# Civil Service Reform and Organizational Practices: Evidence from the Pendleton Act*

Diana Moreira
UC Davis

Santiago Pérez
UC Davis and NBER

Online Appendix

**Abstract**

Competitive exams are a standard method for selecting civil servants. Yet, evidence on the effectiveness of such approach is mixed, and lack of personnel data limits our understanding of the mechanisms underlying this varying success. We digitize personnel and financial data to study the impacts of the 1883 Pendleton Act, which mandated exams for some employees in the largest US customs-collection districts. The reform improved targeted employees' professional background and reduced turnover. However, it did not increase cost-effectiveness in revenue collection. An unintended consequence of the reform was to induce hiring in exempted positions, provoking distortions in districts' personnel structure.

A professionalized bureaucracy is increasingly seen as a key determinant of states' ability to successfully implement public policies (Rauch *et al.* , 1995; Finan *et al.* , 2017). Historically, the standard recipe for achieving such professionalization has been the enactment of civil service reforms introducing recruitment through competitive exams and employee protection from political dismissals. The promise of these reforms is that reducing politicians' control over hiring and firing decisions would allow governments to attract and retain more qualified employees, which would, in turn, lead to improvements in bureaucratic performance.

Although civil service reforms have improved performance in some contexts (Rauch *et al.* , 1995; Ornaghi, 2016; Xu, 2018), they appear to be no silver bullet: In an assessment of 71 reforms funded by the World Bank, only 42% were rated as successful by an independent agency (Webb, 2008).[1] Moreover, there has been a recent push in some developed countries toward greater flexibility for hiring and removing public employees, often fueled by concerns that traditional civil service rules might result in a

*dsmoreira@ucdavis.edu, seperez@ucdavis.edu. We thank Luiza Aires and Lisa Pacheco for outstanding research assistance, and Enrique Pérez for help with data collection. We have benefited from the comments of Oriana Bandiera, Brian Beach, Assaf Bernstein, Sandra Black, Nicolás Caramp, Katherine Eriksson, James Feigenbaum, Robert Gibbons, Michela Giorcelli, Walker Hanlon, Daniel Honig, Rick Hornbeck, Chris Meissner, Andrea Pozas-Loyo, Sarah Quincy, Arman Rezaee, Monica Singhal, Michael Ting, Gergely Ujhelyi, Martin Williams, Chenzi Xu and Noam Yutchmann, as well by seminar participants at *Cooperacion Andina de Fomento*, CEPR STEG Workshop, Cornell University, University of California - Riverside, NEUDC, NBER Organizational Economics Fall 2020 Conference, Columbia University, University of Chicago Harris School of Public Policy and Stanford CASBS.

[1]These assessments were based on a comparison of the administrative capacities of countries before and after a World Bank- funded reform of their civil service. Based on this evidence, Webb (2008) writes that "despite the continued efforts and some modification of the approach, civil service reform has been relatively unsuccessful."

bureaucracy that is unresponsive to citizens' needs.[2] Ultimately, whether reducing political control over the civil service actually enables governments to hire and retain more qualified employees is an empirical question, and one for which there is remarkably limited direct evidence. Opening the black box of the bureaucracy is a necessary step to understand the mechanisms underlying civil service reforms' varying degrees of success.

This paper studies the impacts of the 1883 Pendleton Civil Service Reform act. This act, which introduced competitive exams for the selection of certain federal employees, is widely regarded as the first step towards a professionalized civil service in the US (Van Riper, 1976). Our analysis focuses on the consequences of the act for the functioning of the Customs Service, a key government agency that, by the time of the reform, collected more than half of federal revenue. Although we find that the reform indeed improved targeted employees' professional background and reduced turnover, we show that these changes did not translate into higher cost-effectiveness in customs revenue collection.

Our analysis is based on newly digitized data on the personnel and finances of US customs-collection districts. First, we digitized districts' personnel records spanning 1871 to 1893. These records include employees' names, place of employment, position, and salary. We gathered additional information on the professional background of these employees by linking these data to population censuses, using name-based matching techniques (Abramitzky *et al.* , 2019). Second, we digitized yearly data on districts' expenses and receipts.

Our identification strategy exploits the fact that the requirement to hire some employees through examinations applied only to districts that had 50 or more employees by 1883. We use this feature of the reform in a difference-in-differences design, comparing districts that were subject to the reform to those that were not, before and after 1883. Our identification assumption is that, in the absence of the reform, the outcomes of districts that were above the 50 employees cutoff would have evolved similarly to the outcomes of those that were below. Consistent with this assumption, we show that the outcomes of these two groups of districts evolved similarly prior to the reform. Moreover, there are no differential changes in outcomes post-1883 when we estimate the "effects" of placebo reforms that use alternative employee cutoffs.

In the first part of the paper, we show that the reform worked as expected by its proponents along two main dimensions. First, it led to a sizable (25%) reduction in employee turnover. This reduction was larger for workers in positions subject to exams, as well as in years in which control of the Federal administration changed party hands. Second, it led to an improvement in targeted employees' professional background: New hires in positions requiring exams were 11 percentage points less likely to report working in an unskilled occupation prior to joining the Customs Service, and 9 percentage points more likely to report working in a professional one. Since exams were aimed at testing practical knowledge relevant to an applicant future position (rather than formal academic training), we interpret these changes in occupational background as reflecting a likely improvement in targeted employees' actual qualifications for their job.[3] Indeed, a shortage of workers with a professional background might have

[2] In the US, the Trump administration issued an executive order exempting certain positions from competitive recruitment (https://www.forbes.com/sites/tomspiggle/2020/10/28/trumps-executive-order-would-diminish-civil-service-employment-protections/?sh=4b012c067b3a). Ujhelyi (2014) describes recent state legislation in the US limiting traditional civil service protections..

[3] The focus on practical skills was in contrast to other civil service reforms. For instance, Grindle (2012) argues that in the UK

been a binding constraint in achieving cost-effectiveness: Prior to the reform, there was a strong positive correlation between changes in the share of employees with a professional background and changes in district's revenue.

Ten years after the reform, nearly 60% of employees in reformed districts had been appointed through an exam. We next ask if this change led to increased cost-effectiveness in customs revenue collection. We expect improvements in this regard through three main channels. First, limiting the room for patronage could have curtailed personnel expenses by reducing the number of employees hired solely to reward political loyalty. Second, by creating a separation between bureaucrats and politicians, the reform could have reduced corruption thereby increasing revenue. Third, to the extent that the reform increased bureaucratic expertise, workers might have been better equipped to interpret and enforce the customs laws.

Surprisingly, however, we find that the reform had limited impacts on cost-effectiveness. First, the reform did not lead to a reduction in total expenses (nor in districts' total number of employees): Our point estimates are close to zero and are statistically insignificant. Similarly, we find no statistically significant evidence that the reform led to increased customs revenue. Indeed, when estimating event-study specifications we see little evidence that would suggest an increase in revenue over time: the post-reform estimated effects are sometimes positive and sometimes negative, lacking a clear pattern. Finally, as expected given the limited effects on expenses and revenue, we also see no indication of an improvement in our main measure of cost-effectiveness, the "revenue per dollar spent".

In the last part of the paper, we investigate possible reasons why the changes in personnel outcomes did not appear to have translated into higher cost-effectiveness in revenue collection. We first discuss the potential role played by the reform's incomplete scope across the layers of the bureaucracy. As typical in civil service reforms, the Pendleton Act targeted only a subset of employees. Specifically, it targeted employees in mid-tier positions but exempted those below a salary threshold as well as districts' top managers (the "collectors of customs").[4] This incomplete scope could have been important for two reasons. First, by exempting employees below a salary threshold, the reform created incentives to hire additional workers in low-paid positions. Indeed, we document that the reform led to a near *doubling* in the share of workers in such positions. This shift was likely pernicious for reformed districts' performance, both because it distorted their personnel structure but also because low-paid employees had weaker professional backgrounds. Second, to the extent that collectors mattered for districts' outcomes (which we show to be the case by estimating "collector fixed effects" and exploiting collectors' deaths while in office as in Jones & Olken, 2005 and Besley *et al.* , 2011), not changing their selection method also likely limited the reform's ability to improve cost-effectiveness.

Finally, we discuss three additional potential explanations for the lack of effects on costs effectiveness. First, we find limited evidence that the lack of detectable effects on cost-effectiveness was due to the reform spilling over to the non-reformed districts: proximity to a reformed district does not predict either increases or decreases in revenue in the post-reform period. Second, the effect of the reform on

exams were designed such that their contents would only be accessible to those with "access to elite educations at Oxford and Cambridge". In contrast, in the US the Civil Service Commission "maintained that a common school education was sufficient to pass examination" (Hoogenboom, 1959).

