Civil Service Reform and Organizational Practices: Evidence from the 1883 Pendleton Act*

Diana Moreira
UC Davis

Santiago Pérez
UC Davis and NBER

PRELIMINARY AND INCOMPLETE: DO NOT CIRCULATE

Click here for the most recent version

Abstract

How do civil service reforms affect the personnel outcomes and performance of government organizations? We focus on the 1883 Pendleton Act, which required US customhouses with 50 or more employees to select workers using competitive exams. Using a difference-in-differences strategy comparing reformed and non-reformed customhouses, we find that the reform led to reduced employee turnover and to an improvement in the professional background of new hires. However, these improvements did not translate into higher cost-effectiveness at revenue collection. A key mechanism explaining these results is that the reform induced a reallocation of workers towards exempted positions, ultimately undermining its effects.

* 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 comments of Brian Beach, Assaf Bernstein, Katherine Eriksson, James Feigenbaum, Michela Giorcelli, Walker Hanlon, Rick Hornbeck, Chris Meissner, Sarah Quincy, Monica Singhal, Chenzi Xu and Noam Yutchmann, as well by seminar participants at *Cooperacion Andina de Fomento*, CEPR STEG Workshop.

1 Introduction

A promise of civil service reforms that introduce merit as the main criteria for hiring is that doing so would enable governments to attract and retain more qualified employees. This professionalization, in turn, would translate into improvements in bureaucratic performance. While civil service reforms have been shown to improve bureaucratic performance in some contexts (Ornaghi, 2016; Xu, 2018), they appear to be no silver bullet: In an assessment of 71 civil service reforms funded by the World Bank, only 42% were rated as successful by an independent agency (Webb, 2008).[1] Moreover, these reforms have an a priori ambiguous effect on personnel outcomes, and direct evidence on whether they actually enable governments to hire and retain more qualified employees remains remarkably limited. Opening the black box of the bureaucracy is a necessary first step to understand the mechanisms underlying these varying degrees of success.

This paper studies the impacts of the 1883 Pendleton Civil Service Reform Act on the functioning of US customhouses. This act, which introduced open and competitive exams for the appointment of certain federal employees, is widely regarded as the first step towards the emergence of a professional bureaucracy in the US (Van Riper, 1976). The act established that customhouses with 50 or more employees by 1883 would need to select some of their employees (i.e. mid-tier bureaucrats) using competitive exams, and could no longer dismiss them on the basis of their political affiliation. However, it enabled the smaller customhouses to continue hiring using traditional patronage methods. We compare the functioning of reformed and non-reformed customhouses, before and after the passing of the act.

Our data enable us to measure if the reform led to changes in personnel outcomes, as well as to changes in the cost-effectiveness with which customhouses performed its main function, the collection of customs revenue.[2] First, we have digitized the personnel records of all employees of the US Customs Service from 1871 to 1893. These records include a person's name, place of employment, position and salary. We collected 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 have collected yearly data on the expenses and receipts of each US customhouse from 1874 to 1893.

Our identification strategy exploits the fact that only customhouses that had 50 or more employees by 1883 were required to hire through examinations. We use this feature of the reform to estimate a difference-in-differences model, comparing the functioning of customhouses that were subject to the reform to those that were not, before and after the passing of the reform. Our identification assumption is that, in the absence of the reform, the outcomes of customhouses that were above the 50 employees cutoff by 1883 would have evolved similarly to the outcomes of the customhouses that were below. Consistent with this assumption, we show that the outcomes of these

[1] "Despite the continued efforts and some modification of the approach, civil service reform has been relatively unsuccessful." "Quality of public administration (CPIA 15), which we take as civil service reform, had the lowest success rate, with fewer than 45 percent of borrowers in this area showing improvement." (Webb, 2008).

[2] By the early 1880s, the Customs Service collected more than half of the federal government's total revenue.

two groups of customhouses evolved similarly prior to the reform with respect to all of the main outcomes that we include in our analysis. Moreover, there are no differential changes in outcomes post 1883 when we estimate the "effects" of placebo reforms based on alternative employee cutoffs.

In the first part of the paper, we analyze the effects of the act on the personnel outcomes of reformed customhouses. We find that the reform worked as intended by proponents of civil service reform along two main dimensions. First, it led to a sizable reduction in employee turnover: customhouse employees were 12 percentage points less likely to leave their jobs (a 25% reduction) after the reform. This decline in turnover is particularly strong in years in which the Federal administration went from Republican to Democratic hands (and vice versa), suggesting that the reform was successful in reducing politically motivated dismissals. Also consistent with the decline in turnover operating through a reduction in patronage, we find that the decline is concentrated among workers in positions that were subject to exams (that is, those positions that the law explicitly protected from political influence).

Second, we find that the reform resulted in an improvement in the professional background of customhouse employees. Newly hired employees in reformed customhouses were less likely to report an unskilled or no occupation prior to joining the customs service, and were more likely to have held a professional occupation. This improvement took place *only* among employees who were hired in positions that were subject to examination. In contrast, there is a more modest improvement in the professional background of the *average employee*, which includes employees hired prior to the reform as well as employees hired after the reform but in positions exempted from exams.

We next ask if the reform led to increased cost-effectiveness in revenue collection.[3] We expect the reform to lead to improvements along this margin through three main channels. First, limiting the room for patronage might have led to reduced incentives to hire employees with the sole purpose of rewarding political loyalty, hence curtailing operating expenses. Second, hiring on the basis of merit could have reduced corruption by creating a separation between politicians and bureaucrats, thus making it harder to coordinate on corrupt behavior (Dahlström *et al.* , 2012).[4] This channel is likely to be particularly relevant in this context, given the discretion and corruption that are often involved in customs administration (Sequeira & Djankov, 2014). Third, to the extent that the reform induced a professionalization in their workforce, employees in reformed customhouses might have been in a better position to implement the complex procedures that characterize bureaucracies, as well as to revise existing (possibly inefficient) processes in place (Wilson, 2019).

Surprisingly, however, we find that the reform had very limited impacts on cost effectiveness. First, the reform did not lead to a significant reduction on customhouses total expenses: Our results enable us to rule out a reduction in expenses that was larger than 14%. In other words, the reform

[3] Increasing tariff revenue might have been negative for the US economy to the extent that it reduced international trade and increased domestic prices. In this analysis, we take the goal of collecting revenue as given, and ask whether the reform improved cost effectiveness.

[4] Indeed, meritocratic appointments have been empirically associated to lower corruption among public officials (Rauch & Evans, 2000; Dahlström *et al.* , 2012; Meyer-Sahling *et al.* , 2018).

did not seem to substantially reduce the incentives to inflate the bureaucracy and its operating expenses by rewarding loyalists with a job. Indeed, consistent with this lack of reduction in expenses, the reform did not result in a decrease in the total number of employees deployed in each customhouse. Second, we find no evidence that the reform led to increased revenue. Our point estimates are close to zero and are never statistically significant. In our preferred difference-in-differences specification, we can rule out an increase in revenue that was higher than 18%. Moreover, when estimating event-study specifications we see no evidence of higher gains in revenue towards the end of the sample period: the estimated effects are sometimes positive and sometimes negative, with no clear pattern that would suggest an improvement over time. Finally, as expected given the limited effects on both expenses and receipts, we also see no evidence of an improvement with respect to our main measure of performance, the "cost to collect one dollar".

In the last part of the paper, we explore the question of why revenue per dollar spent did not respond to the improvement in the professional background of employees and the reduction in turnover. We argue that a potentially important explanation for this pattern is related to the role played by *non-merit* hires after the reform. Specifically, the reform targeted employees in mid-tier positions, but exempted employees below a salary threshold as well as those employees at the top of a customhouse's hierarchy. This feature of the Pendleton Act is also present in other civil service reforms, in which patronage and merit-based hires are allowed to coexist within the same organization.[5] Indeed, Dahlström *et al.* (2012) argues that this "formal political discretion" could potentially help explain the lack of success of some of these reforms.

We start by showing that reformed customhouses changed their mix of positions in favor of those that could still be filled through patronage appointments. In particular, we document a sharp increase in the proportion of workers making $900 or less a year (who were exempted from exams), as well as an actual *reduction* in the number of bureaucrats in covered positions (such as clerks or examiners). This shift towards lower-paying positions is likely pernicious for the performance of reformed customhouses, both because it implies a distortion in their personnel structure but also because these employees had weaker professional backgrounds than those hired for higher paying jobs. Hence, while we do see an improvement in the professional background of the average employee, we find that this improvement is smaller than what it would have been had reformed customhouses maintained the distribution of workers across positions as in the pre-reform period.

Unlike mid-tier bureaucrats, "collectors of customs" (the persons at the top of a customhouse hierarchy) continued to be political appointees after the reform. This continuity can help rationalize the lack of improvements in cost-effectiveness through two main channels: First, since collectors were in charge of hiring, politically-aligned collectors might have facilitated the distortion in

[5] Scholars in public sector administration refer to the legal coexistence of patronage and merit hires as "formal political discretion" (Meyer-Sahling, 2006). For instance, Teso *et al.* (2019) documents the existence of patronage in Brazil, a country with civil service exams. Brassiolo *et al.* (2020) documents a similar pattern in the context of Ecuador. Brierley (2019) shows the coexistence of patronage and merit hires in the context of Ghana. See also Grindle (2012), which presents case studies of civil service reforms across the world and similarly documents the widespread coexistence of patronage and merit-based employees.

personnel structure discussed above.[6] Second, to the extent that collectors played a meaningful role in explaining the performance of the customhouses they administered (as argued in the historical literature, see Parrillo, 2013; Rao, 2016), the lack of change in how they were selected might have limited the reform's ability to improve cost-effectiveness. To assess whether collectors indeed mattered for customhouses' outcomes, we use deaths of collectors while in office as a shock to a customhouse leadership, a similar strategy as in Jones & Olken (2005) and Besley *et al.* (2011). Using this approach, we find that the identity of collectors had a substantial effect on customhouse revenue. This finding supports the hypothesis that the lack of change in the selection method for leadership positions might have played a role in explaining the reform limited effects on cost effectiveness.

A complementary explanation that we consider –in addition to that based on the role of non-merit hires– is that some of the changes in personnel outcomes that we document could have actually led to *lower* rather than higher revenue. For example, while reducing turnover would lead to accumulated expertise, it could have also lowered incentives to perform and facilitated collusion. While we do not observe direct measures of effort or collusion, we provide suggestive evidence that some of these opposite forces might have been at play in this context. Specifically, we find that there is a *positive* association between the fraction of employees who are newly hired and revenue, particularly in the more discretionary revenue categories such as fines. Moreover, we also find a positive association between a collector's death and a customhouse revenue. These findings suggest that, in this context, there might have been some adverse consequences of having a "protected" bureaucracy.

Finally, we discuss a number of alternative mechanisms for the absence of an improvement in cost-effectiveness (for which we find more limited evidence). Our results rule out that this lack of improvement was due to the fact that capitalizing the improvements in personnel outcomes required changes (for instance, replacing employees hired through the old patronage regime) that took more than 10 years to implement, or because the reform spilled over to the non-reformed customhouses.

Our data do not enable us to establish if the reform led to improvements in performance along margins other than revenue per dollar spent. For instance, customhouses 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).[7] Yet, while the "cost to collect one dollar" does not capture all potential dimensions of bureaucratic performance, it captures an important aspect of performance in the context of a government 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 higher cost to collect in the US relative to other developed countries).[8] Moreover, similar measures have been used by other

[6] Qualitative evidence suggests that this was indeed the case. See Section 2 for more details.

[7] Although we note that errors leading to higher import duties would have been more likely to be challenged by the merchants themselves.

[8] For instance, the Annual Reports of the Secretary of the Treasury (US Congress, 1874-1893) include such measure in

scholars studying the performance of government units in charge of revenue collection (Xu (2018), Khan *et al.* (2016), Naritomi (2019)). Finally, the share of workers with a professional background, which the reform increased, positively correlates with our cost to collect measure.

