit Systems do not Affect Expenditures or Employment



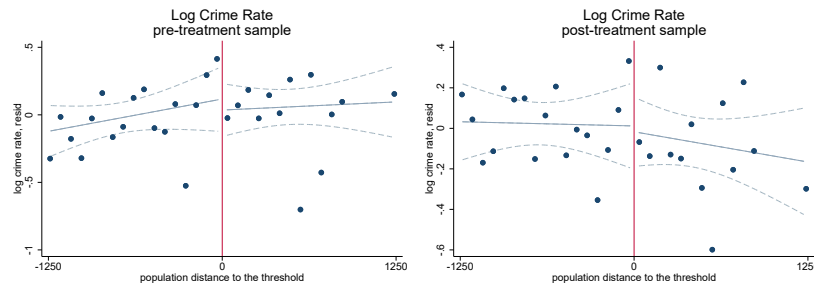
Notes: the graphs show the effect of merit system mandates on expenditures and employment for post-treatment years. Merit systems are mandated for municipalities above the threshold in 1970. Post-treatment years are 1970 to 1979 for expenditures and 1972 to 1979 for employment. The points show the average 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 VIII: 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 IX: Merit Systems do not Affect Crime Rates post-1980, RD graphs



Notes: the graphs show the post-1980 effect of merit system mandates on crime rates. Crime rates are crimes per 100,000 people. 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 | - |

Notes: the table summarized information on other policies changing at the same threshold. The information was collected performing a state-by-state legislative survey searching for the threshold in State Codes and State Session Laws. More details on the procedure are available in Online Appendix B.

Appendix Table II: Descriptives

Panel A: Sample

| | All | Below the Threshold | Above the Threshold |
|--------------------|-----|------------------------|------------------------|
| | (1) | (2) | (3) |
| All States | 139 | 99 | 40 |
| Arizona | 1 | 1 | 0 |
| Illinois | 63 | 46 | 17 |
| Iowa | 14 | 7 | 7 |
| Louisiana | 10 | 5 | 5 |
| Montana | 5 | 5 | 0 |
| Nebraska | 11 | 6 | 5 |
| West Virginia | 0 | 0 | 0 |
| Wisconsin Cities | 29 | 25 | 4 |
| Wisconsin Villages | 6 | 4 | 2 |

Notes: the table reports information about the sample used in the main analysis. The sample is restricted to municipalities within 1250 population distance from the threshold.

Panel B: Descriptive Statistics

| Statistics | N | Mean | SD | N | Mean | SD |
|-------------------------------|---------------|---------|---------|----------------|---------|---------|
| Sample | pre-treatment | | | post-treatment | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Total Crime Rate | 6832 | 117.663 | 134.381 | 9755 | 272.884 | 246.757 |
| Property Crime Rate | 6832 | 110.370 | 125.945 | 9755 | 260.589 | 237.748 |
| Larceny Rate | 6832 | 73.358 | 96.134 | 9755 | 185.699 | 178.985 |
| Vehicle Theft Rate | 6832 | 28.233 | 39.746 | 9755 | 57.795 | 68.373 |
| Robbery Rate | 6832 | 8.780 | 20.284 | 9755 | 17.096 | 40.310 |
| Violent Crime Rate | 6832 | 7.292 | 22.897 | 9755 | 12.295 | 24.652 |
| Murder Rate | 6832 | 0.119 | 1.907 | 9755 | 0.158 | 1.870 |
| Manslaughter Rate | 6832 | 0.061 | 1.692 | 9755 | 0.052 | 1.064 |
| Rape Rate | 6832 | 0.511 | 3.959 | 9755 | 0.676 | 4.582 |
| Robbery Rate | 6832 | 1.538 | 8.929 | 9755 | 2.117 | 7.953 |
| Aggravated Assault Rate | 6832 | 5.125 | 16.591 | 9755 | 9.345 | 21.121 |
| Property Crime Clearance Rate | 4528 | 0.214 | 0.347 | 9470 | 0.198 | 0.298 |
| Violent Crime Clearance Rate | 1073 | 0.617 | 0.488 | 3073 | 0.615 | 0.454 |

Notes: the table reports summary statistics (number of observations, mean and standard deviation) for total, property and violent crime rates and property and violent crime clearance rates for the sample of pre-treatment year (1960-1969, columns 1 to 3) and post-treatment years (1970-1979, columns 4 to 6). Crime rates are crimes per 100,000 people and clearance rates are number of crimes cleared by arrest over total number of crimes. The sample is restricted to municipalities within 1250 population distance from the threshold.