[4] Dahlström *et al.* (2012) refers to the coexistence of patronage and merit hires after civil service reforms as "formal political discretion". Subsequent reforms that were enacted at the state and local levels in the US also led to the coexistence of merit and non-merit employees (Ujhelyi, 2014; Ornaghi, 2016).

cost-effectiveness was limited even over a 20-year horizon. This result is contrary to the hypothesis that to fully capitalize on the benefits of the reform required changes (for instance, fully replacing employees hired through the old regime) that took longer than ten years to implement. Third, we consider the possibility that, although employees hired through exams might have been of better "quality", they might have also exerted less effort (or otherwise be less responsive) than patronage hires. Although we do find some suggestive evidence consistent with this explanation, we note that, unlike some modern civil service protections, the Pendleton Act *did not* provide tenure to employees. Hence, the disincentive effects of the reform might have been less prominent than in other contexts.

Our data do not enable us to establish if the reform led to improvements in performance along margins other than revenue per dollar spent.[5] For instance, reformed districts might have become faster at clearing imports or may have improved on how closely they followed the tariff laws (which would not necessarily lead to higher revenue).[6] Although "revenue per dollar spent" does not incorporate all dimensions of performance, it does capture an important aspect of it in the context of an agency whose primary goal was revenue collection. Indeed, this measure was regularly discussed both in government publications and by proponents of civil service reform, who blamed patronage for the high cost to collect in the US.[7] Moreover, similar measures have been used by other scholars studying the performance of government units in charge of revenue collection (Khan *et al.* , 2016; Xu, 2018; Naritomi, 2019).

This paper makes two main contributions. First, we contribute to the growing literature on the personnel economics of the state (summarized in Finan *et al.* , 2017 and Pepinsky *et al.* , 2017). A core focus of this literature has been on understanding the link between how civil servants are selected and their subsequent performance on the job.[8] Our results provide evidence on the effects of competitive civil service exams, a selection method that has been a defining characteristic of modern bureaucracies.[9] In addition, we show that, by changing the incentives of appointing officers, policies shaping how some bureaucrats are selected might have additional indirect impacts on the hierarchical and occupational structure of government organizations.

Second, we contribute to the literature on civil service reforms. In the US, state and local reforms reduced incumbents' parties chances of reelection (Folke *et al.* , 2011), reduced political budget cycles (Bostashvili & Ujhelyi, 2019), and improved bureaucratic performance (Ornaghi, 2016; Rauch *et al.* , 1995).[10] Remarkably, however, there is very limited evidence on the effects of these reforms on the main objects that they are intended to change: namely, the turnover rate and qualifications of bureaucrats –and

[5]In this analysis, we take the goal of collecting customs revenue as given. However, increasing such revenue might have been negative for the US economy to the extent that it reduced international trade and increased domestic prices.

[6]Although errors leading to higher import duties would have been more likely to be challenged by importers.

[7]The Annual Reports of the Secretary of the Treasury (US Congress, 1874-1893) include such measure in several of their editions. Claims that the "cost to collect" was unusually high in the US were at the core of reformists' arguments.

[8]See for instance Ashraf *et al.* (2020), Dal Bó *et al.* (2013), Deserranno (2019), Estrada (2019), Muñoz & Prem (2020), Dahis *et al.* (2020), Voth & Xu (2019) and Weaver (2020)..

[9]Nearly 80% of countries use formal examinations to select some of their public employees (Teorell *et al.* , 2011).

[10]A growing literature focuses on patronage in developing countries (see for instance Colonnelli *et al.* , 2020, Brassiolo *et al.* , 2020 and Akhtari *et al.* , 2019). Xu (2018) studies the link between patronage and the performance of British Empire's colonial governors. Xu (2018) finds that being connected with the Secretary of State worsens governor's performance but that this effect disappears after a reform that limited patronage. We depart from Xu (2018) in three main ways. First, we study how a reform affected bureaucrats' *selection*, whereas Xu (2018) focuses on patronage's *incentive* effects. Second, our analysis focuses on the reform's consequences for the overall personnel structure and organization of reformed units, whereas Xu (2018)'s analysis of personnel outcomes focuses on top bureaucrats (that is, colonial governors). Third, the reform in Xu (2018) did not establish exam-based recruitment, but rather that appointments would need to be overseen by an independent board.

the existing evidence casts doubts on whether these reforms actually generate these intended changes (Ornaghi, 2016). Our data allow us to investigate how these reforms affect *both* the personnel outcomes and the overall organization and performance of reformed units. Doing so enables us to better unpack the factors mediating a reform's overall success: the reform was binding and it partially succeeded in improving personnel outcomes, yet it led to distortions in personnel structure by incentivizing hiring in exempted positions. Finally, we focus on an important historical and policy context: the Pendleton Act is a landmark reform in US history, and the ability to collect revenue is a key determinant of state capacity.[11]

# 1 Historical Background

## 1.1 The US Customs Service and Customs Revenue Collection

Revenue collected by the Customs Service accounted for more than half of federal revenue by the early 1880s (Schmeckebier, 1924). Most of this revenue was due to the "collection of duties" on imported goods. Hence, to understand why changing the method for selecting employees could have led to improvements in the cost-effectiveness of the Customs Service, it is important to understand the procedure for the collection of these duties in the late 19th-century. The key takeaway of this subsection is that this process was complex and prone to errors and corruption.

Upon arrival to a US port of entry, importers had to present a manifest detailing the articles included in their shipment. After receiving this manifest, officers of the Customs Service had to verify its accuracy and establish the duties payable on the shipment. In essence, the role of customs officers was to guarantee that anyone bringing goods into the US passed through an authorized port and paid duties according to law (Parrillo, 2013).

Import duties depended on the physical quantities of imported products (for goods subject to specific duties), their total monetary value (for those subject to ad-valorem duties), and their product category. In the case of some goods subject to specific duties (for instance, a bushel of wheat), accurately computing the amount owed by the importer simply required weighing the shipment and applying the corresponding rate.[12] In other cases, a detailed examination of the imported goods' physical properties was required. For instance, determining the correct tariff on raw sugar required a chemical test to determine its saccharine content (Schmeckebier, 1924).

Goods subject to ad-valorem tariffs required an "appraisal" –based on the price of similar goods being commercialized in the exporting country– of their monetary value. Hence, it was essential for appraisers to be aware of current market values so as to detect cases in which importers attempted to undervalue the contents of a shipment. According to Parrillo (2013), however, employees hired through patronage "were typically ignorant, sometimes unable to read the foreign languages in which invoices were written", leaving them poorly equipped to assess the accuracy of importers' invoices.

[11] Our study provides the first quantitative evaluation of the impacts of the Pendleton Act which, as discussed in Johnson & Libecap (1994a), had not yet been possible due to lack of adequate data. The Pendleton Act has attracted the attention of scholars in economics, political science, history and public administration (see for instance Hoogenboom (1968); Theriault (2003); Johnson & Libecap (1994a,b); Libecap & Johnson (2007); Van Riper (1976)).

[12] Yet, Schmeckebier (1924) describes a scandal involving a sugar importer who had tampered with the scales so that they would show less weight.

Finally, after being appraised, goods had to be "classified" in order to determine the rate of duty to be levied on them. Since some articles were not explicitly enumerated in the law, the law established that in those cases the rate would correspond to that of the "closest" listed article. This, in turn, implied that importers had incentives to have their goods included in the classification entailing the lowest duties. Schmeckebier (1924) provides several examples of the ambiguities in classifying some goods, and of how importers could exploit such ambiguities.[13]

### 1.2 US Customs Service Before the Pendleton Act

Prior to the reform, hiring decisions in the Customs Service were ruled by the "spoils system." Under this system, appointment to office was primarily based on political and personal connections rather than on "fitness for office" (Ziparo, 2017). Political bosses used these positions to reward supporters as well as to fuel political machines, often requiring a fraction of employees' salaries in the form of "political assessments" (Hoogenboom, 1968).

Proponents of civil service reform targeted the Customs Service as a prime example of the dangers of patronage appointments. In 1877, Congress appointed a number of commissions to investigate the functioning of the major collection ports in the country. As part of its investigations, the *Jay Commission* (in charge of investigating the port of New York) compared the costs and functioning of the US Customs Service to those in other countries (House & of the Treasury, 1877). The Commission concluded that the "cost to collect one dollar" was much higher in the US than in Germany, England or France, where customs' employees were appointed through civil service examinations. The Commission attributed this higher cost to overstaffing (which inflated operating expenses), as well as to "errors and fraud" of patronage employees (which led to revenue losses of up to 40%).[14]

The *Jay Commission* report includes several examples that suggest high levels of inefficiency and corruption in the New York collection district prior to the reform (House & of the Treasury, 1877). The district was overstaffed, so that "many of the weighers and foremen rendered little, if any, service to the government" and "some of the clerks performed no duties at all". Moreover, corruption led to lower revenue since "the law against the acceptance of bribes was a dead letter." Employees' carelessness further contributed to reduced revenue, as "some of the employees didn't have brains enough to do the work, some were incapacitated by ignorance and some by carelessness and indifference."[15] Similar examples can be found in the reports corresponding to the Philadelphia and New Orleans districts, as well as in earlier investigations about New York.

[13] For instance, he describes the case of the importation of "pickle forks" which could plausibly have been assigned to three different classes (each with a different duty payable).