This paper makes three main contributions. First, we contribute to the literature on civil service reforms (Rauch *et al.* , 1995). In the US, state and local civil service reforms have improved electoral accountability (Folke *et al.* , 2011), reduced political electoral cycles (Bostashvili & Ujhelyi, 2019) and improved bureaucratic performance (Ornaghi, 2016).[9] Remarkably, however, there is limited evidence on the consequences of these reforms for the main objects that they are intended to change: namely, the personnel outcomes and characteristics of appointed bureaucrats. In contrast, we are able to show how such reforms affect *both* the personnel outcomes and the overall organization and performance of reformed units. We do so in an important context: the ability to collect revenue is a main determinant of state capacity, and customs remain an important source of revenue in many developing countries.[10] In addition, we empirically document a nuanced consequence of such reforms: when the reform is limited to mid-tier bureaucrats, politicians avoid the constraints to patronage by promoting greater hiring at the exempted tiers of the bureaucracy.[11]

More broadly, we also contribute to the growing literature on the personnel economics of the state (summarized in Finan *et al.* , 2017). A core focus of this literature has been on understanding the link between how government employees are screened and their subsequent performance on the job.[12] Our results show that changes in how some bureaucrats are screened might impact the delivery of public services not only through changes in the characteristics of the selected employees, but also by inducing changes in the personnel structure of the targeted organizations.[13]

Finally, we contribute to the literature on the Pendleton Act, a landmark legislation in US history (Theriault, 2003; Johnson & Libecap, 1994a,b; Libecap & Johnson, 2007). Specifically, our study provides the first quantitative evaluation of the changes brought by the Act which, as discussed in Johnson & Libecap (1994a), has not yet been possible due to lack of adequate data. Overall,

some of their editions. Claims that the "cost to collect" was higher in the US than in other countries that had adopted merit reforms were at the core of the argument of proponents of civil service reform (see section 2 for more details).

[9]There is a growing literature that studies the political gains of patronage practices in the context of developing economies (see for instance Teso *et al.* , 2019 and Brassiolo *et al.* , 2020). Xu (2018) shows that a reform that prohibited patronage in the appointment of leadership positions in the British Empire led to improved bureaucratic performance. We depart from Xu (2018) in two main ways. First, we look at how civil service reforms affect the *selection* of bureaucrats, while the research design in Xu (2018) exploits within-person differences in patronage connections (i.e. it keeps selection fixed). Second, we study how these reforms affect the personnel outcomes of an entire government unit (rather than just of those in leadership positions).

[10]For instance, out of 109 countries for which the World Bank reports customs revenue as a percent of total revenue, there are 39 where this proportion is above 10%, and 17 where it is above 20%. See https://data.worldbank.org/indicator/GC.TAX.IMPT.ZS. Moreover, customs collection that relies on the physical inspection of goods remains common in the developing world. Laajaj *et al.* (2019) studies a reform that introduced the computerization of import transactions in Colombia, and finds that it led to improvements in the outcomes of importing firms (suggesting a reduction in transaction costs).

[11]This finding is parallel to the result that state politicians in the US deviated spending towards local governments when civil service reform was passed at the state level (Ujhelyi, 2014).

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

[13]Our results are also related to the literature aimed at understanding the determinants of tax collection. See for instance Khan *et al.* (2016).

our results suggest a nuanced assessment of the consequences of the Act on the functioning of US customhouses; while the Act led to reduced politically-motivated dismissals and an overall professionalization of the workforce, there is more limited evidence that it resulted in more efficient revenue collection.[14]

2 Historical Background

2.1 The US Customs Service

The US Customs Service was established by the Tariff Act of 1789. Initially, the Act established 59 collection districts covering eleven US states.[15] By 1883, this number had been expanded to include a total of 117 collection districts spanning the entire US territory. Figure A1 shows a map of custom collection districts in 1883, together with a list of their associated "ports of entry".[16] The term "customhouse" was used to refer to the physical building where the customs transactions occurred, as well as to the administrative unit in charge of conducting these transactions.

The law established that each collection district would need to be administered by a "collector of customs". Collectors were appointed by the President for four-year terms and had to be confirmed by the Senate. The law provided collectors with significant prerogatives, including the ability to appoint and to fire employees.[17] Collectors remained to be political appointees until the abolition of this position in 1965.[18]

Revenue collected by the Customs Service accounted for the vast majority of federal revenue in the early decades of the 19th century, and continued to account for a significant proportion throughout the 19th century.[19] Most of the revenue generated by the Customs Service was due to the "collection of duties" on imported goods. Hence, to understand why changing the method for appointing employees could have potentially led to improvements in cost-effectiveness, it is important to understand the procedure through which these duties were collected in the late 19th century. The key take away of this subsection is that the process was complex and prone to errors and corruption along its several steps.[20]

[14] There is a substantial historical literature on the Pendleton act and the civil service reform movement. See for instance Hoogenboom (1968). Van Riper (1976) presents a history of the US civil service which includes a discussion of the Pendleton act. Finally, Skowronek *et al.* (1982) discusses the development of the American state in general and includes a discussion on the development of the civil service.

[15] In 1912, Congress authorized the President to reorganize districts, which led to a consolidation of some of the smaller districts (Schmeckebier, 1924).

[16] In most cases collection districts were named after their corresponding port of entry. Some collection districts had "ports of delivery" in addition to ports of entry. These were ports that were authorized to receive goods after they had first been entered into the country through a port of entry.

[17] The act established that a collector functions included "to employ proper persons as weighers, gaugers, measurers and inspectors at the several ports within his district, together with such persons as shall be necessary to serve in the boats which may be provided for securing the collection of the revenue to provide at the public expense". Some collection districts also had a "naval officer" and a "surveyor".

[18] https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title5a-node84-leaf158num=0edition=prelim

[19] The customs service remained the principal source of revenue until 1913, when the Sixteenth Amendment established the modern income tax.

[20] This section follows Schmeckebier (1924) closely.

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 amount of duties that would need to be charged on the shipment. In essence, the role of customs officers was to guarantee that anyone bringing goods into the US passed through a customhouse and paid duties according to law (Parrillo, 2013).

Import duties fell into three classes: specific, ad valorem, or both specific and ad valorem. In the case of some goods that were 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.[21] In other cases, a detailed examination of the goods' physical properties was required. For instance, the tariff on raw sugar depended on its saccharine content, thus requiring a chemical test to establish the correct amount to be levied (Schmeckebier, 1924).

Ad-valorem duties posed the greatest challenge in terms of accurately establishing the amount of duties. This is because, in the case of these goods, establishing this amount required an "appraisal" of the goods' monetary value. In our period of analysis, this value was determined (for custom collection purposes) with reference to the price of similar goods being commercialized in the market of the exporting country. Hence, to successfully implement this job, it was essential for appraisers to be aware of current market values so as to detect cases in which the importer attempted to undervalue the contents of a shipment.[22] According to Parrillo (2013), however, employees hired through patronage "were typically ignorant, sometimes unable to read the foreign languages in which invoices were written", making them poorly equipped to assess the accuracy of the invoices presented by importers.

After establishing their monetary value, goods had to be "classified" so as to determine the rate of duty that would need to be levied on them. Since some articles were not explicitly enumerated in the law, the Tariff acts stated that in those cases the duty charged 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) describes several examples of the difficulties in establishing the correct classification of some goods, and of how importers could take advantage of these ambiguities in order to minimize the amount of duties that they paid.[23]

2.2 US Customs Service before the Pendleton Act

Prior to the reform, hiring decisions were ruled by the "spoils system": Appointment to office was based primarily on political considerations and personal connections. As described by Aron (1987), "who an applicant knew counted at least as much as the skills he or she could demonstrate".

[21] Yet, Schmeckebier (1924) describes a scandal involving a large importer of sugar who had tampered the scales so as to reduce the amount owed in tariffs.

[22] Appraisers were supposed to gather this information by routinely checking trade publications (Schmeckebier, 1924).

[23] For instance, he describes the case of the importation of two pickle forks under the Tariff act of 1894. These forks could be plausibly assigned to three different classes of goods, each of which had a different duty (ranging from 35 to 45%).

Members of Congress played an important role as brokers of positions within their districts of influence (Fish, 1905).

Proponents of civil service reform targeted customhouses as a prime example of the dangers of patronage appointments.[24] In 1877, US Congress appointed a number of commissions to investigate the functioning of the major customhouses in the country. As part of its investigations, the *Jay Commission* (in charge of investigating the New York customhouse) compared the costs and functioning of US customhouses to those in other countries (House & of the Treasury, 1877). The commission concluded that the "cost of collection" in proportion to the amounts collected was much higher in the US than in Germany, England or France (where customs' employees were appointed through competitive examinations). The commission attributed this higher cost to over staffing (which inflated operating expenses), as well as to "errors and fraud" of patronage employees (which led to losses in revenue).[25]

The *Jay Commission* reports include several examples that highlight the inefficiency and corruption reigning in the New York customhouse prior to the Pendleton Act (House & of the Treasury, 1877). The customhouse was overstaffed, so that "many of the weighers and foremen rendered little, if any, service to the government, and that 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."[26] Employees' carelessness further contributed to reduced revenue, as "some of the employees didn't have brains enough to do the work, that some were incapacitated by ignorance and some by carelessness and indifference."[27] Similar examples can be found in the reports corresponding to the Philadelphia and New Orleans customhouses.[28]

The findings of the *Jay Commission* led in 1878 to the removal of Chester Arthur, the collector of the Port of New York (Newcomer, 1937).[29] The commission concluded by recommending a 20% reduction in the number of employees of the New York customhouse and the adoption of a merit reform. Indeed, the New York customhouse adopted a merit reform in March of 1879

[24] Reformers blamed patronage for substantial losses in government revenue: "How far are the losses of revenue due to the existing system of appointment at the request of political leaders and associations throughout the country?" (Sparks *et al.* , 1878).

[25] "Due to errors and frauds in the New York Custom-House. The investigating commission appointed by President Grant in l87l reported that the loss was probably twenty-five per cent. The New York Chamber of Commerce assured Sherman's commission that it had risen to forty per cent in l877." (Rogers, 1921).

[26] "Officer Cornell remarked on the violations of law by the acceptance of gratuities and by complicity in frauds".

[27] For instance, the report describes how weighers did not correctly perform their job, leading to lower revenue. "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)

[28] The commission in charge of investigating the Philadelphia customhouse asserted regarding the customhouse workforce that "Our conclusion that it can be somewhat reduced without injury to the service." The Philadelphia commission lists examples of corruption leading to lower cost-efficiency "as practices of taking the time of the Government for private business; of receiving presents of wines from the officers or agents of steamship lines; also the practices of delegating the appointing power and of making appointments, on political grounds, without sufficient assurance of the character and capacity of the appointee". The New Orleans commission similarly concluded that "the customs revenues at this port can be collected with a reduction of nearly 25 percent in the force employed".

[29] Ironically, Chester Arthur would later on become US president and sign the Civil Service Reform Act into a law in January of 1883. Arthur, who was elected as vice president in the 1880 elections, replaced James Garfield in 1881 after his assassination in the hands of a disappointed office seeker.

(four years prior to the Pendleton act) that introduced competitive exams for the appointment of certain employees (Brown, 1882). Our results, however, are similar when we exclude the New York customhouse from the analysis (see Figure B2 in the Appendix).

2.3 Civil Service Reform Movement and the Pendleton Act of 1883

During the course of the 19th century, the growth in the size of government drastically increased the cost of negotiating and monitoring patronage positions: From 1816 to 1881, the Federal government went from 5,000 to nearly 100,000 employees (Libecap & Johnson, 2007). Inspired by earlier reforms that took place in Europe, since the mid 1860s a civil service reform movement started pushing for "rational and bureaucratic means of staffing the federal departments" (Aron, 1987).[30]

While pressure was mounting for the adoption of a merit reform, the exact timing of the passing of the law is related to two political events. First, in July of 1881, newly elected 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 the reformists with a powerful example of the evils of the spoils system. Only three months after Garfield's death, Democratic senator George H. Pendleton introduced a bill with the aim of reforming the civil service. Second, Democrats took control of the House in March of 1882. Fearing that they would lose the 1884 presidential election, Republicans supported the civil service reform bill with the aim of protecting Republican office holders from politically-motivated dismissals (Hoogenboom, 1959). In January of 1883, president Chester Arthur signed the Pendleton Civil Service Reform Act into a law.