Appendix Table III: Covariate Balance Test

| Census year | 1970 | | | | 1980 | | | |
|--------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Population Growth | 0.047 (0.356) | 0.201 (0.390) | -0.042 (0.295) | 0.180 (0.395) | -0.315 (0.311) | -0.095 (0.249) | -0.231 (0.216) | -0.314 (0.383) |
| Observations | 90 | 114 | 138 | 68 | 75 | 104 | 132 | 58 |
| Bandwidth | 750 | 1000 | 1250 | 602 | 750 | 1000 | 1250 | 636 |
| Male | -0.004 (0.007) | -0.001 (0.006) | -0.002 (0.005) | -0.005 (0.007) | 0.002 (0.010) | 0.004 (0.008) | 0.002 (0.007) | -0.001 (0.009) |
| Observations | 90 | 114 | 138 | 95 | 77 | 106 | 134 | 85 |
| Bandwidth | 750 | 1000 | 1250 | 794 | 750 | 1000 | 1250 | 832 |
| Non-white | 0.003 (0.035) | 0.001 (0.031) | -0.001 (0.026) | 0.004 (0.037) | -0.015 (0.022) | -0.004 (0.023) | -0.056* (0.031) | -0.064*** (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 (0.023) | -0.007 (0.020) | -0.003 (0.017) | -0.010 (0.026) | 0.009 (0.010) | 0.009 (0.008) | 0.006 (0.007) | 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 (0.053) | 0.049 (0.044) | 0.029 (0.039) | 0.053 (0.053) | -0.038 (0.045) | -0.012 (0.041) | -0.025 (0.031) | -0.035 (0.052) |
| Observations | 90 | 114 | 138 | 88 | 77 | 106 | 134 | 63 |
| Bandwidth | 750 | 1000 | 1250 | 731 | 750 | 1000 | 1250 | 641 |
| Unemployed | 0.010 (0.013) | 0.008 (0.011) | 0.006 (0.009) | 0.011 (0.013) | 0.014 (0.018) | 0.000 (0.015) | -0.001 (0.013) | 0.015 (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 (0.025) | 0.032 (0.022) | 0.037* (0.020) | 0.038 (0.025) | -0.009 (0.019) | -0.015 (0.016) | -0.017 (0.015) | -0.009 (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 (1.442) | 1.322 (1.217) | 0.567 (1.092) | 1.447 (1.800) | 0.011 (2.406) | 2.319 (2.640) | 0.341 (1.885) | -0.695 (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 | -0.049 (0.105) | -0.192** (0.081) | -0.165** (0.076) | -0.039 (0.095) |
| Clusters | 77 | 107 | 136 | 74 |
| Observations | 8760 | 12300 | 15600 | 8400 |
| Bandwidth | 750 | 1000 | 1250 | 722 |