[14] According to Rogers (1921), "the investigating commission appointed by President Grant in 1871 reported that the loss was probably twenty-five percent. The New York Chamber of Commerce assured Sherman's commission that it had risen to forty percent in 1877." Similarly, the 1866 "Report of the Special Commissioner of the Revenue" claimed that $12 to $25 million in revenue were annually lost just in the New York district due to "frauds, waste and incompetency". These losses were mostly driven by the "undervaluation of invoices" (Wells, 1866). For reference, the entire Customs Service collected around 130 US$ million in 1866.

[15] For instance, they described how "In fact, much of the weighable merchandise was not weighed at all. The Custom-House employees either took their figures from the city weighers or copied off the foreign marks of weight found on the packages." (Rogers, 1921)

## 1.3 Civil Service Reform Movement and the Pendleton Act of 1883

While pressure for the adoption of a merit reform had been mounting since the 1860s (Hoogenboom, 1968), the exact timing of the passing of the Pendleton act is related to two events. First, in July of 1881, President James A. Garfield was shot by a disappointed office-seeker (Garfield would die by September). This assassination put civil service reform at the center of the political stage, and provided reformists with a powerful example of the evils of the "spoils system." Only three months after Garfield's death, Democratic senator George H. Pendleton reintroduced a civil service reform bill. Second, Democrats took control of the House in 1882. Fearing that they would lose the 1884 presidential election, Republicans supported the bill with the aim of protecting Republican officeholders from politically motivated dismissals. In January of 1883, President Chester A. Arthur signed Pendleton's bill into law.

The act's main provision was to establish that employees in certain "classified" positions within the federal administration would need to be selected through "open, competitive examinations" (United States Civil Service Commission, 1883). Hence, after the passing of the act, applicants to these classified positions were required to complete an exam, and only those who obtained a passing grade were deemed eligible for appointment. On the opening of a vacancy, the "Examining Board" in each collection district produced a list of the *top four* candidates (based on the exam results) from which the appointing officer would need to choose.[16]

Exams were aimed at testing practical knowledge relevant to an applicant's future position rather than formal academic training. In addition, the act included a number of provisions intended to ensure political neutrality in the appointment process. First, exam administration was overseen by a bipartisan "Civil Service Commission". Second, the exams were to be graded anonymously. Third, exams could not include questions aimed at eliciting an applicant's political orientation. Figure A2 shows an example question of the arithmetic exam for applicants to the position of clerk.

The "classified" (that is, subject to exams) Customs Service was initially restricted to collection districts with *at least 50* employees, and to positions earning $900 or more within these districts: By 1883, 11 districts met this threshold.[17] From 1883 to 1893, nearly 22,000 applicants completed an exam to join the Customs Service, out of which about 2,800 had been appointed by 1893 (United States Civil Service Commission, 1893, p.240). By 1893, nearly 60% of the workforce in classified collection districts had been appointed through examinations.

Although it changed the method of appointment for some federal employees, it is important to emphasize that the act *did not* grant tenure to these employees: classified workers remained open to the possibility of removal as administrations changed (Johnson & Libecap, 1994a).[18] Later reforms further increased the stability of federal government employment by introducing the notion that employees could only be removed for "just causes" (Johnson & Libecap, 1994a).

The conventional wisdom is that, although initially covering only a limited number of federal civil

[16] This number was further reduced to three in March of 1888 (Commission, 1886, p.128).

[17] These were New York, NY; Boston, MA; Philadelphia, PA; San Francisco, CA; Baltimore, MD; New Orleans, LA; Chicago, IL; Burlington, VT; Portland, ME; Detroit, MI and Port Huron, MI. Districts that fell below 50 employees after the reform was implemented had to nevertheless remain within the classified service. The law further divided the classified civil service into the "Classified Departmental Service" for federal employees in DC, and the "Classified Postal Service" for postal workers.

[18] "The power to remove for even the most partisan and selfish reasons remains unchanged" (United States Civil Service Commission, 1883). The only exception is that employees could no longer be removed for refusing to perform a political service or paying a political assessment, although this provision of the law applied to all positions, not just the classified ones).

service positions, the reform improved the efficiency of the federal bureaucracy in general and of the Customs Service in particular. For instance, Hoogenboom (1959) argues that "service in post offices and customhouses was vastly improved". Similarly, Johnson & Libecap (1994b) describe how the reform "improved the performance of federal workers in the positions that were covered by the law" (although they acknowledge that these claims are based on limited quantitative data). The annual reports of the Civil Service Commission (arguably, an interested party) contain a number of accounts of customs collectors describing improvements in the functioning of their agencies after the reform (see pages 32-39 in Commission (1884) and page 38 in Commission (1885)). Indeed, in its 15th report, the Commission claimed that the reform had led to a 25% decline in the cost of collecting customs revenue, although the basis of this calculation is unclear (Commission, 1897).[19]

## 2 Data

**Personnel Records.** We digitized Customs Service personnel records using the *Official Registers of the United States* (Department of the Interior, 1871-1893). This biennial publication contains the name of every federal employee, their job title, state or country of birth, US state of appointment, the location of their post, and their compensation. Our data include information on approximately 50,000 Customs Service employee-years, and span 1871 to 1893. Figure A3 shows an example page listing employees of the New York collectors' office in 1883.

We gathered additional information on the professional background of these employees by linking the Registers to US population censuses. Specifically, we used name-based matching techniques (Abramitzky *et al.* , 2019) to link each of the 1871-1893 Registers to the 1850, 1860, 1870 and 1880 censuses. Through this procedure, we obtained information on employees' occupations *prior* to their employment in the federal government. While we provide further details on the linking strategy and sensitivity checks in Online Appendix Section A.1, we note here that: (1) the reform does not affect the likelihood of finding an individual in the census (Table A1), and (2) the results that *do not* require the linked data are very similar when estimated in this linked sample (Table A2).

**Financial Records.** We collected yearly data on the annual receipts and expenses of each collection district from 1874 to 1893.[20] These data come from the *Annual Report of the Secretary of the Treasury on the State of the Finances* (US Congress, 1874-1893). This report was published annually and includes detailed information on the revenues and expenditures of the different branches of the federal government, including the Customs Service.

Our measure of expenses corresponds to the amount reported as "expenses for collecting the revenue from customs" (US Congress, 1874-1893). Our baseline measure of total receipts corresponds to the sum of receipts from "customs", "fines, penalties and forfeitures", "emolument fees", "services of United States officers", "labor, drayage and storage" and "weighing fees".[21] Finally, our baseline measure of

[19] The quote is as follows: "Officials in charge of collecting the customs duties of the Government have emphatically stated that there has been a saving of about one-fourth in the cost of gathering this part of the public revenue." (Commission, 1897). Indeed, Fish (1905)'s analysis of the Pendleton Act states (regarding the savings claimed by the Civil Service Commission) that "such definite statements, however, lack a firm basis."

[20] In most cases, these data are reported at the collection-district level. The one exception is that we have separate information on the receipts and expenses for those "ports of delivery" in which the surveyor acted as a collector.

[21] Receipts from customs accounted for more than 95% of total receipts in our baseline sample of districts.

cost-effectiveness is the ratio between total receipts and expenses.

## 3 Empirical Strategy

In our main analysis, we estimate a difference-in-differences model comparing the outcomes of classified and non-classified collection districts, before and after 1883. We estimate:

$$y_{ct} = \alpha_c + \alpha_t + \beta Classified_c \times After_{ct} + \gamma X_{ct} + \epsilon_{ct} \tag{1}$$

where $y_{ct}$ is an outcome of district $c$ in year $t$, $\alpha_c$ are district fixed effects, and $\alpha_t$ are year fixed effects. When we focus on personnel outcomes, we estimate an analogous equation but using employees as the unit of analysis.[22] Our interaction of interest is $Classified_c \times After_{ct}$, which takes a value of one in the post-reform period (that is, after 1883) for districts employing at least 50 employees by 1883.[23] For some specifications, $X_{ct}$ includes census region-year fixed effects as well interactions between a districts' number of employees in 1883 and year dummies. Throughout the paper, we cluster standard errors at the district level.

In addition to this baseline model, we estimate event-study specifications in which we allow the difference between the control and treatment groups to vary over time. We estimate:

$$y_{ct} = \alpha_c + \alpha_t + \sum_{t=1874}^{1893} \beta_t Classified_c \times \alpha_t + \gamma X_{ct} + \epsilon_{ct} \tag{2}$$

where the $\beta_t$ coefficients capture the differential evolution in outcomes in the classified and non-classified districts during our sample period.

A first concern with our strategy is that the 50 employees cutoff might have been chosen so as to include or exclude certain districts. Similarly, districts might have manipulated their number of employees in anticipation of the reform. There are three pieces of evidence that suggest that this possibility is unlikely. First, there are no districts that, based on their number of employees in 1879 (prior to the introduction of the bill), would have been part of the classified service but which downsized their workforce in order to remain within the unclassified service. Moreover, there is no evidence of districts manipulating the *growth* in their number of employees so as to remain below 50 employees: there are no districts that would have been above this cutoff had their employee numbers continued growing at the same rate after 1879 as they did from 1871 to 1879, but that ended up below 50 by 1883 (Panel (a) of Figure B8). Finally, Panel (b) of Figure B8 shows no evidence of districts bunching right below the 50 employees cutoff in 1883.