The Pendleton Act established a Civil Service Commission in charge of holding *"open, competitive examinations for testing the fitness of applicants"* (United States Civil Service Commission, 1883). Exams were aimed at testing practical knowledge relevant to an applicant's future position (rather than formal academic training), and could not include questions aimed at eliciting an individual's political orientation. To further ensure political neutrality, exams were graded by a "Board of Examiners" who did not know the identities of those who completed the exams.[31] After a vacancy occurred, the Civil Service Commission produced a list of the top four candidates from which the head of the Department (which, in the case of the customs service, would typically correspond to the collector) filling the vacancy would need to choose.[32] Figure A2 shows an example question of the arithmetics exam (one of the subjects that was required for individuals who applied for the position of clerk).[33]

[30] For instance, the United Kingdom established a Civil Service commission in 1855 in charge of holding "fair and competitive" examinations. An even earlier example of the use of exams for the appointment of state officials can be traced back to the Imperial Examinations in Imperial China.

[31] In the case of the classified customs service, these boards were appointed locally within each of the classified customs districts.

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

[33] Applicants to the position of clerk were examined in the following subjects: (1) Orthography, (2) Penmanship, (3) Copying, (4) Letter writing, (5) Arithmetic, (6) Elements of bookkeeping and accounts, and (7) Elements of geography, history and government of the US (Commission, 1886).

The law divided the “classified” (that is, subject to exams) Civil Service into three branches: the “Classified Departmental Service”, the “Classified Postal Service”, and the “Classified Customs Service” (our focus in this paper) . The Custom and Postal classified services were initially restricted to custom districts and post-offices who had at least 50 employees by 1883, and to employees making no less than $900 within these offices.[34] At the time of the enactment of the law, 11 custom districts met the 50 employees threshold, which meant that about 2,500 employees in the custom service were added to the classified service.[35] Figure 1 shows the location of US customhouses in 1874, distinguishing between those that would become part of the classified service by 1883 and those that would not. Finally, unclassified positions in the civil service remained to be filled via traditional patronage appointments. It is important to emphasize that the reform *did not* grant tenure to employees, but rather just prohibited politically-motivated dismissals.[36]

From 1883 to 1893, nearly 22,000 applicants completed an exam with the goal of joining the Customs Service, out of which about 2,800 had been appointed by 1894 (United States Civil Service Commission, 1893, p.240).[37] As a results, by 1893 nearly 60% of the workforce in the classified Customs Service had been appointed through open and competitive examinations.[38]

The conventional wisdom is that this reform improved the efficiency of 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”, and Johnson & Libecap (1994b) argue that: “There is evidence that the reforms introduced by the Pendleton Act improved the performance of federal workers in the positions that were covered by the law”.[39] The annual reports of the Civil Service Commission (arguably, an interested party) contain several instances in which customs collectors describe 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 argued that the reform had led to a 25% decline in the cost of collecting customs revenue, although the source of this calculation is unclear (Commission, 1897).[40]

[34] Customhouses that subsequently fell below the 50 employees cutoff would nevertheless have to remain as part of the classified service.

[35] These were New York City, 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.

[36] Later reforms to the civil service further increased the stability of federal government employment (Johnson & Libecap, 1994a).

[37] About 58% of the applicants obtained a passing grade in this period.

[38] The Customs Service employed an average of 4,200 employees in any given year during our sample period, out of which 3,300 were employees in the classified Customs Service.

[39] Another example of similar quotes is Fish (1905).

[40] The exact 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. If their estimate is correct, this item alone shows a saving of nearly $2,000,000 per annum” (Commission, 1897). It is not clear, however, how this number was calculated. First, as we show below, we find no evidence of a reduction in expenses in our difference-in-differences analysis. Moreover, a simple comparison of total expenses in 1897 to total expenses in 1883 yields an actual *increase* of 8% in total nominal expenses, which corresponds to an even higher increase in real terms since the US price level actually declined in this time period. Fish (1905) states (regarding the savings claimed by the Civil Service Commission) that “such definite statements, however, lack a firm basis”.

3 Data

Our analysis combines data from multiple sources. First, we have digitized customhouses' personnel records using the "Official Registers of the United States" (Department of the Interior, 1871-1893). This biennial publication was first published in 1816 and contains the name of every federal employee, their job title (including which government unit they were employed with), 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 employee-years in the US Customs Service, and span the 1871 to 1893 period. Figure A3 shows an example page listing employees of the New York's customhouse in 1883.

We collected additional information on the professional background of these employees by linking the Registers to US censuses of population, using name-based matching techniques (Abramitzky *et al.* , 2019). Specifically, we linked each of the 1871-1893 Registers to the 1850, 1860, 1870 and 1880 censuses.[41] Through this procedure, we are able to obtain information about an individual's occupation *prior* to his or her employment in the federal government.[42] While we provide further details on the linking strategy and sensitivity checks in Online Appendix section A.1, we note that: (1) there is no correlation between the reform and the likelihood of finding an individual in the census (Table A1), and (2) the results that *do not* require the Register-to-census linked data are very similar when we estimate them in this linked sample (Table A2).

Second, we have collected data on the expenses and receipts of each customhouse from 1874 to 1893.[43] Specifically, for each customhouse in the US we have collected data on: (1) its annual receipts (separately by source of revenue), (2) its annual expenditures, and (3) its number of employees, their occupations and the compensations that they received. 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 different branches of the US federal government, including the Customs Service. Appendix section **??** provides further details about each of the types of revenue that were collected by the Customs Service.

Finally, the data on collectors who died while in office come from the "Journal of the Executive Proceedings of the United States Senate" (Senate, 1875). This publication lists all nominations made by the President for the position of customhouse collector, as well as the reason why a new collector had to be nominated (death, removal, resignation or end of term). We searched these data for all instances in which a new collector was nominated due to the death of a previous collector. 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 customhouse.

[41] We chose these years since 1850 is the first US population census to list persons individually, and there are no surviving individual-level records for the 1890 census.

[42] Unlike the "Departmental Service" in Washington (Ziparo, 2017), the customs service employed extremely few women.

[43] We start from 1874 because the earlier data corresponding to some customhouses do not cover expenses and receipts for the entire fiscal year.

Table 1 shows some basic information on the personnel structure and receipts and expenses of customhouses, separately for the classified and the non-classified ones. Panel (a) focuses on statistics computed at the customhouse level (spanning 1874-1894), whereas Panel (b) focuses on the employee-level records (spanning 1871-1893). Revenue and expenses have a skewed distribution, with a few customhouses (in particular the New York customhouse) accounting for the majority of revenue and expenses.[44] Similarly, the distribution of the number employees is skewed: The median non-classified customhouse has about 10 employees, whereas the median classified customhouse has close to 200. Median wages are about 20% percent higher in the classified customhouses, and there is a smaller proportion of workers who made less than $900 a year. On average (across the entire sample period), 47% of the employees in non-classified customhouses and 34% in the classified customhouses who are listed in an Official Register in a given year were not listed in the following Register (our measure of employee turnover). Our differences-in-difference identification strategy (described in detail below) enables us to take into account these baseline differences in *levels*.

4 Empirical Strategy

In our main analysis, we estimate a difference-in-differences model comparing the outcomes of classified customhouses to those of non-classified customhouses, before and after 1883. Specifically, we estimate:

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

where y_{ct} is an outcome for customhouse c in year t, α_c are customhouse fixed effects, and α_t are year fixed effects. Our main variable of interest is $Classified \times After_{ct}$, which takes a value of one for those customhouses that were required to hire using exams after 1883. As described above, the classified customs service was initially limited to those customhouses employing at least 50 employees by 1883. The list of classified customhouses remained the same throughout our period of analysis (up to 1894), which implies that all our treated units received their "treatment" at the same time.[45] Throughout the analysis, we focus on those customhouses that were already in operation by 1874 and that remained in operation throughout the period. In some specifications, X_{it} include census region-year fixed effects. We cluster standard errors at the customhouse level.[46]

In addition to the difference-in-differences model, we also estimate event-study specifications in which we enable the difference between the control and the treatment groups to vary over time.

[44]As discussed above, in some robustness checks we exclude the New York customhouse from the analysis.

[45]In 1894, he classification was extended to include ports having as many as twenty employees. However, the cutoff was rapidly further reduced in May of 1896 to customhouses having as many as five employees (13th report, page 101). After this extension, only 63 employees in the customs service remained outside of the classified service (13th report, page 80). As a consequence, these later expansions of the classified customs service are not useful to study the effects of the reform since by 1896 nearly all employees of the customs service were within the classified system.

[46]There are not enough customhouses for us to estimate a regression discontinuity model using the number of employees in 1883 as the running variable.

We estimate:

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

where the β_t's capture the differential evolution in outcomes of the classified and non-classified customhouses during our sample period, and D_t are year dummies.

A first concern with our estimation strategy is that the 50 employees cutoff might have been chosen so as to include or exclude certain customhouses. Alternatively, customhouses 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, the list of customhouses that had 50 or more employees was nearly the same in 1879 (with the exception of Chicago's omission) as in 1883. In other words, there are no customhouses that would have been part of the classified service by 1879 (prior to the introduction of the bill in 1881), but which managed to reduce their number of employees so as to remain within the unclassified service. Moreover, there is no evidence of customhouses manipulating the *growth* in their number of employees so as to remain below the 50 employees cutoff: there are no customhouses that would have been above this cutoff had their number of employees continued growing at the same rate as it did from 1871 to 1879, but that ended up below 50 employees by 1883 (see panel (a) of Figure 2). Finally, panel (b) of Figure 2 shows no evidence of customhouses bunching right below the 50 employees cutoff. This figure shows the distribution of the number of employees across customhouses by 1883, focusing on those with at most 100 employees (so as to more easily visualize the distribution around the 50 employees cutoff).

A second concern with our identification strategy is that the outcomes of smaller customhouses (which were exempted from using exams) might have been on a different time trend relative to those of the larger customhouses. We present a number of tests for the common trends assumption. First, Table 2 reports the F-test p-values corresponding to the hypothesis that all of the pre-1883 event study coefficients from equation 2 are equal to zero (Columns 1-4). Each row corresponds to each of the main outcomes that we use in the analysis, whereas each column corresponds to a different comparison group (based on a customhouse number of employees in 1883). Starting from customhouses with 10 employees or more in 1883 (column 3), we do not reject the hypothesis that all coefficients are zero for most outcomes. With this in mind, we use customhouses with 10 employees or more by 1883 as the main comparison group in our analysis (but our main results on personnel outcomes are similar when use alternative control groups, see appendix table B3). To further deal with potential differential trends based on a customhouse size, we also estimate models in which we include interactions between a customhouse number of employees in 1883 and year dummies, thus allowing customhouses to be on differential trends based on their pre-reform size.

Finally, to further validate our empirical strategy, we estimate our baseline difference-in-differences specification using alternative "placebo" cutoffs to determine whether a customhouse would have been classified, and find no effects of these placebo reforms (see appendix table B2). We discuss a number of additional specification and robustness checks at the end of section 5.1.

5 Results

5.1 Personnel Outcomes in Reformed Customhouses

The standard argument in favor of protecting the bureaucracy from political influence is that this protection will improve personnel outcomes along three main dimensions: 1) reducing employee turnover which, in turn, facilitates the accumulation of bureaucratic expertise; 2) improving the quality of employees by prioritizing skills over political connections; 3) reducing patronage-related "excess" hiring, thus deflating operational expenses. In this section, we focus on understanding whether reformed customhouses improved along the first two dimensions. We discuss the evidence on whether the reform reduced “excess hiring” in section 6.

Turnover. We start by asking if the reform led to reduced employee turnover. As discussed in section 2, the reform did not provide full job security to employees, but rather just prohibited politically motivated dismissals. Hence, it is not obvious ex-ante that the reform would have led to reduced turnover. On the one hand, patronage employees who under performed in their jobs might have been less likely to be fired relative to merit hires (who were less likely to enjoy the protection of a political boss). On the other hand, prohibiting politically motivated dismissals should have removed an important channel of turnover.