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) |
| Crime Rate | -48.586 (34.532) | -30.830 (28.621) | -9.913 (26.255) | -53.828 (37.552) | -189.165** (84.692) | -142.528** (68.957) | -118.170** (57.267) | -118.170** (57.267) |
| Clusters | 80 | 101 | 123 | 68 | 89 | 113 | 137 | 137 |
| Observations | 6407 | 8183 | 9923 | 5441 | 9054 | 11444 | 13928 | 13928 |
| Bandwidth | 750 | 1000 | 1250 | 648 | 750 | 1000 | 1250 | 1264 |
| Control Mean | 114.759 | 113.738 | 117.663 | 121.459 | 298.638 | 284.254 | 272.884 | 272.884 |
| Crime Rate, Inverse Hyperbolic Sine | -0.559* (0.309) | -0.429 (0.271) | -0.190 (0.266) | -0.578* (0.343) | -0.680*** (0.253) | -0.523** (0.214) | -0.477** (0.195) | -0.594** (0.235) |
| Clusters | 80 | 101 | 123 | 69 | 89 | 113 | 137 | 104 |
| Observations | 6407 | 8183 | 9923 | 5561 | 9054 | 11444 | 13928 | 10451 |
| Bandwidth | 750 | 1000 | 1250 | 672 | 750 | 1000 | 1250 | 901 |
| Control Mean | 4.665 | 4.623 | 4.632 | 4.750 | 5.985 | 5.923 | 5.847 | 5.952 |
| Total Crime | -1.909 (1.603) | -1.092 (1.305) | 0.072 (1.223) | -2.496 (1.641) | -10.049** (4.404) | -7.711** (3.767) | -7.421** (3.274) | -7.421** (3.274) |
| Clusters | 80 | 101 | 123 | 73 | 89 | 113 | 137 | 137 |
| Observations | 6407 | 8183 | 9923 | 5876 | 9054 | 11444 | 13928 | 13928 |
| Bandwidth | 750 | 1000 | 1250 | 708 | 750 | 1000 | 1250 | 1258 |
| Control Mean | 5.003 | 5.031 | 5.560 | 5.116 | 15.489 | 14.734 | 14.803 | 14.803 |
| Log Total Crime | -0.322 (0.198) | -0.214 (0.168) | -0.056 (0.167) | -0.271 (0.217) | -0.614*** (0.224) | -0.520*** (0.194) | -0.478*** (0.174) | -0.487** (0.245) |
| Clusters | 80 | 101 | 123 | 65 | 89 | 113 | 137 | 66 |
| Observations | 5811 | 7413 | 8929 | 4666 | 8906 | 11242 | 13623 | 6732 |
| Bandwidth | 750 | 1000 | 1250 | 638 | 750 | 1000 | 1250 | 584 |
| Control Mean | 1.331 | 1.332 | 1.365 | 1.388 | 2.373 | 2.314 | 2.239 | 2.497 |

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 Crime Rate | -0.174 (0.182) | -0.082 (0.153) | 0.031 (0.149) | -0.024 (0.220) |
| Clusters | 76 | 96 | 118 | 58 |
| Observations | 4560 | 5835 | 7112 | 3357 |
| Bandwidth | 750 | 1000 | 1250 | 603 |
| Control Mean | 4.356 | 4.368 | 4.433 | 4.450 |

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: Effect of Merit System Mandates on Probability of Reporting at Least One Incident

| Sample | pre-treatment | | | | post-treatment | | | |
|----------------------------------|-------------------|-------------------|-------------------|-------------------|---------------------|-------------------|-------------------|-------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Reported Positive Violent Crimes | -0.067 (0.074) | -0.049 (0.062) | -0.035 (0.056) | -0.066 (0.075) | -0.174** (0.088) | -0.114 (0.075) | -0.104 (0.068) | -0.068 (0.062) |
| Clusters | 80 | 101 | 123 | 79 | 89 | 113 | 137 | 169 |
| Observations | 6407 | 8183 | 9923 | 6360 | 9054 | 11444 | 13928 | 16633 |
| Bandwidth | 750 | 1000 | 1250 | 737 | 750 | 1000 | 1250 | 1547 |
| Control Mean | 0.186 | 0.175 | 0.187 | 0.188 | 0.353 | 0.335 | 0.315 | 0.322 |