A second concern is that the outcomes in smaller districts could have been on a different trend relative to those in the larger districts. We present a number of tests for the common trends assumption. First, Table B1 reports the F-test p-values corresponding to the hypothesis that all of the pre-1883 event-study

[22] Our personnel results are nevertheless similar if we use data collapsed at the district-year level (Table B6).

[23] The list of "classified" districts remained the same throughout our baseline analysis period (up to 1894). In 1894, the classification was extended to include ports having as many as 20 employees. However, the cutoff was further reduced in May of 1896 to districts having as many as five employees (Commission, 1897). After this extension, only 63 employees in the entire Customs Service remained outside of the classified service. Hence, these later expansions are not useful to study the effects of the reform since by 1896 nearly all Customs Service's employees were within the classified civil service.

coefficients from equation 2 are equal to zero. Each row corresponds to one of our main outcomes, and each column corresponds to a different comparison group (based on a district's number of employees in 1883). Starting from districts with ten employees or more (column 3), we do not reject the hypothesis that all pre-reform coefficients are zero for all outcomes. With this in mind, we use districts with ten employees or more by 1883 as the main comparison group in our analysis (but we note our results are similar when using alternative control groups with fewer or more employees, see Table B2).

# 4 Results

## 4.1 Personnel Outcomes

A standard argument in favor of limiting political influence over the bureaucracy is that doing so will lead to improvements along three main dimensions: (1) reducing employee turnover, thus facilitating the accumulation of bureaucratic expertise; (2) improving employee qualifications by prioritizing skills over political connections; and (3) reducing patronage-related "excess" hiring, thus deflating expenses. In this subsection, we investigate whether reformed districts improved along the first two dimensions. We discuss if the reform reduced "excess hiring" in Section 5.

**Turnover**. As discussed in Section 1, the reform did not provide tenure to civil servants but rather just changed the appointment method in certain classified positions. Yet, limiting politicians' discretion to hire might have removed an important incentive to remove employees in the first place, thus potentially leading to lower turnover.[24] To assess whether this was the case, we estimate:

$$Turnover_{ict} = \alpha_c + \alpha_t + \beta Classified_c \times After_{ct} + \gamma X_{ict} + \epsilon_{ict} \quad (3)$$

where $Turnover_{ict}$ is an indicator that takes a value of one if employee $i$ in district $c$ who was listed in year's $t-2$ Register was no longer listed in year's $t$ (Registers were collected every two years).[25]

Table 1 shows that the average employee was less likely to leave (either through removal or resignation) after the reform. The reduction in turnover is sizable: employees in reformed districts were 12.6 percentage points less likely to be out of their job by the next Register, relative to a control group mean of 47%.[26] This finding implies that employees in reformed districts had a longer time horizon over which to accumulate bureaucratic expertise.

If the reduced turnover that we document was due to the reform, we should observe the strongest responses among employees in positions not exempted from exams. Column 2 presents evidence consistent with this prediction. In this column, we report the triple interaction between working in a position subject to exams and working in a classified district in the post-reform period. This interaction is negative and statistically significant: there is an additional 7 percentage points decline in turnover among individuals in these positions. Interestingly, however, there is also evidence of a decline in turnover even

[24] Indeed, the reports of the Civil Service Commission anticipated this possibility: "when a removal cannot be followed by the appointment of a favorite pressing for the vacancy to be made, most of the temptations to make unjustifiable removals are themselves removed."

[25] Since our personnel data starts in 1871, 1873 is the first year in which we observe if employees had left their job.

[26] For comparison, the *annual* turnover rate in the US federal government was 16% in 2019 https://www.bls.gov/news.release/jolts.t16.htm.

among workers in *exempted* positions.[27]

Similarly, we should also see the largest declines in turnover in years in which there was a party transition at the federal level. To test this hypothesis, we interact the $Classified_c$ indicator with a variable that takes a value of one in 1887 (the first Register year after the presidency went from Chester Arthur (R) to Grover Cleveland (D)) and in 1891 (the first after it went from Cleveland (D) to Benjamin Harrison (R)). Column 3 of Table 1 shows that the declines in turnover are indeed more pronounced in these years: the interaction term is negative and close to 18 percentage points. In contrast, the main effect is close to zero and statistically insignificant, suggesting that the reform did not significantly affect turnover in years with no party transitions. Figures 1a and 1b show a similar pattern (that is, a large negative gap between the treatment and the control groups in 1887 and 1891) both in the raw data or when we implement an event-study regression.

**Employees' Professional Background**. Classified districts had to hire some of their employees through open and competitive exams. We expect this change to lead to an improvement in employees' professional background if, in the pre-reform period, districts traded-off expertise for political loyalty when screening employees.[28] Alternatively, if hiring individuals with inadequate qualifications was costly for appointing officers, they might have placed a heavy emphasis on expertise even when hiring was discretionary (Brierley, 2019). Finally, the reform –by potentially increasing bureaucrats' job stability–might have increased the appeal of a career in government, thus leading to an improvement in the applicant pool.

To test if the reform led to improvements in employees' professional background, we use the data linking Customs Service personnel records to population censuses. Note that, since we only observe previous occupations for those employees that we can successfully link to an adult observation in the census, the sample size is smaller than in the previous exercise. We estimate:

$$y_{ict} = \alpha_c + \alpha_t + \beta Classified \times After_{ct} + \gamma X_{ict} + \epsilon_{ict} \tag{4}$$

where $y_{ict}$ is a characteristic of employee $i$ who worked in district $c$ in year $t$. In cases in which we link an employee to more than one census, we focus on the most recent census year (among those prior to Register year $t$).

In Panel (a) of Table 2, we focus on the probability that an employee was listed in the census as having a professional occupation. Professional occupations are those with a value of less than 100 in the 1950 Census Bureau occupational classification system. Examples of such occupations include accountants, lawyers and teachers.[29] Column 1 focuses on the *stock* of employees, which reflects a combination of employees hired before and after the reform, in exempted and non-exempted positions. In this case, we

[27] The Civil Service Commission reports mention that some collectors avoided removing workers even if they worked in positions exempted from exams, presumably due to concerns about employee morale: "The only exception to this rule is that afforded by Collector Saltonstall, at Boston, who apparently refused to make removals in the excepted and unclassified places of his office for reasons which would not warrant the removal of men from the classified places." (United States Civil Service Commission, 1890)

[28] For instance, Colonnelli *et al.* (2020) and Oliveros & Schuster (2018) find that politically-connected bureaucrats tend to be less qualified.

[29] Unfortunately, US censuses prior to 1940 do not include information on individual income or years of schooling. Hence, we cannot directly assess if the reform brought workers who were more highly educated or who earned more prior to joining the Customs Service.

observe a small increase in the likelihood that an employee would have previously held a professional occupation, which is not significant at the conventional levels.

In Column 2, we instead focus on the *flow* of newly hired employees. Here, we see a larger increase of 6.4 percentage points in the likelihood of previously holding a professional occupation, which is significant at the 5% level. In Columns 3 and 4, we continue to focus on new hires but further distinguish employees based on whether or not they were hired in positions that required an exam. Consistent with the exams helping select more qualified employees, we find the strongest effects among newly hired employees in positions that were subject to exams (Column 4). Specifically, newly-appointed employees in non-exempted positions were 9 percentage points more likely to have held a professional occupation. In contrast, we find a small and statistically insignificant effect among those hired in exempted positions (Column 3).

Panel (b) shows that the increase in the likelihood of hiring employees with a previous professional occupation was driven by a decline in the likelihood of appointing employees who in the census listed either none or an unskilled occupation.[30] Similar to the pattern in Panel (a), we find a relatively small decline in the share of such workers when focusing on the employee stock (Column 1), but a larger decline when focusing on new hires in non-exempted positions (Column 4).

Overall, these results suggest that the reform led to a professionalization of classified districts' workforces; employees had stronger professional backgrounds to begin with, as well as a longer time horizon over which to accumulate bureaucratic expertise.

### 4.2 Cost-Effectiveness in Revenue Collection

In this subsection, we use the data on districts' receipts and expenses to ask if the changes in personnel outcomes that we documented above translated into higher cost-effectiveness in customs revenue collection.

Figure 2 provides preliminary evidence suggesting that the reform did not lead to higher cost-effectiveness. The figure shows average log expenses (Panel (a)), log receipts (Panel (b)) and log receipts over expenses (Panel (c)) in classified and non-classified districts, from 1874 to 1893. Prior to the reform, both expenses and receipts evolved in a parallel fashion in both groups of districts. Moreover, the figure shows no clear trend break after 1883 neither with respect to expenses nor to receipts.