To formally test if the reform led to changes in turnover, we estimate:

$$Turnover_{ict} = \alpha_c + \alpha_t + \beta Classified \times After_{ct} + \gamma X_{ict} + \epsilon_{ict} \tag{3}$$

where $Turnover_{ict}$ is an indicator that takes a value of one if employee i in customhouse c who is listed in year t's Official Register is no longer listed in the following Register (Registers were collected every two years).[47]

Table 3 shows that the average employee was less likely to leave after reformed customhouses switched to merit-based appointments. The reduction in turnover is sizable: employees in reformed customhouses were 12.6 percentage points less likely to be out of their jobs by the next register, relative to a control group mean of about 47%.[48] This finding implies that employees in reformed customhouses had a longer time horizon over which they could accumulate bureaucratic expertise.

If the reduced turnover that we document was due to the reform, we should observe that the effects are concentrated among individuals who worked in positions non-exempted from exams (since these are the positions that enjoyed protection from politically motivated dismissal). Col-

[47]Since our data on personnel records covers the 1871 to 1893 period, 1891 is the last year for which we are able to observe whether employees had left their job by the following Register.

[48]For comparison, the annual turnover rate in the US federal government was 16% in 2019 https://www.bls.gov/news.release/jolts.t16.htm. A turnover rate of 47% over a two years period implies an annualized turnover rate of about 27%.

umn 2 presents evidence consistent with this prediction. In this column, we report the interaction between working in a position subject to exams and working in a classified customhouse in the post-reform period. This interaction is negative and statistically significant: there is an additional 7 percentage points decline in the likelihood of turnover among individuals who were employed in positions subject to exams.

As discussed in section 2, "collectors", who were nominated by the President and confirmed by the Senate, had the ability to hire and remove employees. Hence, if the reduced turnover that we observe is due to employees being protected from political influence, we should see the largest declines in turnover in those years in which there was a party transition at the Federal level. Figure 3 shows graphical evidence suggesting that the reform led to reduced turnover, and especially so in years in which the Presidency went from a Republican to a Democrat or vice versa. The y-axis in this figure shows the proportion of employees who are listed in the Official Register of year t but who will no longer be listed on the following Register. The vertical dashed lines mark years in which the Presidency changed party hands.[49]

To more formally test this hypothesis, we interact the $Classified$ indicator with a variable that takes a value of one in 1885 (when the presidency went from Chester Arthur (R) to Grover Cleveland (D)) and in 1889 (when it went from Cleveland (D) to Benjamin Harrison (R)). Column 3 of Table 3 shows that the declines in the likelihood of turnover are particularly pronounced in those years in which the presidency switched party hands: the interaction term is negative and close to 18 percentage points. In contrast, the main effect is closer to zero and statistically insignificant, suggesting that the reform did not significantly affect turnover in years with no party transitions. Figure 4 shows a similar pattern (that is, a large negative gap between the treatment and the control groups in 1885 and 1889) when we implement an event-study regression.

Employees' professional background. Classified customhouses had to hire some of their employees on the basis of the results of open and competitive exams. We expect this change to lead to an improvement in the professional background of employees if, in the pre-reform period, customhouses traded off expertise with political loyalty when screening employees.[50] Alternatively, if hiring individuals with inadequate qualifications was costly for customhouse administrators, they might have placed a heavy weight on expertise even when hiring was discretionary (Brierley, 2019). Finally, the reform –by reducing politically motivated dismissals– might have increased the appeal of a career in government, potentially leading to an improvement in the applicant pool.

To test whether the reform led to an improvement in the professional background of customhouse employees, we use the data in which we link customhouses' personnel records to earlier population censuses.[51] We estimate:

[49]Unfortunately, we do not observe any party transitions in the pre-reform years for which we have data (1873-1883).

[50]For instance, Teso *et al.* (2019) and Oliveros & Schuster (2018) find that politically-connected public employees tend to be less qualified.

[51]Since we can only observe previous occupations for those employees that we can successfully link to the census, the sample size is smaller than in the previous exercise. We provide further details on linking in Online Appendix Section A.1.

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

where y_{ict} is a characteristic of employee i, who worked in customhouse c in year t. In all cases, we focus on the most recent census year for which we can observe person's i characteristics prior to register year t.

Each of the panels of Table 4 focuses on the occupational background of customhouse employees.[52] In panel (a), we focus on the probability that an employee is listed as having a professional occupation in the census.[53] Column 1 focuses on the stock of customhouse employees in a given register year. This stock reflects a combination of employees appointed prior to the reform and employees hired after, in exempted and non-exempted positions.[54] When focusing on the stock of employees, we observe a small increase in the likelihood that an employee would have held a professional occupation, which is significant at the 10% level.

In column 2, we instead focus on the *flow* of newly hired employees. This flow reflects a combination of workers hired in positions subject to exams and in positions that did not require one. Here, we see a larger increase of 6.3 percentage points in the likelihood of 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 they were hired in positions that required an exam or not. When focusing on employees hired in positions not requiring an exam (Column 3), we see a small and statistically insignificant increase in the likelihood that an employee would have held a professional occupation. Consistent with the effects of the reform operating through the exams selecting 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 positions non exempted from exams are 9 percentage points more likely to have held a professional occupation in an earlier census.

Panel (b) show that the increase in the likelihood of hiring employees with a previous professional occupation is driven by a decline in the likelihood of appointing employees who listed either no or an unskilled occupation in an earlier census.[55] Similar to the pattern in panel (a), we find a relatively small decline when focusing on the stock of employees (column (1)), but a larger decline when considering new hires in jobs subjects to exams. In this case, we also find a negative coefficient among new hires on exempted jobs, although the effect is statistically insignificant. Overall, these results suggest that classified customhouses became more conducive to the accumulation of professional expertise among their employees.

[52] Unfortunately, US censuses prior to 1940 do not include any information on a person's 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 workforce.

[53] 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.

[54] Employees hired prior to the reform in non-exempted were nevertheless also protected from political dismissal. This importation of patronage appointments into the classified system is often referred to as "blanketing in".

[55] 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.

Robustness of personnel results. Table B2 in the Online Appendix shows that the effects of the "reform" on personnel outcomes are all small and statistically insignificant when we use placebo cutoffs of 20, 30 or 40 employees (instead of 50) for the minimum number of employees above which a customhouse would have been classified. In this table, we focus on customhouses with less than 50 employees and estimate our main difference-in-differences specification using these placebo cutoffs.[56]

In Figure B1, we implement a randomization inference approach for computing p-values. Specifically, we estimate the effects of 1,000 placebo "reforms" in which we randomly choose 11 customhouses as being "classified". We then compare the estimated effects of these placebo reforms to the effects that we obtain when using the actual set of reformed customhouses in the estimation. Our estimated effects are always significantly larger in absolute value than the ones corresponding to the placebo reforms.

Since we have a relatively small number of classified customhouses, a concern is that the effects of the reform might have been driven by changes taking place in one specific classified customhouse (particularly in the very large ones since they account for a large number of observations when using the employee-level data). In Figure B2, however, we show that the results are similar when we estimate our baseline difference-in-differences specification excluding one classified customhouse at a time. The x-axis in this figure indicates the customhouse that we exclude from the regression, and the y-axis shows our estimated coefficients of the effect of the reform on each personnel outcome. The figure shows that the results are very stable regardless of which specific customhouse we exclude. Hence, our findings are unlikely to be driven by concurrent changes unrelated to the reform that took place in a specific customhouse.

In our baseline specification, our control group is comprised of customhouses with 10 or more employees by 1883. We use this control group since it has similar pre-trends than the classified customhouses with respect to all of our main personnel and cost-efficiency outcomes. However, Table B3 shows that we continue to find very similar results if we use alternative control groups with fewer (where we do not impose any restrictions on the minimum number of 1883 employees), or more (where we use a cutoff of 20+ employees) employees by 1883.

Finally, Table B4 shows that our personnel results are also robust to controlling for: (1) census region-year fixed effects, (2) interactions between the number of employees in 1883 and year fixed effects, or (3) both at the same time. These results make it unlikely that our findings would be driven by pre-existing differential trends between customhouses of different size or by differential trends across broad US regions.

5.2 Cost-Effectiveness in Revenue Collection

As discussed in section 2, proponents of civil service reform argued that the higher "cost to collect" in the US relative to other countries was primarily due to its lack of a meritocratic workforce. In this

[56]We restrict the sample to customhouses with fewer than 50 employees because otherwise the "placebo" treatment group would mechanically include the actual set of reformed customhouses (i.e. those with 50 or more employees).

subsection, we use the data on receipts and expenses to ask if the changes in personnel outcomes that we documented above resulted in improvements in the cost-effectiveness with which these customhouses collected revenue.

Figure 5 shows average yearly log expenses (panel (a)), log receipts (panel (b)) and log receipts over expenses (panel (c)) in classified and non-classified customhouses, from 1874 to 1893. Panels (b) and (c) include receipts from all sources. In all cases in this figure, the sample of non-classified customhouses is restricted to those that had at least 10 employees by 1883. The figures show that, prior to the reform, both expenses and receipts evolved in a parallel fashion in classified and non-classified customhouses. Moreover, the figure shows no clear break in the trend after 1883 neither with respect to expenses nor to receipts.

Table 5 shows the results of estimating equation 1 using the same outcome variables as in Figure 5. In the odd columns of this table, we add region-year fixed effects to account for potential differential trends across broad US regions. Columns 1 and 2 show that there is no statistically significant effect of the reform on customhouse expenses: the effect size is very close to zero and statistically insignificant when we do not include region-time fixed effects, and more negative but also statistically insignificant when we include them. These estimates enable us to rule out a reduction in expenses that was more than 18%.

In columns 3 and 4, we instead focus on a customhouse total revenue. We find no statistically significant evidence of an increase in total revenue in classified customhouses. Specifically, the effect on revenue is small (0.025) and insignificant when not including region-time fixed effects, and actually *negative* (but also statistically insignificant) when including them.

Finally, columns 5 and 6 focus on our main measure of cost-efficiency, log receipts over expenses. Consistent with the previous results, we see a small and positive point estimate (that is statistically insignificant) when not including region-time fixed effects, and a *negative* point estimate when we include them.

Figure 6 shows event-study estimates corresponding to estimating equation 2 using each of the outcomes discussed above as dependent variables. The omitted category in these figures are customhouses in 1874, the first year in our receipts and expenses data. The event-study results confirm the pattern of limited effects documented above: none of the pre-reform or post-reform event study coefficients are statistically different from zero (regardless of the outcome that we consider). Indeed, the event-study coefficients do not seem to show a pattern that would be consistent with a systematic improvement (nor with a deterioration): they are sometimes positive and sometimes negative, both in the pre- and in the post-reform periods. Indeed, Table B1 shows that we find similar effects if we estimate our baseline difference-in-differences model while excluding the first 5 years of post-reform data from the analysis (so as to allow for a longer time horizon over which the reformed customhouses could have implemented changes).

6 Mechanisms

Employees in classified customhouses stayed longer in their jobs and had stronger professional backgrounds. Yet, this professionalization did not translate into measurable improvements in the cost-effectiveness with which reformed customhouses collected revenue. In this section, we investigate possible reasons that might explain this lack of improvement. We find evidence that *non-merit* hires played an important role. First, the reform led to a large inflow of workers in exempted, lower paying positions. Second, the lack of change in the appointing procedure for leadership positions meant a missed opportunity to affect revenue. In contrast, we find no evidence that the reform spilled over to the non-reformed customhouses, or that capitalizing the improvements in personnel outcomes required changes that took longer to implement.

6.1 The role of non-merit hires

Distortions in customhouses personnel structure. The first hypothesis we investigate is that the reform might have led to distortions in customhouses personnel organization. In classified customhouses, 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 classified customhouses wanted to retain discretion in hiring, we should observe an increase in the proportion of workers below this salary cutoff after the reform. 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 manipulation.[57]

Figure 7 shows evidence consistent with this prediction. The figure shows the proportion of employees making below $900 a year in classified and non-classified customhouses, before and after the reform. Panel (a) focuses on the stock of employees, whereas panel (b) focuses on the flow of new hires. From this figure, it is evident that there was a sharp increase in the proportion of the employees below the classification threshold after 1883, and that this increase was particularly pronounced among newly hired employees (panel (a) versus panel (b)).