Notes: *** p<0.01, ** p<0.05, * p<0.1. The tables shows the reduced form effect of merit systems on the extensive margin of crime. It presents RD estimates on the probability of reporting at least one violent crime 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 VIII: Crime-by-crime Effect of Merit System Mandates on Property Crimes

| Sample | pre-treatment | | | | post-treatment | | | |
|-------------------------------------|-------------------|-------------------|------------------|---------------------|----------------------|---------------------|---------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log Burglary and Vehicle Theft Rate | -0.048 (0.168) | -0.025 (0.140) | 0.027 (0.126) | 0.053 (0.206) | -0.410* (0.218) | -0.265 (0.181) | -0.220 (0.158) | -0.432** (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 |
| Control Mean | 3.811 | 3.791 | 3.827 | 3.823 | 4.262 | 4.215 | 4.210 | 4.222 |
| Log Larceny Rate | -0.189 (0.182) | -0.084 (0.137) | 0.019 (0.135) | -0.328** (0.157) | -0.570*** (0.212) | -0.457** (0.180) | -0.380** (0.159) | -0.627*** (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 |
| Control Mean | 4.204 | 4.223 | 4.263 | 4.242 | 5.028 | 4.986 | 4.961 | 5.062 |

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 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. 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 IX: Effect of Merit Systems on Reporting

| Sample | pre-treatment | | | | post-treatment | | | |
|------------------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Monthly Crime Report Missing | 0.074 (0.131) | 0.043 (0.114) | 0.029 (0.103) | 0.044 (0.110) | 0.053 (0.056) | 0.042 (0.045) | 0.013 (0.041) | 0.042 (0.046) |
| Clusters | 90 | 114 | 138 | 120 | 90 | 114 | 138 | 103 |
| Observations | 10800 | 13680 | 16560 | 14400 | 10560 | 13380 | 16260 | 12120 |
| Bandwidth | 750 | 1000 | 1250 | 1064 | 750 | 1000 | 1250 | 864 |
| Control Mean | 0.420 | 0.425 | 0.425 | 0.428 | 0.148 | 0.159 | 0.158 | 0.155 |

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.

Notes: The table shows that police departments did not differentially report monthly crime data to the UCR program 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). The sample is balanced: the difference between the number of observations in the pre-treatment and post-period is driven by five Montana municipalities being excluded from the sample 1975-1979, because the policy was expanded to all municipalities in the state. 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 Xa: Effect on Crime and Clearance Rates, Robustness to Data Cleaning, Population Dynamics and Specification

| Specification | post-treatment | | | | | | | | |
|-------------------------------|---------------------|----------------------|---------------------|---------------------|--------------------|---------------------|---------------------|--------------------|----------------------|
| | Drops | Uses UCR | Yearly | Controls | Drops | Quasi- | Includes | Estimates | SE |
| | outliers | population | Data | for 1980 | Population | Balanced | controls | DID | Clustered |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | Two-way |
| Log Crime Rate | -0.450** (0.187) | -0.490*** (0.184) | -0.463** (0.200) | -0.393** (0.192) | -0.385* (0.213) | -0.476** (0.199) | -0.430** (0.178) | -0.276* (0.149) | -0.451** (0.180) |
| Clusters | 113 | 113 | 113 | 113 | 89 | 92 | 113 | 113 | 113 |
| Observations | 11233 | 11242 | 945 | 11242 | 8702 | 9814 | 11242 | 18655 | 11242 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 5.357 | 5.351 | 7.788 | 5.356 | 5.236 | 5.355 | 5.356 | 4.921 | 5.356 |
| Log Property Crime Rate | -0.459** (0.180) | -0.501*** (0.178) | -0.460** (0.186) | -0.404** (0.185) | -0.398* (0.204) | -0.486** (0.192) | -0.436** (0.173) | -0.279* (0.147) | -0.461*** (0.173) |
| Clusters | 113 | 113 | 113 | 113 | 89 | 92 | 113 | 113 | 113 |
| Observations | 11205 | 11215 | 943 | 11215 | 8676 | 9799 | 11215 | 18517 | 11215 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 5.313 | 5.308 | 7.751 | 5.312 | 5.194 | 5.310 | 5.312 | 4.882 | 5.312 |
| Log Violent Crime Rate | 0.086 (0.327) | 0.020 (0.298) | -0.396 (0.347) | 0.067 (0.314) | 0.152 (0.375) | 0.126 (0.353) | -0.073 (0.132) | -0.028 (0.192) | 0.087 (0.298) |
| Clusters | 113 | 113 | 113 | 113 | 89 | 92 | 113 | 112 | 113 |
| Observations | 3766 | 3780 | 794 | 3780 | 2643 | 3166 | 3780 | 4950 | 3780 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 3.446 | 3.448 | 4.660 | 3.445 | 3.451 | 3.429 | 3.445 | 3.452 | 3.445 |
| Property Crime Clearance Rate | 0.030 (0.032) | - - | 0.020 (0.030) | 0.032 (0.030) | 0.039 (0.033) | 0.024 (0.033) | 0.028 (0.032) | 0.011 (0.044) | 0.028 (0.030) |
| Clusters | 113 | - | 113 | 113 | 89 | 92 | 113 | 113 | 113 |
| Observations | 11079 | - | 943 | 11215 | 8676 | 9799 | 11215 | 16785 | 11215 |
| Bandwidth | 1000 | - | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 0.195 | - | 0.193 | 0.193 | 0.194 | 0.191 | 0.193 | 0.198 | 0.193 |
| Violent Crime Clearance Rate | 0.100** (0.046) | - - | 0.150*** (0.052) | 0.103** (0.042) | 0.075 (0.050) | 0.110** (0.045) | 0.063 (0.045) | 0.168** (0.080) | 0.106*** (0.039) |
| Clusters | 113 | - | 113 | 113 | 89 | 92 | 113 | 112 | 113 |
| Observations | 3712 | - | 794 | 3780 | 2643 | 3166 | 3780 | 4818 | 3780 |
| Bandwidth | 1000 | - | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 0.626 | - | 0.660 | 0.619 | 0.649 | 0.609 | 0.619 | 0.625 | 0.619 |