Table 3 confirms the pattern of limited effects of the reform on expenses and receipts. This table shows the results of estimating equation 1 using the same outcome variables as in Figure 2. In the even columns, we add region-year fixed effects to account for potential differential trends in economic activity across broad US regions (which would have influenced revenue levels). Columns 1 and 2 show no statistically significant effect of the reform on expenses: the effect size is very close to zero and statistically insignificant when we do not include region-time fixed effects, and negative but also statistically insignificant when we include them. These estimates allow us to rule out a reduction in expenses that was larger than 17%.

In Columns 3 and 4, we show that the reform did not lead to a statistically significant increase in total revenue. Specifically, the effect on revenue is small (0.025) and insignificant when not including

[30] Unskilled occupations are those with a value of 700 or more in the 1950 Census Bureau occupational classification system (for instance, laborers and janitors).

region-time fixed effects, and actually *negative* (but also statistically insignificant) when including them. These point estimates enable us to rule out an increase in total revenue that was higher than 21%.

Moreover, not surprisingly given the lack of effects on expenses and receipts, we also see no effects of the reform on our main measure of cost-effectiveness, log receipts over expenses (Columns 5 and 6). Specifically, we find a small and positive point estimate (that is statistically insignificant) when not including region-time fixed effects, and again a negative point estimate when we include them.

Figure 3 shows event-study estimates of the effect of the reform on each of these outcomes. The event-study results confirm the pattern of limited effects documented above: regardless of the outcome we consider, we find that none of the pre- or post-reform event-study coefficients are statistically different from zero. Indeed, the event-study coefficients do not show a pattern that would be consistent with a systematic improvement (nor with a deterioration) over time: they are sometimes positive and sometimes negative, both in the pre- and post-reform periods.

# 5 Mechanisms

Employees in classified districts stayed longer in their jobs and had stronger professional backgrounds. Yet, this professionalization did not translate into measurable improvements in customs revenue collection cost-effectiveness. This is despite the fact that, prior to the reform, there was a strong positive correlation between changes in districts' share of employees with a professional occupation and receipts collected (Table B10).[31]

## 5.1 Fixed Middle, Broken Tails? The Role of Non-Merit Hires

**Distortions in Districts' Personnel Structure**. In classified districts, employees paid at a rate of less than $900 a year were exempted from examinations and could be hired through traditional patronage methods. To the extent that appointing officers wanted to retain hiring discretion, we should observe an increase after the reform in the proportion of workers below this salary cutoff. Indeed, the reports of the Civil Service Commission warned that using a salary-based rule to determine which employees were subject to exams opened the room for this kind of distortion.[32]

Figure 4 shows evidence consistent with this prediction. The figure shows the proportion of employees making below $900 a year in classified and non-classified districts, before and after the reform. Panel (a) focuses on the stock of employees, whereas Panel (b) focuses on new hires. From this figure, it is evident that there was a sharp increase in the proportion of the employees below the $900 threshold after 1883, and that this increase was particularly pronounced among new hires.

Table 4 confirms that the reform led to a sharp increase in the share of workers in exempted positions. Specifically, employees in reformed districts were 18 percentage points more likely to work in an

[31] In this table, we use the pre-reform data to estimate a regression in which the dependent variables are districts' expenses, receipts and receipts over expenses, and the main independent variable of interest is the share of workers with a professional occupation. The regressions further include district and year fixed effects.

[32] "Turning to the custom-houses, the Commission is able to present much less satisfactory tables. The classification of the Customs Service has always been very imperfect. It has been classified by salary rather than by employment, and has been possible to take the employees out of the classified grades by lowering their salaries or by changing their designations" (United States Civil Service Commission, 1890).

exempted position (26 percentage points among new hires). For comparison, in the pre-reform period less than 15% of the employees in classified districts worked in positions that would continue not requiring exams after 1883.[33] Columns 3 to 6 further split exempted positions into those reserved for political appointees (such as the collector) and their direct assistants, and those exempted due to the $900 salary cutoff. The reform permitted only a limited number of political appointees, but did not limit the number of employees that could earn less than $900.[34] Indeed, the increase in the overall share of employees in exempted positions came exclusively from an increase in the proportion of employees below the $900 cutoff, with no change in the proportion of those in other exempted positions. Figure **??** presents event-study estimates confirming the finding of an increased proportion of the employee stock earning below the $900 cutoff after 1883.[35]

Two pieces of evidence suggest that this distortion might have led to a worsening in the cost-effectiveness of reformed districts' . First, Table B11 shows that the reform led to an actual *reduction* in the number of employees in non-exempted positions. Column 1 focuses on the total number of employees in a district, whereas Columns 2 and 3 distinguish between those in exempted and non-exempted positions. Column 2 shows that classified districts ended up with fewer employees in the more technical positions -such as clerks, examiners or inspectors- that were subject to exams, a 20% reduction (p-value: 0.11). At the same time, classified districts experienced a large increase in the number of workers in exempted positions, a 62% increase (p-value: 0.000) (Column 3). The fact that the reform did not lead to an overall reduction in the number of employees is not entirely surprising: districts' budget depended on an appropriation of Congress and the reform did not change this amount (Schmeckebier, 1924).[36] Hence, districts could reallocate any savings stemming from a reduction in patronage appointments in positions subject to exams toward patronage appointments in exempted positions (rather than toward non-personnel expenses).[37]

Second, the professional backgrounds of employees hired in positions paying less than $900 were, unsurprisingly, inferior to the backgrounds of those hired for higher-paying positions. Table B12 shows the association between an indicator that takes a value of one if an employee earned less than $900 and measures of professional background. Relative to other Customs Service employees, workers in these positions were less likely to have been employed in a professional occupation, more likely to have been employed in an unskilled one, and less likely to be literate.

Finally, Table B13 provides direct evidence showing that this distortion dampened the effects of the reform on personnel outcomes. In this table, we focus on the effects of the reform on turnover (Columns 1 and 2) and on the likelihood that an employee would have held a professional occupation (Columns 3

[33] Since the reform reduced turnover of employees in non-exempted positions, the increase in the proportion of new hires in exempted positions could be a mechanical consequence of this reduction. However, note that we also observe an increase in the *stock* of employees in exempted positions. Similarly, the fact that there is an increase in the proportion of workers making less than $900 among *new hires* suggests that the effects do not simply capture a reduction in the salaries of existing employees.

[34] Besides the collector, the following positions were filled by political appointees and their staff: deputy collectors; cashier of the collector; assistant cashier of the collector; auditor of the collector; chief acting disbursing officer; deputy naval officers; deputy surveyors; one private secretary or one confidential clerk.

[35] Online Appendix Section shows the robustness of this result to a number of alternative specifications similar to those described at the end of Section 4.1.

[36] This appropriation was set in $5,500,000 in 1871. Districts were also allowed to keep the receipts that they collected as "fines, penalties and forfeitures" and "labor, drayage and storage" for their operating expenses (Schmeckebier, 1924).

[37] Due to the higher average wage of merit hires (as compared to patronage appointments), each merit hire could be replaced by more than one low-paid non-merit hire without increasing total personnel expenses. Indeed, Column 1 in Table B11 shows that the point estimate on the total number of employees is positive (although statistically not significant).

and 4). In the even columns, we add *position* fixed effects, where a position is defined as the combination of an occupation and a salary (for instance, a $1200 clerk). When we add position fixed effects (thus fixing the composition of workers across positions), we find a *larger* reduction in turnover as well as a larger increase in the likelihood that an employee would have held a professional occupation. These findings suggest that the improvements in personnel outcomes would have been stronger had districts maintained the same occupational structure as in the pre-reform period.

**The Role of Collectors**. Each collection district was administered by a "collector of customs". Collectors had significant prerogatives, including the ability to appoint and remove employees.[38] However, the reform did not introduce any changes with respect to how collectors would be selected and appointed: until the abolition of this position in 1965, collectors continued to be political appointees, nominated by the President and confirmed by the Senate.

This continuity could be important in explaining the lack of improvements in cost-effectiveness. First, the fact that collectors continued to be political appointees likely facilitated the distortions in personnel structure documented above. Second, if a district's performance depended on the characteristics of the person at the top of its hierarchy (as argued in the historical literature, see Parrillo, 2013; Rao, 2016), then the lack of change in how this person was selected might have limited the reforms' ability to improve cost-effectiveness. Indeed, civil service reforms in US states and cities that maintained political discretion to select hierarchical positions of the bureaucracy have been associated with more limited improvements than those which also reduced discretion at the top (Ornaghi, 2016; Ujhelyi, 2014).

Although we cannot directly test if the lack of improvement in cost-effectiveness can be partly attributed to the lack of change in how collectors were selected, we can test a necessary condition for this hypothesis to be true: Namely, that the identity of collectors mattered for districts' outcomes. We implement two empirical tests to assess whether this was the case.