Table 6 formally tests the hypothesis that the reform might have prompted a change in the employee mix (so as to retain discretion in hiring). The odd columns focus on the stock of employees, whereas the even columns focus on the flow of newly hired employees. Columns (1) and (2) focus on the likelihood that an employee would be employed in any position that was exempted from exams. Both columns show a very large increase in the proportion of employees in exempted positions: employees in reformed customhouses were 18 percentage points more likely to be in an exempted position, and 26 percentage points more likely when focusing on new hires. For comparison, in the pre-reform period less than 15% of the employees in classified customhouses worked

[57] "Turning to the custom-houses, the Commission is able 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 the classified grades by lowering their salaries or by changing their designations. Thus, at Burlington, Vt., at the beginning of the administration of the head of the office appointed by President Cleveland there were twenty-one classified places subject examination. At the time the rule concerning excepted plans was changed in March, 1888, there were only three."(United States Civil Service Commission, 1890).

in positions that would keep not requiring exams post 1883.[58]

Columns (3) to (6) further split positions exempted from exams into two groups: Those that were exempted because they were reserved for political appointees (such as the collector) and their direct assistants, and those exempted due to being below the $900 salary cutoff. The reform permitted only a limited number of political appointees, but did not impose any limits on the number of employees that could earn below the salary cutoff.[59] Indeed, we see that the increase in the proportion of employees in exempted positions came exclusively from an increase in the proportion of employees below the salary cutoff, with very limited change in the proportion of employees in other positions exempted from exams. Figure 8 confirms the pattern of an increased proportion of the employee stock making below the $900 cutoff when we estimate an event-study regression.

Two pieces of evidence suggest that this distortion might have led to a worsening of reformed customhouses cost-efficiency. First, Table B5 in the appendix shows that the reform led to an actual *reduction* in the number of employees in positions that were subject to exams. Column (1) focuses on the total number of employees in a customhouse, whereas columns (2) and (3) distinguish between those in non-exempted and exempted positions. Column (2) shows that classified customhouses ended up with actually fewer employees in the more technical positions -such as clerks, examiners or inspectors- that were subject to exams, but saw a very large increase in the number of workers in exempted positions (Column 3). The fact that the reform did not lead to an overall reduction in the number of employees is not entirely surprising: Customhouses' budget depended on an appropriation of Congress and the reform did not change this appropriation (Schmeckebier, 1924).[60] Hence, customhouses had an incentive to reallocate any savings from a reduction in patronage appointments in positions subject to exams towards patronage appointments in exempted positions (rather than reallocating resources towards non-personnel expenses).

Second, employees hired in positions paying less than $900 were, not surprisingly, worse in terms of their professional background than employees hired for higher paying positions.[61] As a consequence, increasing the number of employees earning below this cutoff implied an inflow of employees with relatively worse professional background in the reformed customhouses. Table B6 shows the association between an indicator that takes a value of one if an employee earned less than $900, and measures of professional background. Employees in these positions were 2.8 percentage points less likely to have been employed in a professional occupation and 9 percentage

[58]Since the reform led to reduced turnover among employees in protected positions, a concern is that the increase in the proportion of new hires in exempted positions could be a mechanical consequence of this reduced turnover. However, the proportion of workers in exempted positions goes up also when considering the *stock* of employees.

[59]In addition to the collector, the following occupations were filled by political appointees and the staff of the appointee's choice: deputy collectors who do not also act as inspectors, examiners, or clerks; 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. These positions were exempted from exams.

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

[61]This result is consistent with the findings in Dal Bó *et al.* (2013), which shows that experimentally increasing the offered wages for public employees leads to an improvement in the applicant pool.

points more likely to have held an unskilled occupation relative to other customs service employees, had occupational income scores that were 6 percent lower, and were 2 percentage points less likely to be literate. For instance, since the reform increased the proportion of new hires below the $900 cutoff by 28 percentage points, in the absence of this increase the proportion of new hires with a professional occupation would have been 0.8 percentage points higher.

Overall, these findings suggest that the professionalization of reformed customhouses was in part outdone by a distortion in the customhouse personnel structure. This, in turn, might explain why the improvements in cost-efficiency were not as pronounced.

The Role of Collectors. As described above, each customhouse was administered by a "collector of customs". The reform did not introduce any changes with respect to how collectors were selected and appointed: collectors remained to be political appointees, nominated by the President and confirmed by the Senate.

This lack of change could be important for two reasons. First, the fact that collectors remained to be political appointees likely facilitated the distortions in personnel structure documented above. Indeed, civil service reforms in US states and cities that kept political discretion for hierarchical positions of the bureaucracy have been associated with weaker institutional changes (Ornaghi, 2016; Ujhelyi, 2014). Second, if the performance of customhouses depended on the characteristics of the person at the top of the hierarchy, then the lack of change in how this person was selected might have limited the ability of the reform to improve cost-effectiveness.

While we cannot directly test if the lack of improvement in cost-effectiveness was partly due to the lack of change in how collectors were selected, we can test a necessary condition for this hypothesis to be true: Namely, that collectors mattered for the outcomes of customhouses. To assess whether this is the case, we study the link between collectors and customhouses' outcomes . To deal with the potential endogeneity of appointment decisions, we use deaths of collectors while in office as a shock to a customhouse leadership, a similar strategy as in Jones & Olken (2005) and Besley *et al.* (2011).

The test we implement asks whether there are systematic differences in outcome associated with different leaders. Specifically, we employ the Wald test first proposed by Jones & Olken (2005). We explain the specification we use and then explain how these estimates feed into the Wald test. We estimate:

$$y_{ct} = \alpha_c + \alpha_t + \beta_z PRE_z + \gamma_z POST_z + \epsilon_{ct} \tag{5}$$

where y_{ct} is an outcome for customhouse c in year t, and z indexes collectors' deaths. Year and customhouse fixed effects are included through α_t and α_c respectively. For each collector death z, there is a separate set of PRE_z and $POST_z$ dummies. PRE_z is a dummy equal to 1 in the T years prior to collector z's death in that collector's customhouse. $POST_z$ is a dummy equal to 1 in the T years after leader z's death. In the baseline, we use $T = 3$.[62] We also follow Jones & Olken (2005)

[62] We also include deaths that are close to the limits of the 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

and do not include the transition year in neither the PRE nor the $POST$ dummies. We include all customhouses in the sample as this helps estimate the other parameters in the regression. This specification estimates coefficients β_z and γ_z for each death.

Under the null hypothesis that the identity of collectors does not matter for outcomes, we have:

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

$$\overline{PRE_z} = \frac{1}{T}\sum_t Y_{zt}^{PRE}$$

$$\overline{POST_z} = \frac{1}{T}\sum_t Y_{zt}^{POST}$$

Where the variance $\sigma_{\epsilon i}^2$ is customhouse specific. We construct Wald test statistic J. Based on the distribution of $\overline{POST_z} - \overline{PRE_z}$ under the null hypothesis:

$$J = \frac{1}{N_z}\sum_{z=1}^{N_z} \frac{(\overline{POST_z} - \overline{PRE_z})^2}{\frac{2\sigma_{\epsilon i}^2}{T}} \tag{6}$$

Then, under the null hypothesis $N_z \times J \sim \chi(N_z)$.

We found 56 deaths in all customhouses in the 30 years 1873-1903. We take several steps to select the collector's death events that we use in the analysis. We exclude 8 deaths that occur either at the very beginning or end of the sample, since for those we cannot estimate the corresponding PRE and POST dummies. There are an additional 15 deaths for which the outcomes are not available for at least one of the 30 years of the sample. We need many observations over the years to estimate the variance of $\sigma_{\epsilon i}^2$. Finally, we exclude an additional 5 deaths as they corresponds to customhouses that have less than 5 employees and leads to large outlier in the expenditure or receipts variables. For example, one of such customhouse has expenditure multiplied by 7 in a one-year horizon.

We use a total of 28 deaths in a 30 years period. As a comparison Jones & Olken (2005) had 57 deaths in a 40 years period.[63] Table 7 presents the Wald tests for the cost-effectiveness measures. We reject the null hypothesis that collectors do not matter for receipts (p-value:0.0004). The J statistic is 2.15, so the variance of the outcome is 115 percent higher around the collector's death than it would be normally. A collector's death does not seem to matter for expenditure (p-value=0.62).

Overall the results suggest that the lack of change in the selection of collectors contributed to the lack of improvements in cost-effectiveness. This result also highlights that decisions made at the customhouses have direct consequences for revenue. In other words, revenue is not entirely determined by business cycle, tariff structure and other elements not under the control of customhouses employees. Finally, these findings also suggest the importance of individuals (as opposed

number of periods that we observe before and after. For example, if a death occurred in 1974, each dummy will assume a value equal to one, for only one observation. The results are similar if we exclude those deaths.

[63]We are in the process of digitizing new data to expand the number of collector's deaths that can be used in the analysis

to organizations) as a source of state effectiveness (as highlighted in Best *et al.* (2017)).

6.2 Was there a trade-off between a "protected" and a "responsive" bureaucracy?

While patronage is often seen as resulting in an inefficient allocation of public resources (Teso *et al.* , 2019; Xu, 2018), it could in theory lead to better bureaucratic performance. First, allowing for discretion in hiring might improve selection if principals are able to use private information to screen applicants (Voth & Xu, 2019). Second, politically aligned bureaucrats might exert higher levels of effort.[64]

The Role of Reduced Turnover. A standard argument in favor of a stable bureaucracy is that such stability enables bureaucrats to develop expertise, which then translates into improvements in performance. However, long tenure in office could also lead to reduced employee motivation, as well as facilitating "collusive" corruption– a concern that is particularly relevant in the context of revenue collection. Indeed, Rao (2016) describes several instances in which customs officials used their discretion to impose lower or higher duties, often giving merchants the "benefit of the doubt".[65] As a consequence, whether stability on the job increases or decreases performance in this context is ultimately an empirical question.

We implement two tests to assess the relationship between employee turnover and customhouse revenue outcomes. First, we estimate the association between turnover (measured as the fraction of a customhouse' workforce who is newly hired) and receipts and expenses. Specifically, we estimate:

$$y_{ct} = \alpha_c + \alpha_t + \beta Share\,New\,Hires_{ct} + \gamma X_{ct} + \epsilon_{ct} \tag{7}$$

where α_c are customhouse fixed effects, α_t are year fixed effects, and $Share\,New\,Hires_{ct}$ is the fraction of a customhouse personnel that is newly hired (i.e. that shows up for the first time in year t Register). Since a higher share of new hires might just reflect growth in a customhouse total personnel, X_{ct} includes a customhouse number of employees in year t as an additional control variable.

The main concern with this test is that, even if we condition on customhouse and year fixed effects, the share of new hires could be endogenous to customhouses' revenue outcomes. For instance, employees might be more likely fired when these outcomes are worsening. Alternatively, a customhouse that is actively looking to improve its performance might both replace a higher fraction of its personnel and implement other changes.

To deal with this concern, we use a second test in which we exploit the exogenous variation in turnover that is generated by the death of collectors while in office. While a limitation of this exercise is that we are limited to considering turnover at the top of the hierarchy rather than across

[64]For instance, in his description of the consequences of the Pendleton Act, Fish (1905) argues that the merit system offers "the advantage of steady, light employment at a moderate remuneration and attracts the steady-going and uninmaginative".

[65]Similarly, Xu *et al.* (2018) find that, in the context of the Indian civil service, bureaucrats assigned to their home state have worst performance.

the full employee distribution, it has the advantage that in this case turnover is driven by reasons unrelated to revenue outcomes. We also note that, since collectors had the ability to hire employees, there is also a a strong association between the death of a collector and overall turnover (so the reduced form effect of a collector death includes the increased turnover that this death generates). We estimate:

$$y_{ct} = \alpha_c + \alpha_t + \beta Death_{ct} + \epsilon_{ct} \tag{8}$$

where $Death_{ct}$ is an indicator that takes a value of one if a collector died in year t.

Table B7 shows the result of these two sets of regressions. We find a *positive* association between our two measures of turnover (the fraction of new hires and whether the collector died while in office) and measures of revenue. The association is stronger when focusing on a more discretionary source of revenue, namely the amount collected through fines and forfeitures.[66]

Overall, these results suggest that the reduction in turnover induced by the reform might not have been by itself a source of improvement in outcomes; if anything, we tend to find a positive association between higher turnover and receipts.