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 yearly data; (4) using UCR population to calculate crime rates; (5) controlling for 1980 population; (6) restricting the sample of municipalities reporting at least half of the times; (7) including baseline controls; (8) estimating a DID specification; (9) 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 8. State-month fixed effects are included in all columns (state-year fixed effects are included in column (3)).

Appendix Table Xb: Effect on Crime and Clearance Rates, Robustness to Overlapping Legislation

| Sample State being excluded | post-treatment | | | | | | | |
|--------------------------------|---------------------|---------------------|----------------------|---------------------|---------------------|---------------------|---------------------|----------------------|
| | AZ | IL | IA | LA | MT | NE | WI CITY | WI VILL |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log Crime Rate | -0.451** (0.187) | -0.443** (0.226) | -0.491** (0.200) | -0.460** (0.191) | -0.459** (0.189) | -0.439** (0.200) | -0.423* (0.221) | -0.466** (0.191) |
| Clusters | 113 | 60 | 101 | 103 | 108 | 105 | 91 | 110 |
| Observations | 11242 | 5968 | 9923 | 10574 | 10995 | 10330 | 8779 | 10883 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 5.356 | 5.356 | 5.356 | 5.356 | 5.356 | 5.356 | 5.356 | 5.356 |
| Log Property Crime Rate | -0.461** (0.180) | -0.442** (0.225) | -0.505*** (0.191) | -0.465** (0.182) | -0.467** (0.182) | -0.453** (0.192) | -0.439** (0.214) | -0.476*** (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 |
| Control Mean | 5.312 | 5.312 | 5.312 | 5.312 | 5.312 | 5.312 | 5.312 | 5.312 |
| Log Violent Crime Rate | 0.087 (0.320) | 0.027 (0.264) | 0.094 (0.335) | 0.123 (0.352) | 0.082 (0.320) | 0.077 (0.326) | 0.066 (0.325) | 0.087 (0.320) |
| Clusters | 113 | 60 | 101 | 103 | 108 | 105 | 91 | 110 |
| Observations | 3780 | 1338 | 3409 | 3376 | 3742 | 3543 | 3506 | 3766 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 3.445 | 3.445 | 3.445 | 3.445 | 3.445 | 3.445 | 3.445 | 3.445 |
| Property Crime Clearance Rate | 0.028 (0.031) | 0.078 (0.057) | 0.028 (0.033) | 0.032 (0.032) | 0.026 (0.031) | 0.016 (0.032) | 0.013 (0.030) | 0.027 (0.032) |
| 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 |
| Control Mean | 0.193 | 0.193 | 0.193 | 0.193 | 0.193 | 0.193 | 0.193 | 0.193 |
| Violent Crime Clearance Rate | 0.106** (0.042) | 0.258*** (0.068) | 0.101** (0.044) | 0.091** (0.044) | 0.103** (0.042) | 0.102** (0.044) | 0.099** (0.044) | 0.105** (0.042) |
| Clusters | 113 | 60 | 101 | 103 | 108 | 105 | 91 | 110 |
| Observations | 3780 | 1338 | 3409 | 3376 | 3742 | 3543 | 3506 | 3766 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 0.619 | 0.619 | 0.619 | 0.619 | 0.619 | 0.619 | 0.619 | 0.619 |