First, we follow Besley *et al.* (2011) and estimate "collector fixed-effects" from the following regression:

$$y_{lct} = \alpha_l + \alpha_c + \alpha_t + \epsilon_{lct} \tag{5}$$

where $y_{lct}$ is an outcome of district $c$ in year $t$ under the leadership of collector $l$. Testing the hypothesis that collectors do not matter is equivalent to testing whether the collector fixed effects $\alpha_l$ are all equal to zero.[39] Note that identification of $\alpha_l$ is possible since the data include collectors who served only for a subset of our sample years. Indeed, Panel (a) of Table 5 shows that the collector-fixed effects are highly jointly significant, regardless of the outcome (receipts, expenses or receipts over expenses) we consider. This finding suggests that collectors mattered for districts' financial outcomes.[40]

A concern with this test is that collectors' transitions might have been endogenous to districts' out-

[38] Collectors' functions included employing " proper persons as weighers, gaugers, measurers and inspectors at the several ports within his district".

[39] In this analysis, we exclude year-districts in which the collector changed and hence there were multiple collectors in a year; the "collector fixed effects" are not well defined in that case.

[40] These results are related to the literature estimating "leader effects" in other contexts such as CEOs in the private (Bertrand & Schoar, 2003) and public (Janke *et al.* , 2019) sectors, or sports coaches (Berry & Fowler, 2021). An important difference between our findings and those in this literature is that we do not observe a given collector in multiple districts. Hence, we are closer to Besley *et al.* (2011) who also identify leaders fixed effects out of leaders who stay in power for a subset of all the years in a given country.

comes. For instance, collectors might have been more likely to be replaced in districts in which cost effectiveness was already deteriorating. To assess this possibility, we collected data from the *Journal of the Executive Proceedings of the United States Senate* (Senate, 1875) on each nomination to the position of collector of customs, and the reason why a new collector had to be nominated (death, removal or suspension, resignation or end of term of the previous collector). Table B14 categorizes all transitions into each of these groups, whereas Figure B5 shows the number of each such transitions per year. The single most common reason for transitions (38% of cases) is term expiration. In particular, it was very likely that the terms of collectors would not be renewed if their term expired after the presidency had changed party hands. The second most common motive (32%) are cases in which the collector was removed or suspended. Similar to the case of term expirations, most removals also occurred in years in which the presidency changed party hands. This pattern suggests that both non-renewals and removals were mostly driven by political considerations and party affinity rather than by district performance.

To more formally evaluate the relationship between collector's turnover and districts' financial outcomes, we estimate:

$$Transition_{ct} = \alpha_l + \alpha_c + \beta y_{ct-1} + \epsilon_{ct} \tag{6}$$

where $Transition_{ct}$ takes a value of one if there was a change in the identity of the collector heading district $c$ in year $t$, and $y_{ct-1}$ are the total receipts, total expenses or receipts over expenses of district $c$ in year $t-1$. Consistent with the hypothesis that collectors' transitions were mostly driven by political consideration rather than by districts' performance, Table B15 shows that there is no correlation between past performance in revenue collection (or expenses) and the likelihood that a collector would leave or be removed in the following year.

Nonetheless, to further deal with the concern of endogenous collectors' transitions, we implement a second test in which we use deaths of collectors while in office as an exogenous shock to leadership, the same strategy as in Jones & Olken (2005).[41] This method compares average districts' outcomes in the $T$ years before and after a collector's death. The subindex $z$ represents a particular death:

$$\overline{PRE_z} = \frac{1}{T}\sum_{t}^{T} y_{zt}^{PRE}$$

$$\overline{POST_z} = \frac{1}{T}\sum_{t}^{T} y_{zt}^{POST}$$

Under the null hypothesis that the identity of collectors does not matter, there should not be any systematic difference in districts' outcomes around a collector's death, hence:

$$\overline{POST_z} - \overline{PRE_z} \sim N(0, \frac{2\sigma_{\epsilon i}^2}{T})$$

where the variance $\sigma_{\epsilon i}^2$ is district specific. To implement this test, we estimate:

$$y_{ct} = \alpha_c + \alpha_t + \beta_z PRE_z + \gamma_z POST_z + \epsilon_{ct} \tag{7}$$

[41] Figure A4 shows an example case in which a new collector ("T. Jefferson Jarrett") is nominated to replace a deceased collector ("Peter F. Cogbill") in the Petersburgh, VA district.

where $z$ indexes collectors' deaths. For each collector's death $z$, there is a separate set of $PRE_z$ and $POST_z$ dummies: $PRE_z$ is a dummy equal to one in the $T$ years prior to collector $z$'s death in that collector's district, whereas $POST_z$ is a dummy equal to one in the $T$ years after.[42] We follow Jones & Olken (2005) with respect to two choices. First, we *do not* include the year of the death itself in neither the $PRE$ nor the $POST$ dummies (so as to exclude any immediate effect of the potential disruption caused by collectors' deaths). Second, we include all districts in the sample (even those without collector's deaths), as doing so helps estimate the other parameters in the regression (for instance, the year fixed effects). Finally, since the death of collectors in office is a relatively rare event, we include all districts (regardless of their number of employees in 1883) and consider a longer time period than in our baseline analysis (1874-1903 rather than 1874-1893).[43] Our sample includes a total of 33 deaths over a 30 years period.[44]

This specification estimates separate coefficients $\beta_z$ and $\gamma_z$ for each collector's death. We use these estimates to construct a test of the equality of the mean of the outcome variables before and after all collectors' deaths. Specifically, we use the Wald statistic:

$$J = \frac{1}{N_z} \sum_{z=1}^{N_z} \frac{(\hat{\beta}_z - \hat{\gamma}_z)^2}{\frac{2\sigma_{\epsilon i}^2}{T}} \tag{8}$$

where $N_z$ is the total number of deaths that we include in our analysis. Under the null hypothesis, $N_z \times J$ follows a $\chi^2_{N_z}$.

Panel (b) of Table 5 presents the Wald tests for each of our measures of expenses and receipts. We reject the null hypothesis that collectors do not matter for expenses (p-value: 0.006). The J-statistic is 1.76, implying that the variance of expenses is 76 percent higher around collectors' deaths than what it would normally be. Similarly, the variance of receipts is also higher around collectors' deaths, although the evidence that the identity of collectors matter in this case is weaker statistically (p-value: 0.14). Finally, our main measure of performance (receipts over expenses) also exhibits excess variability around collectors' deaths (p-value: 0.085).

Overall, these findings suggest that the identity of collectors mattered for districts' outcomes. Hence, the lack of change in collectors' selection method might have constituted a missed opportunity to improve districts' cost-effectiveness.

# 6 Conclusions

Despite several attempts at reform, a well-functioning civil service remains an elusive goal for many developing countries (Schuster, 2017).[45] The historical experience of the US in its transition toward a

[42] In our baseline analysis, we use $T = 3$. We use a shorter time window before and after deaths than in Jones & Olken (2005) because our sample spans a shorter time period. We also include deaths that are close to the limits of the time window we study for which we observe less than three years before and after the death. In this case, the $PRE$ and $POST$ dummies are equal to one only for the minimum number of periods that we observe before and after. For example, if a death occurred in 1875, each dummy takes a value of one for only one observation before and after. There are only two deaths that fall under that category and the J-statistic continues to be statistically significant at the 5% level for expenses if we exclude those observations.

[43] In Section **??**, we use these data to investigate longer-term consequences of the reform.

[44] As a comparison Jones & Olken (2005) had 57 deaths in a 40 years period.

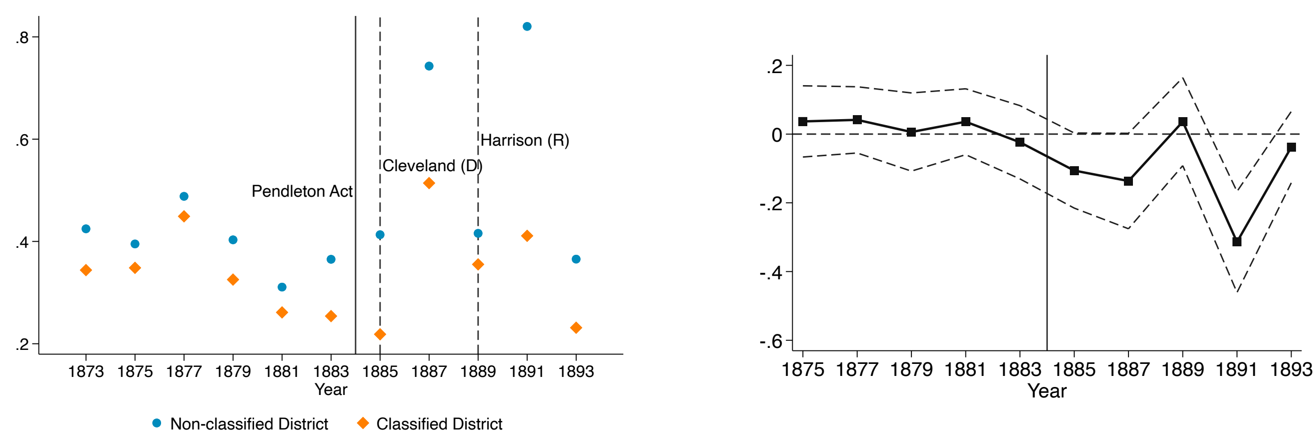
[45] Schuster (2017) shows that there is little association between the existence of a law mandating merit-based hiring in the public sector and survey responses regarding whether hiring in the public sector is actually meritocratic. Moreover, despite

professionalized bureaucracy offers a window into some of the challenges involved in these reforms. The US experience is particularly relevant from a development perspective, as it illustrates these challenges in a context where party patronage was "fully embedded in political reality" (Grindle, 2012).