6.3 Spillovers from classified to non-classified customhouses

One channel through which the reform might have affected revenue is by causing a diversion of trade away (or into) the classified customhouses. On the one hand, if the reform led to reduced "collusive" corruption, merchants who were interested in importing goods at a lower cost by engaging in corrupt behavior might have avoided the classified customhouses (thus reducing receipts). On the other, if the reform led to lower "coercive" corruption or greater efficiency and speed in processing imports, then merchants might have diverted their business into the classified customhouses (Sequeira & Djankov, 2014).[67] By a similar logic, the reform might have also affected personnel outcomes in non-classified customhouses (for instance, if workers that could not be hired for patronage jobs in classified customhouses were instead hired by non-classified customhouses).

To test this possibility, we analyze whether proximity to a classified customhouse led to different outcomes in the pre- and post-reform periods. Intuitively, if the reform led to diversion away from classified customhouses, we should observe that being close to a classified customhouse leads to *higher* receipts in the post-reform period. Alternatively, if the reform implied that classified customhouses attracted more business, we should see that proximity to a classified customhouse leads to *lower* receipts after 1883.

[66] Parrillo (2013) provides direct evidence that there was substantial discretion in the collection of fines. Specifically, he discusses how, during the 1860s, a change in the monetary incentives for imposing fines on merchants led to a nearly seven-fold increase in the annual amount of fines being charged.

[67] Sequeira & Djankov (2014) define "collusive" corruption as that that "occurs when public officials and private agents collude to share rents generated by the illicit transaction, thus reducing firm-level trade costs." and "coercive" as that that "takes place when a public bureaucrat coerces a private agent into paying an additional fee above the official price of the clearing service, which increases firm-level trade costs."

Figure B3 shows little evidence that the reform led to any diversion away or into classified customhouses. The figure (which uses the sample of non-classified customhouses) shows the correlation between a customhouse distance to the closest classified customhouse and its total revenues, before and after the reform. The figures shows that there is relatively little correlation between proximity to a classified customhouse and receipts in the pre-reform period and, more importantly, that this correlation does not seem to change after the reform. To more formally test for the presence of spillovers, we estimate:

$$y_{ct} = \alpha_c + \alpha_t + \beta log(Distance\, to\, Nearest\, Classified) \times After_{ct} + \epsilon_{ct} \tag{9}$$

where y_{ct} is an outcome of customhouse c in year t, $alpha_c$ are customhouse fixed effects $alpha_t$ are year fixed effects, and $Distance\, to\, Nearest\, Classified$ is the physical distance of a customhouse to the closest classified customhouse.

Table B8 shows that there is no evidence that would suggest spillovers in either direction: proximity to a classified customhouse does not predict either decreases or increases in a customhouse receipts after the reform. Similarly, Table B9 shows that we do not observe any spillover effects with respect to our main personnel outcomes (employee turnover, the proportion of workers in exempted occupations and the proportion of workers with professional background).

As an alternative approach to assess whether spillovers could explain our findings, we estimate our baseline difference-in-differences model but excluding from the control group those non-classified customhouses that were in close proximity to a classified customhouse (and hence were more likely to be affected by spillovers). Table B10 shows that excluding customhouses within a 50, 100 or 200 miles radius from a classified customhouse does not affect our conclusion of limited effects of the reform on cost effectiveness.

6.4 Adjustment costs

A final possibility is that the reform did not lead to improvements in cost effectiveness in its first ten years because reaping its benefits required changes that took longer to implement. For instance, fully replacing the employees who had been hired prior to the reform and who were added to the classified service ("blanketed in") despite not having been appointed through competitive exams. To test this possibility, we use data on receipts and expenses that span a longer period of time (from 1874 to 1903 rather than from 1874 to 1893).[68] Yet, using this longer time period we continue to find no evidence of an improvement in cost effectiveness; if anything, the point estimates become smaller in this expanded sample (Table B11).

[68]We note that we do not use this longer time period in our main analysis because, as discussed in section 2, the classified customs service was further expanded to include customhouses with 20 employees or more in 1894. The minimum number of employees was further reduced to only five in 1896, after which only 63 employees of the entire US customs service remained outside of the classified service. Hence, when using this expanded sample we lack a clearly defined "control group" post 1896.

7 Conclusions

Despite several attempts of reform, meritocratic civil services remain an elusive goal for many developing countries (Schuster, 2017).[69] The historical experience of the US in its transition towards a professional bureaucracy offers a window to understand some of the challenges involved in these reforms. The US experience is particularly relevant from a development perspective, as it illustrates the challenges of this transition in a context in which party patronage was "fully embedded in political reality" (Grindle, 2012).

In this paper, we focused on the consequences of the 1883 Pendleton Act, a milestone legislation that marked the emergence of a modern civil service in the US and still serves as the legal basis for its organization.[70] Specifically, we studied the degree to which moving towards merit-based hiring improved the functioning of US customhouses. Our identification strategy exploited the fact that only customhouses that had 50 or more employees by 1883 had to hire their employees through competitive exams, enabling us to compare the functioning of reformed customhouses to those that did not have to adhere to the merit reform. Our results show that the reform led to improvements in personnel practices: employees in reformed customhouses were more likely to stay longer in their positions and had stronger professional backgrounds. However, we find limited evidence that the reform resulted in gains in the cost-efficiency of revenue collection. We argue that part of the explanation for the lack of efficiency improvements is that the reform led customhouses to change their personnel structure so as to retain discretion in hiring, thus ultimately undermining the effects of the reform.

What are the broader implications of these findings? First, most countries lie on a continuum between public employees' appointments being purely based on merit and being purely based on patronage (Grindle, 2012; Brierley, 2019). Our results highlight the importance of understanding how civil service reforms interact with existing patronage structures. Second, they also highlight the potential dangers of a "piecemeal" approach to civil service reform. While this approach might be appealing to politicians because it enables them to retain discretion in the short run, such discretion might hamper the ultimate effectiveness of reforms.

References

ABRAMITZKY, RAN, BOUSTAN, LEAH PLATT, ERIKSSON, KATHERINE, FEIGENBAUM, JAMES J., & PÉREZ, SANTIAGO. 2019. *Automated Linking of Historical Data*. Tech. rept. NBER WP 25825.

ARON, CINDY SONDIK. 1987. *Ladies and gentlemen of the civil service: Middle-class workers in Victorian America*. Oxford University Press.

[69] Schuster (2017) reports that, out of 117 countries for which data are available, 94 had laws establishing merit-based hiring in the public sector by 2015. Schuster (2017) also shows that there is very little association between the formal existence of a law mandating merit-based hiring in the public sector and survey answers to the question of whether hiring in the public sector is actually based on merit.

[70] See for instance https://www.forbes.com/sites/tomspiggle/2020/10/28/trumps-executive-order-would-diminish-civil-service-employment-protections/4b28724a7b3a.

ASHRAF, NAVA, BANDIERA, ORIANA, DAVENPORT, EDWARD, & LEE, SCOTT S. 2020. Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review*, **110**(5), 1355–94.

BESLEY, TIMOTHY, MONTALVO, JOSE G, & REYNAL-QUEROL, MARTA. 2011. Do educated leaders matter? *The Economic Journal*, **121**(554), F205–227.

BEST, MICHAEL CARLOS, HJORT, JONAS, & SZAKONYI, DAVID. 2017. *Individuals and organizations as sources of state effectiveness*. Tech. rept. National Bureau of Economic Research.

BOSTASHVILI, D., & UJHELYI, GERGELY. 2019. Political Budget Cycles and the Civil Service: Evidence from Highway Spending in US States. *Journal of Public Economics*, **175**(C), 17–28.

BRASSIOLO, PABLO, ESTRADA, RICARDO, & FAJARDO, GUSTAVO. 2020. My (Running) Mate, the Mayor: Political Ties and Access to Public Jobs in Ecuador.

BRIERLEY, SARAH. 2019. Combining patronage and merit in public sector recruitment. *Journal of Politics*.

BROWN, W. 1882. *Civil-service Reform in the New York Custom-house*. Publications of the Civil-Service Reform Association. G.P. Putnam's Sons.

COMMISSION, UNITED STATES CIVIL SERVICE. 1884. *Report of the United States Civil-Service Commission*. US Government Printing Office.

COMMISSION, UNITED STATES CIVIL SERVICE. 1885. *Report of the United States Civil-Service Commission*. US Government Printing Office.

COMMISSION, UNITED STATES CIVIL SERVICE. 1886. *Report of the United States Civil-Service Commission*. US Government Printing Office.

COMMISSION, UNITED STATES CIVIL SERVICE. 1897. *Report of the United States Civil-Service Commission*. US Government Printing Office.

DAHLSTRÖM, CARL, LAPUENTE, VICTOR, & TEORELL, JAN. 2012. The merit of meritocratization: Politics, bureaucracy, and the institutional deterrents of corruption. *Political Research Quarterly*, **65**(3), 656–668.

DAL BÓ, ERNESTO, FINAN, FREDERICO, & ROSSI, MARTÍN A. 2013. Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, **128**(3), 1169–1218.

DEPARTMENT OF THE INTERIOR. 1871-1893. *Official Register of the United States*. US Government Printing Office.

DESERRANNO, ERIKA. 2019. Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda. *American Economic Journal: Applied Economics*, **11**(1), 277–317.

ESTRADA, RICARDO. 2019. Rules versus discretion in public service: Teacher hiring in Mexico. *Journal of Labor Economics*, **37**(2), 545–579.

FINAN, FREDERICO, OLKEN, BENJAMIN A, & PANDE, ROHINI. 2017. The personnel economics of the developing state. *Pages 467–514 of: Handbook of Economic Field Experiments*, vol. 2. Elsevier.

FISH, CARL RUSSELL. 1905. *The civil service and the patronage*. Vol. 11. Longmans, Green, and Company.

FOLKE, OLLE, HIRANO, SHIGEO, & SNYDER, JAMES M. 2011. Patronage and elections in US states. *American Political Science Review*, **105**(3), 567–585.

GRINDLE, MERILEE S. 2012. *Jobs for the Boys*. Harvard University Press.

HOOGENBOOM, ARI. 1959. The Pendleton Act and the civil service. *The American Historical Review*, **64**(2), 301–318.

HOOGENBOOM, ARI ARTHUR. 1968. *Outlawing the spoils: a history of the civil service reform movement, 1865-1883*. Vol. 50. University of Illinois Press.

HOUSE, UNITED STATES. CONGRESS., & OF THE TREASURY, UNITED STATES. DEPARTMENT. 1877. *Commissions to Examine Certain Custom-houses of the United States. Letter from the Secretary of the Treasury, Transmitting, the Report of Certain Commissioners Appointed to Examine Custom-houses, and Recommending Appropriations for Their Pay. October 25, 1877. – Referred to the Committee of Ways and Means and Ordered to be Printed*. Ex. doc.

JOHNSON, RONALD N, & LIBECAP, GARY D. 1994a. Patronage to merit and control of the federal government labor force. *Explorations in Economic History*, **31**(1), 91–119.

JOHNSON, RONALD N, & LIBECAP, GARY D. 1994b. The" Problem of Bureaucracy". *Pages 1–11 of: The federal civil service system and the problem of bureaucracy*. University of Chicago Press.

JONES, BENJAMIN F, & OLKEN, BENJAMIN A. 2005. Do leaders matter? National leadership and growth since World War II. *The Quarterly Journal of Economics*, **120**(3), 835–864.

KHAN, ADNAN Q, KHWAJA, ASIM I, & OLKEN, BENJAMIN A. 2016. Tax farming redux: Experimental evidence on performance pay for tax collectors. *The Quarterly Journal of Economics*, **131**(1), 219–271.

LAAJAJ, RACHID, ESLAVA, MARCELA, & KINDA, TIDIANE. 2019. The costs of bureaucracy and corruption at customs: Evidence from the computerization of imports in Colombia. *Documento CEDE*.

LIBECAP, GARY D, & JOHNSON, RONALD N. 2007. *The Federal Civil Service System and the Problem of Bureaucracy: The Economics and Politics of Institutional Change*. University of Chicago Press.

MEYER-SAHLING, JAN-HINRIK. 2006. The institutionalization of political discretion in post-communist civil service systems: The case of Hungary. *Public Administration*, **84**(3), 693–715.