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 Xc: Effect on Crime and Clearance Rates, Robustness to the Estimation

| Sample | post-treatment | | | | | |
|-------------------------------|------------------------------|--------------------------------|---------------------------|---------------------------|--------------------------------------|---|
| Estimation | LLR, Triangular Kernel | LLR, Epanechnikov 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 Crime Rate | -0.550*** (0.207) | -0.541*** (0.202) | -0.632** (0.274) | -0.428 (0.291) | -0.403** (0.202) | -0.452** (0.187) |
| Clusters | 113 | 113 | 113 | 113 | 113 | 113 |
| Observations | 11242 | 11242 | 11242 | 11242 | 11243 | 11242 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 5.356 | 5.356 | 5.356 | 5.356 | 5.356 | 5.356 |
| Log Property Crime Rate | -0.558*** (0.197) | -0.548*** (0.193) | -0.640** (0.259) | -0.451* (0.269) | -0.422** (0.192) | -0.462** (0.180) |
| Clusters | 113 | 113 | 113 | 113 | 113 | 113 |
| Observations | 11215 | 11215 | 11215 | 11215 | 11216 | 11215 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 5.312 | 5.312 | 5.312 | 5.312 | 5.312 | 5.312 |
| Log Violent Crime Rate | 0.108 (0.389) | 0.081 (0.364) | 0.128 (0.454) | 0.342 (0.485) | 0.132 (0.268) | 0.092 (0.320) |
| Clusters | 113 | 113 | 113 | 113 | 113 | 113 |
| Observations | 3780 | 3780 | 3780 | 3780 | 3941 | 3780 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 3.445 | 3.445 | 3.445 | 3.445 | 3.445 | 3.445 |
| Property Crime Clearance Rate | 0.013 (0.032) | 0.016 (0.032) | -0.007 (0.040) | 0.008 (0.039) | 0.033 (0.034) | 0.028 (0.031) |
| Clusters | 113 | 113 | 113 | 113 | 113 | 113 |
| Observations | 11215 | 11215 | 11215 | 11215 | 11216 | 11215 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 0.193 | 0.193 | 0.193 | 0.193 | 0.193 | 0.193 |
| Violent Crime Clearance Rate | 0.082* (0.044) | 0.088** (0.044) | 0.067 (0.057) | 0.056 (0.059) | 0.176*** (0.063) | 0.099** (0.042) |
| Clusters | 113 | 113 | 113 | 113 | 113 | 113 |
| Observations | 3780 | 3780 | 3780 | 3780 | 3941 | 3780 |
| Bandwidth | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| Control Mean | 0.619 | 0.619 | 0.619 | 0.619 | 0.619 | 0.619 |