This paper focused on the Pendleton Act, a milestone legislation that marked the emergence of a modern US civil service. Specifically, our analysis studied the consequences of the act for the functioning of the Customs Service, a key government agency in charge of the collection of customs revenue –the main source of federal revenue in late 19th century US. Our empirical strategy exploited the fact that the reform mandated exam-based hiring only in collection districts with 50 or more employees, enabling us to compare districts above and below such threshold before and after the reform. We find that the reform worked as its proponents expected in improving targeted employees' professional background and reducing turnover. However, there is limited evidence that it increased cost-effectiveness in customs revenue collection. Moreover, the reform had the unintended consequence of inducing districts to distort their personnel structure so as to retain discretion in hiring.

**FIGURE 1: PROBABILITY OF TURNOVER**

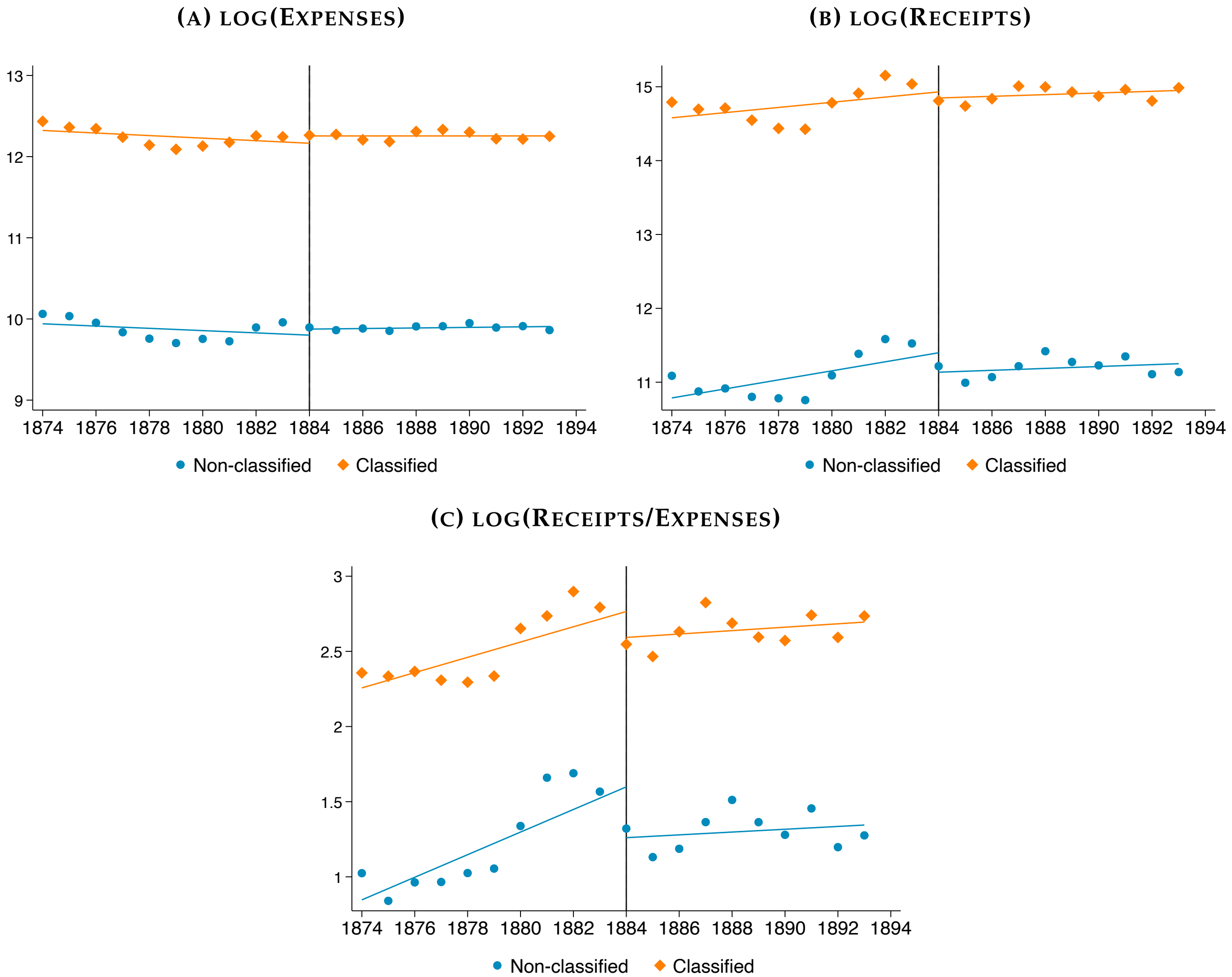
**(A) AVERAGE TURNOVER** **(B) EVENT-STUDY**




Notes: The y-axis shows the proportion of employees listed in the "Official Register" of year $t-2$ as working in district $c$ who were no longer listed in year's $t$. The dashed vertical lines correspond to years in which the Presidency went from a Republican to a Democrat or vice versa. The sample is restricted to districts with at least 10 employees by 1883. The figure in panel (b) shows the estimated coefficients corresponding to an interaction between a "Classified" indicator and year dummies. Standard errors clustered at the district level.

substantial heterogeneity in bureaucratic performance across countries, "meritocratic" civil services are, at least in theory, the norm: out of 117 countries with available data, 94 had laws establishing merit-based hiring in the public sector by 2015.

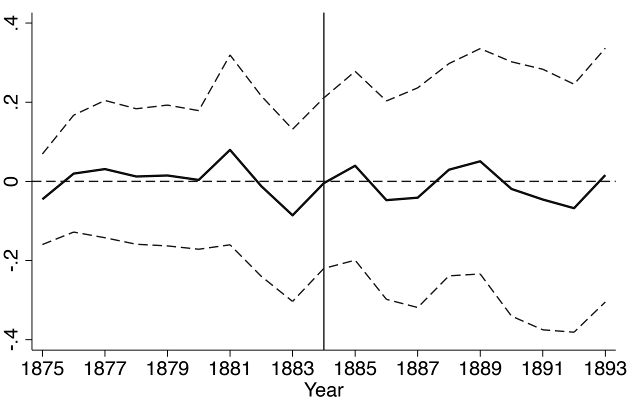
**FIGURE 2: AVERAGE EXPENSES AND RECEIPTS, 1874-1893**



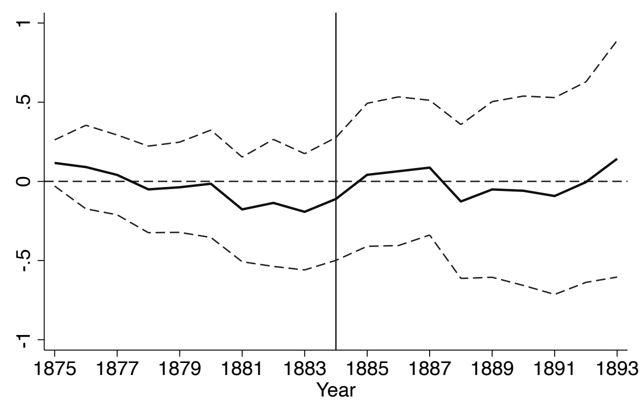

Notes: This figure uses the data on receipts and expenses from the "Annual Reports of the Secretary of the Treasury" (US Congress, 1874-1893). The figure shows yearly average log receipts (Panel (a)), log expenses (Panel (b)) and log of receipts over expenses (Panel (c)), separately for classified and non-classified districts. The sample is restricted to districts with at least 10 employees by 1883.