MEYER-SAHLING, JAN-HINRIK, MIKKELSEN, KIM SASS, & SCHUSTER, CHRISTIAN. 2018. Civil service management and corruption: What we know and what we don't. *Public Administration*, **96**(2), 276–285.

NARITOMI, JOANA. 2019. Consumers as tax auditors. *American Economic Review*, **109**(9), 3031–72.

NEWCOMER, LEE. 1937. CHESTER A. ARTHUR: THE FACTORS INVOLVED IN HIS REMOVAL FROM THE NEW YORK CUSTOMHOUSE. *New York History*, **18**(4), 401–410.

OLIVEROS, VIRGINIA, & SCHUSTER, CHRISTIAN. 2018. Merit, tenure, and bureaucratic behavior: Evidence from a conjoint experiment in the Dominican Republic. *Comparative Political Studies*, **51**(6), 759–792.

ORNAGHI, ARIANNA. 2016. Civil service reforms: Evidence from US police departments. *Job Market Paper*.

PARRILLO, NICHOLAS R. 2013. *Against the Profit Motive: The Salary Revolution in American Government, 1780-1940*. Yale University Press.

RAO, GAUTHAM. 2016. *National duties: custom houses and the making of the American state*. University of Chicago Press.

RAUCH, JAMES E, & EVANS, PETER B. 2000. Bureaucratic structure and bureaucratic performance in less developed countries. *Journal of public economics*, **75**(1), 49–71.

RAUCH, JAMES E, *et al.* . 1995. Bureaucracy, Infrastructure, and Economic Growth: Evidence from US Cities during the Progressive Era. *American Economic Review*, **85**(4), 968–979.

ROGERS, EMMA. 1921. *Chester A. Arthur: Man and President*. University of Wisconsin–Madison.

SCHMECKEBIER, LAURENCE FREDERICK. 1924. *The Customs Service: Its History, Activities and Organization*. Johns Hopkins Press.

SCHUSTER, CHRISTIAN. 2017. Legal reform need not come first: Merit-based civil service management in law and practice. *Public Administration*, **95**(3), 571–588.

SENATE, UNITED STATES. 1875. *Journal of the Executive Proceedings of the Senate of the United States of America*. Green.

SEQUEIRA, SANDRA, & DJANKOV, SIMEON. 2014. Corruption and firm behavior: Evidence from African ports. *Journal of International Economics*, **94**(2), 277–294.

SKOWRONEK, STEPHEN, *et al.* . 1982. *Building a new American state: The expansion of national administrative capacities, 1877-1920*. Cambridge University Press.

SPARKS, J., EVERETT, E., LOWELL, J.R., & LODGE, H.C. 1878. *The North American Review*. American periodical series, 1800-1850, no. v. 127. O. Everett.

TESO, EDOARDO, COLONNELLI, EMANUELE, & PREM, MOUNU. 2019. Patronage and Selection in Public Sector Organizations.

THERIAULT, SEAN M. 2003. Patronage, the Pendleton Act, and the Power of the People. *The Journal of Politics*, **65**(1), 50–68.

UJHELYI, GERGELY. 2014. Civil service rules and policy choices: evidence from US state governments. *American Economic Journal: Economic Policy*, **6**(2), 338–80.

UNITED STATES CIVIL SERVICE COMMISSION. 1883. *Report of the United States Civil-Service Commission*. US Government Printing Office.

UNITED STATES CIVIL SERVICE COMMISSION. 1890. *Report of the United States Civil-Service Commission*. US Government Printing Office.

UNITED STATES CIVIL SERVICE COMMISSION. 1893. *Report of the United States Civil-Service Commission*. US Government Printing Office.

US CONGRESS. 1874-1893. Annual Report of the Secretary of the Treasury on the State of the Finances.

VAN RIPER, PAUL P. 1976. *History of the United States civil service*. Greenwood Press.

VOTH, HANS-JOACHIM, & XU, GUO. 2019. Patronage for Productivity: Selection and Performance in the Age of Sail.

WEAVER, JEFFREY. 2020. Jobs for sale: Corruption and misallocation in hiring. *Available at SSRN 3590721*.

WEBB, STEVEN BENJAMIN. 2008. *Public sector reform: What works and why? An IEG evaluation of World Bank support*. The World Bank.

WILSON, JAMES Q. 2019. *Bureaucracy: What government agencies do and why they do it*. Basic Books.

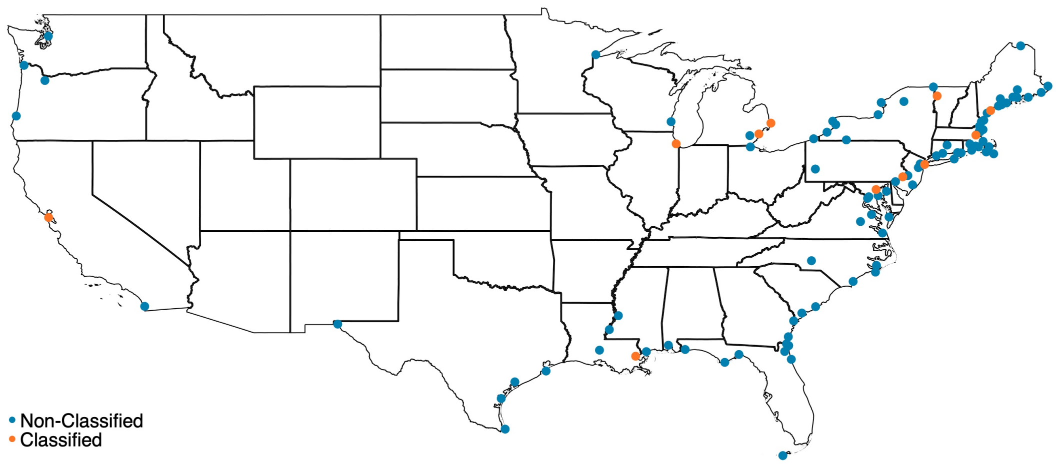
XU, GUO. 2018. The costs of patronage: Evidence from the british empire. *American Economic Review*, **108**(11), 3170–98.

XU, GUO, BERTRAND, MARIANNE, & BURGESS, ROBIN. 2018. *Social Proximity and Bureaucrat Performance: Evidence from India*. Tech. rept. National Bureau of Economic Research.

ZIPARO, JESSICA. 2017. *This Grand Experiment: When Women Entered the Federal Workforce in Civil War–Era Washington*. UNC Press Books.

Figures

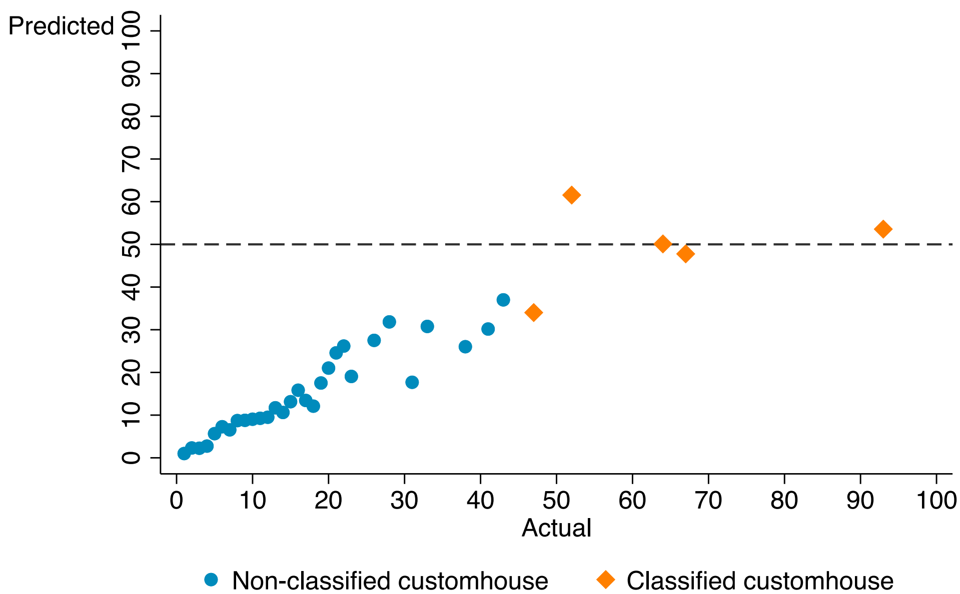
FIGURE 1: LOCATION OF US CUSTOM HOUSES



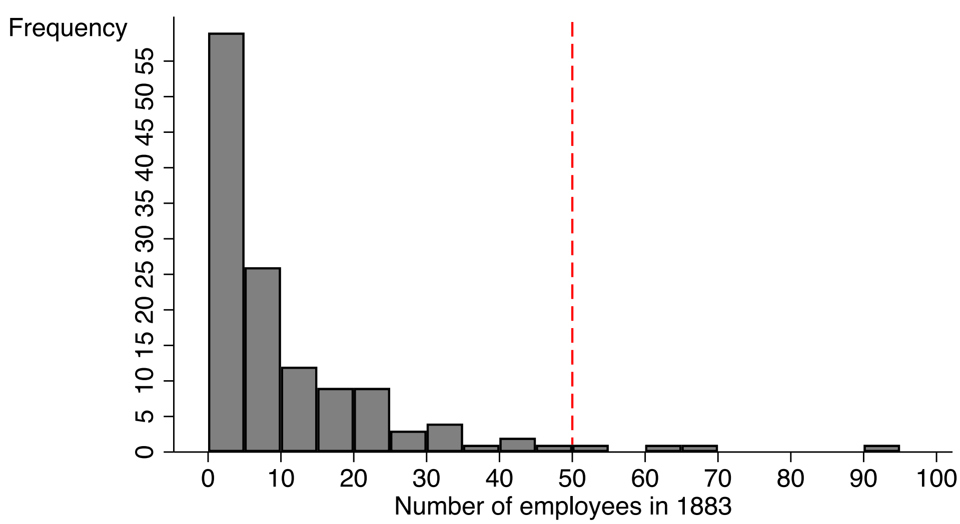
Notes: This map shows the location of the US customhouses that were in operation by 1874.