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 Epanechnikov 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 XI: 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 employment 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 XII: 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 (12.74) | 37.220 (12.47) | 34.010 (10.41) | 37.730 (12.34) |
| Highest grade achieved | 11.210 (2.416) | 13.390 (2.119) | 14.940 (1.964) | 13.080 (2.663) |
| Finished high school | 0.414 (.493) | 0.712 (.454) | 0.738 (.441) | 0.614 (.487) |
| Finished two years of college | | 0.616 (.487) | 0.713 (.453) | 0.455 (.498) |
| Veteran status | 0.583 (.494) | 0.434 (.497) | 0.475 (.5) | 0.499 (.5) |

Notes: This table reports descriptive statistics for policemen characteristics. Columns 1 to 3 report information for a specific census, while column 4 reports information for the pooled sample. 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 XIII: Effect on Demographic Composition of Police Departments, Individual Level Regressions

| Sample | post-treatment | | | |
|----------------------|-------------------|-------------------|-------------------|-------------------|
| | (1) | (2) | (3) | (4) |
| Finished High School | -0.088 (0.097) | 0.009 (0.093) | -0.002 (0.076) | -0.069 (0.101) |
| Clusters | 136 | 176 | 221 | 134 |
| Observations | 400 | 500 | 650 | 400 |
| Bandwidth | 750 | 1000 | 1250 | 716 |
| Age | 1.362 (2.412) | -1.594 (2.220) | -0.993 (2.054) | -0.892 (2.245) |
| Clusters | 136 | 176 | 221 | 191 |
| Observations | 400 | 500 | 650 | 550 |
| Bandwidth | 750 | 1000 | 1250 | 1105 |
| Veteran | -0.019 (0.093) | 0.028 (0.087) | 0.054 (0.074) | 0.036 (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 that 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 XIV: Effect of Merit Systems on Turnover and Wages

| Sample | post-treatment | | | |
|-------------------|-------------------|------------------|------------------|------------------|
| | (1) | (2) | (3) | (4) |
| Fraction New Hire | -0.000 (0.127) | 0.098 (0.123) | 0.088 (0.109) | - - |
| Clusters | 119 | 159 | 191 | - |
| Observations | 99 | 129 | 155 | - |
| Bandwidth | 750 | 1000 | 1250 | - |
| Average Wage | 0.395 (0.636) | 0.628 (0.529) | 0.651 (0.463) | 0.579 (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 certainly 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 XV: Heterogeneous Effects by Whether Chief under Merit System

| Sample | post-treatment | | | |
|------------------------------------|----------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) |
| Log Crime Rate*Chief Not Protected | -0.480** (0.238) | -0.363* (0.201) | -0.303* (0.175) | -0.535** (0.246) |
| Log Crime Rate*Chief Protected | -0.817*** (0.228) | -0.640*** (0.208) | -0.584*** (0.191) | -0.883*** (0.239) |
| Clusters | 89 | 113 | 137 | 76 |
| Observations | 8906 | 11242 | 13623 | 7730 |
| Bandwidth | 750 | 1000 | 1250 | 661 |
| Control Mean | 5.410 | 5.356 | 5.313 | 5.453 |
| Pvalue | 0.127 | 0.150 | 0.112 | 0.123 |

Notes: *** p<0.01, ** p<0.05, * p<0.1. The table shows that heterogeneous effects by whether the police chief was also covered by a merit system. The table presents RD estimates on crime rates interacted with a dummy for being in a state whether the chief was or was not covered for post-treatment years (1970 to 1979). 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. The p-value reported is from a test of equality of the two coefficients. State-month fixed effects are included in all columns.

Appendix Table XVI: Effect of Merit System Mandates on Reporting post-1980

| Sample | post-treatment | | | |
|------------------------------|------------------|------------------|------------------|------------------|
| | (1) | (2) | (3) | (4) |
| Monthly Crime Report Missing | 0.005 (0.053) | 0.041 (0.059) | 0.056 (0.053) | 0.060 (0.053) |
| Clusters | 74 | 103 | 130 | 49 |
| Observations | 8880 | 12360 | 15600 | 5880 |
| Bandwidth | 750 | 1000 | 1250 | 553 |
| Control Mean | 0.070 | 0.071 | 0.101 | 0.073 |

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.