**FIGURE 3: EFFECTS OF THE REFORM ON RECEIPTS AND EXPENSES, EVENT-STUDY**

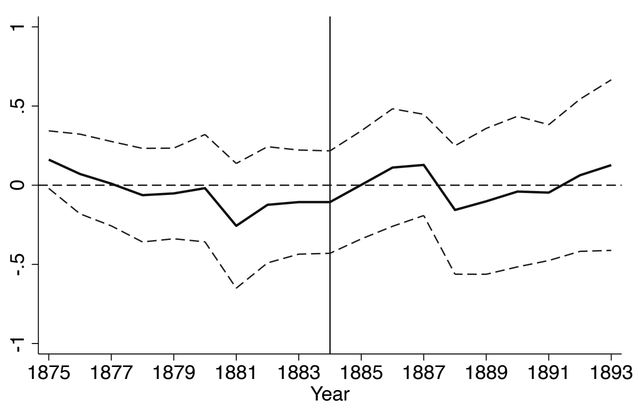
**(A) LOG(EXPENSES)**




**(B) LOG(RECEIPTS)**



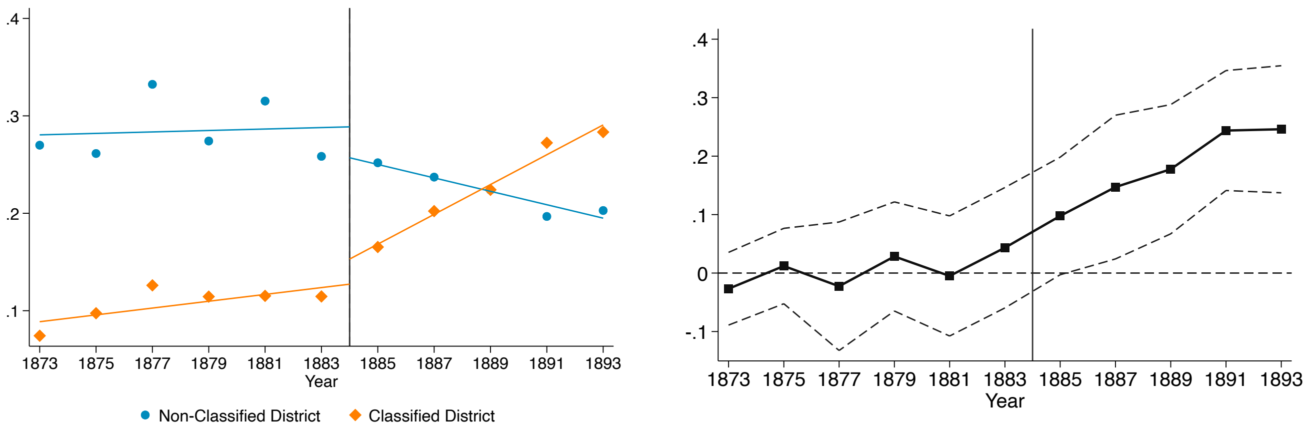

**(C) LOG(RECEIPTS/EXPENSES)**




Notes: This figure uses the data on receipts and expenses from the Annual Reports of the Secretary of the Treasury (US Congress, 1874-1893). The figure shows the estimated coefficients corresponding to an interaction between a "Classified" indicator and year dummies. The sample is restricted to districts with at least 10 employees by 1883. Standard errors clustered at the district level.

### FIGURE 4: SHARE OF EMPLOYEES BELOW THE EXAM CUTOFF

(A) AVERAGE | (B) EVENT STUDY




Notes: The y-axis shows the proportion of employees listed in the Official Register of year $t$ who made less than $900 a year (the cutoff above which employees were subject to exams in classified districts). The figure in panel (b) shows the estimated coefficients corresponding to an interaction between a "Classified" indicator and year dummies. The sample is restricted to districts with at least 10 employees by 1883. Standard errors clustered at the district level.

### TABLE 1: PROBABILITY OF TURNOVER

| | (1) | (2) | (3) |
|---|---|---|---|
| Classified X After | -0.126*** | -0.0834* | -0.0528 |
| | (0.0271) | (0.0473) | (0.0355) |
| Classified X After X Exam | | -0.0681* | |
| | | (0.0393) | |
| Classified X After X Party Turnover | | | -0.188*** |
| | | | (0.0701) |
| District FE | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| Observations | 41267 | 41267 | 41267 |
| Mean of dep. var. | 0.467 | 0.467 | 0.467 |

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. The dependent variable is an indicator that takes a value of one if employee $i$ who is listed in the "Official Register" of year $t-2$ as working in district $c$ is no longer listed in year's $t$ (the Registers were published every two years). $Classified \times After$ takes a value of one for districts that were part of the classified Customs Service after 1883. $Classified \times After \times Exam$ adds an interaction term that takes a value of one for employees working in non-exempted positions. $Classified \times After \times PartyTurnover$ adds an interaction terms for years in which the Presidency went from a Republican to a Democrat or vice versa. The sample is restricted to districts with at least 10 employees by 1883. Standard errors clustered at the district level.

### TABLE 2: EMPLOYEES' PROFESSIONAL BACKGROUND

#### (A) PROFESSIONAL OCCUPATION IN CENSUS

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Classified X After | 0.0278 | 0.0642** | 0.0368 | 0.0915*** |
| | (0.0176) | (0.0270) | (0.0448) | (0.0317) |
| District FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Observations | 7652 | 2766 | 911 | 1855 |
| Sample | All | New hires | New hires, no exam | New hires, exam |

#### (B) UNSKILLED OR NO OCCUPATION IN CENSUS

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Classified X After | -0.0434 | -0.0917** | -0.0726 | -0.109** |
| | (0.0313) | (0.0439) | (0.0741) | (0.0478) |
| District FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Observations | 7652 | 2766 | 911 | 1855 |
| Sample | All | New hires | New hires, no exam | New hires, exam |

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table uses the data linking the "Official Registers" to earlier population censuses. An observation corresponds to an employee-year. $Classified \times After$ takes a value of one for employees in. districts that were part of the classified Customs Service after 1883. The sample in column 1 of each of the panels includes the stock of Customs Service employees in year $t$. The sample in column 2 focuses on newly hired employees. Columns 3 and 4 further split new hires based on whether they work in a position exempted (column 3) or non-exempted from examinations (column 4). Panel (a) focuses on the likelihood that an employee is listed as holding a professional occupation in the census. Professional occupations are those with a value of less than 100 in the 1950 Census Bureau occupational classification system. Examples of such occupations include accountants, lawyers and teachers. Panel (b) focuses on the likelihood than an employee is listed as having an unskilled or no occupation in the census. Unskilled occupations are those with a value of 700 or more in the 1950 Census Bureau occupational classification system. Examples of such occupations include laborers and janitors. The sample in both panels is restricted to employees in districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

**Table 3: Effects of the Reform on Receipts and Expenses, Difference-in-Differences**

| | log(Expenses) | | log(Receipts) | | log(Receipts/Expenses) | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Classified X After | -0.0108 | -0.0827 | 0.0250 | -0.137 | 0.0358 | -0.0543 |
| | (0.0799) | (0.0600) | (0.185) | (0.174) | (0.137) | (0.138) |
| District FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Region X Time FE | No | Yes | No | Yes | No | Yes |
| Observations | 940 | 940 | 940 | 940 | 940 | 940 |

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. The dependent variable in columns 1 and 2 is the log of total expenses, in columns 3 and 4 it is the log of total receipts, and in column 5 and 6 it is the log of the ratio between total receipts and expenses. $Classified \times After$ takes a value of one for districts that were part of the classified Customs Service after 1883. Odd columns include year and district fixed effects. Even columns also include $Region \times Time$ fixed effects. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

**Table 4: Share of Employees in Positions Exempted from Exams**

| | No Exam | | Political Appointees | | Below cutoff | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Classified X After | 0.188*** | 0.259*** | 0.0101 | -0.0220 | 0.178*** | 0.281*** |
| | (0.0290) | (0.0314) | (0.0157) | (0.0169) | (0.0294) | (0.0305) |
| District FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 44922 | 15257 | 44922 | 15257 | 44922 | 15257 |
| Sample | All | New hires | All | New hires | All | New hires |

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. An observation corresponds to an employee-year. $Classified \times After$ takes a value of one for employees in districts that were part of the classified Customs Service after 1883. The odd columns focus on the stock of employees, whereas the even columns focus on the flow of newly hired employees. The dependent variable in columns (1) and (2) is an indicator that takes a value of one if an employee works in a position that was exempted from exams. Columns (3) to (6) further split exempted positions into those that correspond to leadership positions and their appointees (columns (3) and (4)), and those that are exempted because they are below the $900 classification cutoff (columns (5) and (6)). The sample is restricted to districts with at least 10 employees by 1883. Standard errors clustered at the district level.

**TABLE 5: COLLECTORS AND DISTRICT PERFORMANCE**

**(A) COLLECTOR FIXED EFFECTS**

| | (1) F-stat | (2) p-value | (3) Observations | (4) R-squared |
|---|---|---|---|---|
| log(Expenses) | 6.088 | 0.000 | 626 | 0.986 |
| log(Receipts) | 8.154 | 0.000 | 626 | 0.979 |
| log(Receipts/Expenses) | 6.310 | 0.000 | 626 | 0.944 |

**(B) COLLECTOR'S DEATH WALD TEST**

| | (1) J | (2) p-value(chi) | (3) $N_z J$ |
|---|---|---|---|
| log(Expenses) | 1.716 | 0.006 | 56.634 |
| log(Receipts) | 1.263 | 0.143 | 41.669 |
| log(Receipts/Expenses) | 1.353 | 0.085 | 44.656 |

Notes: Panel (a) estimates the specification in equation 5 and presents results of a F-test testing the null hypothesis that all collector fixed effects are jointly equal to zero. The sample is restricted to district-years in which there is only one collector– so as to be able to associate a performance metric to a single collector. Therefore, it excludes district-years where multiple collectors were in charge of the district for different months of the year (295 out of a total of 960 district years). Panel (b) presents the Wald test estimate defined in equation 7, testing whether there is excess variability of districts' financial outcomes around the collector's death. We use 33 deaths of collectors that occurred between 1875-1904.