FIGURE 2: NO EVIDENCE OF MANIPULATION OF 1883 NUMBER OF EMPLOYEES

(A) PREDICTED VERSUS ACTUAL NUMBER OF EMPLOYEES IN 1883

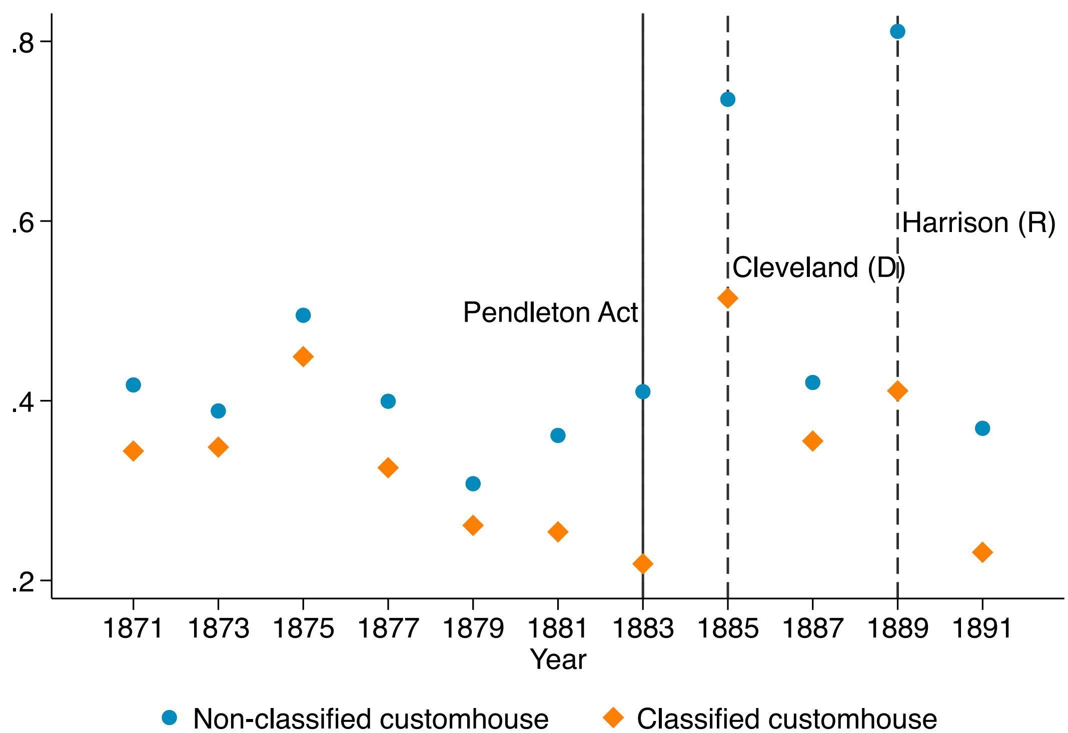


(B) DISTRIBUTION OF THE NUMBER OF EMPLOYEES IN 1883



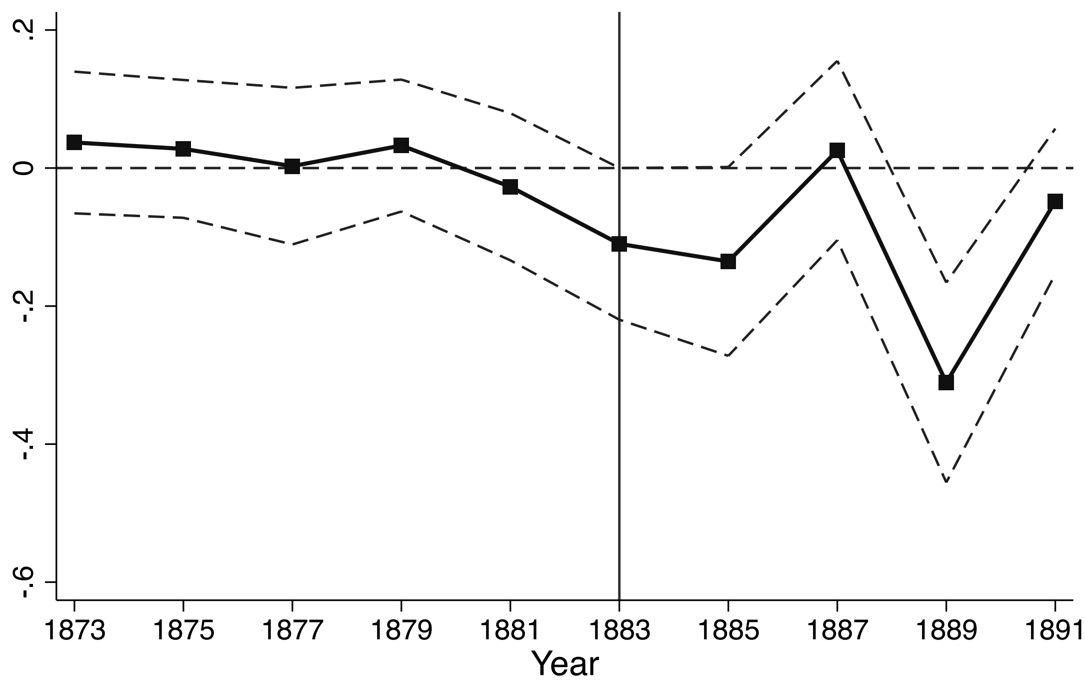
Notes: Panel (a) shows the actual (x-axis) and predicted (y-axis) number of employees of each US customhouse in 1883. The predicted number of employees is computed by extrapolating the 1879 number of employees using the 1871-1879 observed growth rate in their number. The figure shows that there are no customhouses that were predicted to be above the 50 employees threshold by 1883, but that ended up below. Panel (b) presents a histogram showing the empirical distribution of the number of employees across customhouses in 1883. The vertical dashed line corresponds to the 50 employees cutoff. The histogram is restricted to customhouses with at most 100 employees in 1883 so as to more easily visualized the distribution around the cutoff. Customhouses are grouped in bins of 5 employees (1-5, 6-10, etc.).

FIGURE 3: PROBABILITY OF TURNOVER



Notes: The y-axis shows the proportion of employees listed in the Official Register of year t as working in customhouse c who will no longer be listed on the following Register (in year $t+2$). The dashed vertical lines correspond to years in which the Presidency went from a Republican to a Democrat or vice versa.

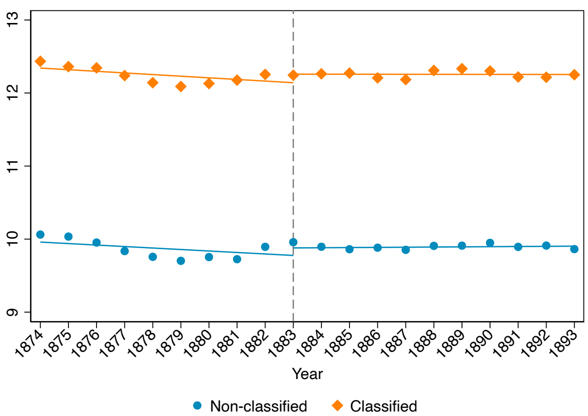
FIGURE 4: PROBABILITY OF TURNOVER



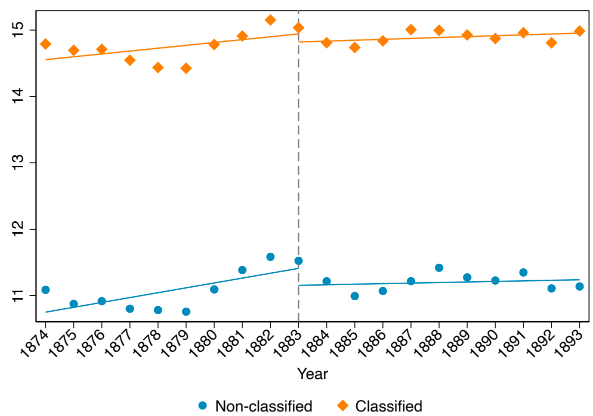
Notes: The dependent variable is an indicator that takes a value of one if employee i in customhouse c who is listed in the Official Register of year t as working in customhouse c is not listed on the following Register of year $t + 2$ (the Registers were published every two years). The figure shows the estimated coefficients corresponding to an interaction between a "Classified" indicator and year dummies.

FIGURE 5: EXPENSES AND RECEIPTS, 1874-1893

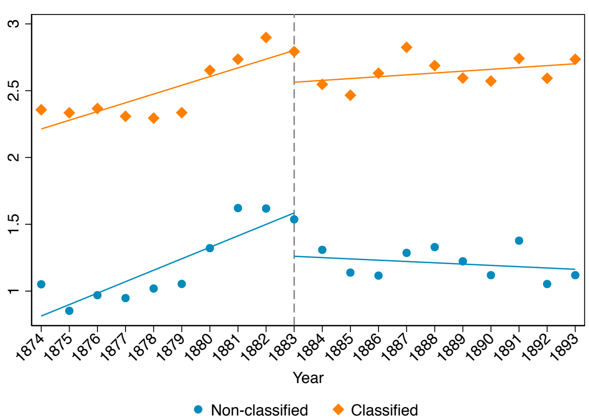
(A) LOG(EXPENSES)



(B) LOG(RECEIPTS)



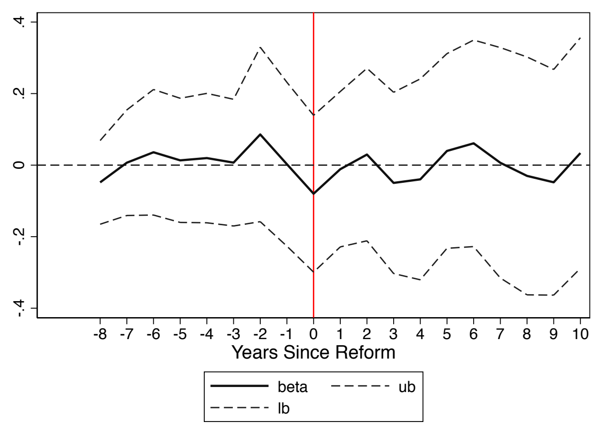
(C) LOG(RECEIPTS/EXPENSES)



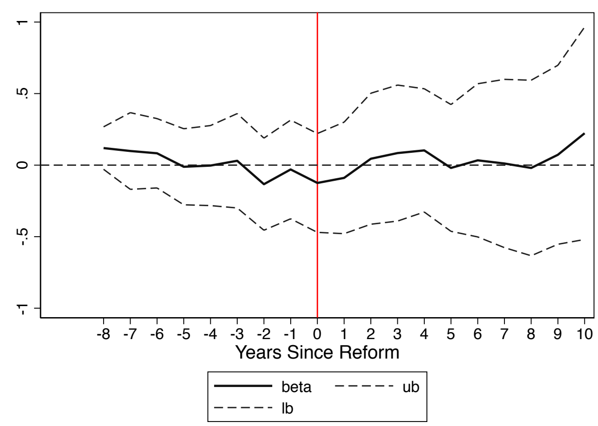
Notes: This figure uses the data on receipts and expenses from the Annual Reports of the Secretary of the Treasury (US Congress, 1874-1893). The figure shows yearly average log receipts (panel (a)), log expenses (panel (b)) and log of receipts over expenses (panel (c)), separately for classified and non-classified customhouses from 1874 to 1893.

FIGURE 6: EFFECTS OF CIVIL SERVICE REFORM ON RECEIPTS AND EXPENSES, EVENT STUDY REGRESSIONS

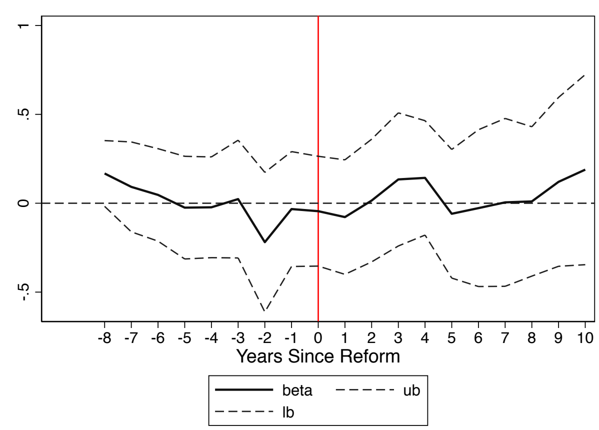
(A) LOG(EXPENSES)



(B) LOG(RECEIPTS)



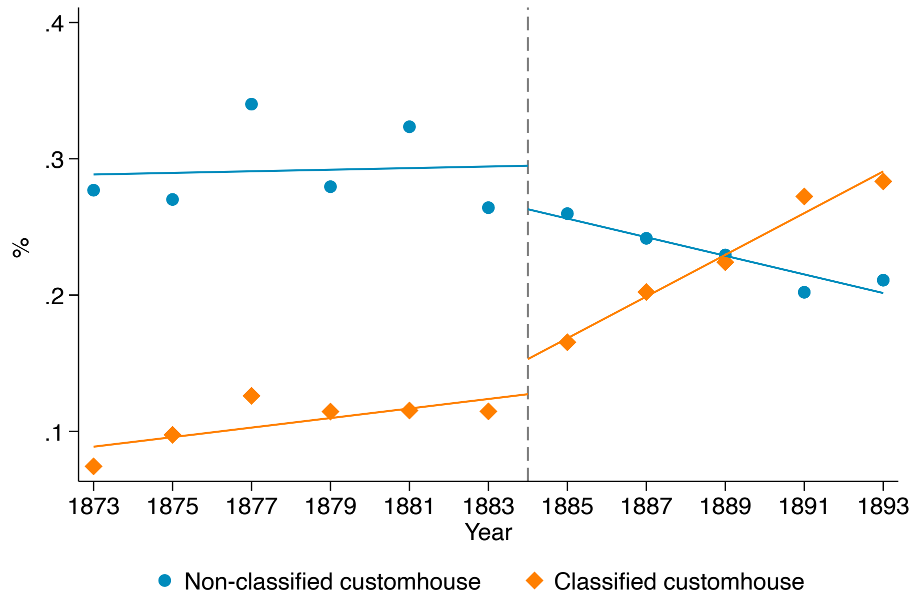
(C) LOG(RECEIPTS/EXPENSES)



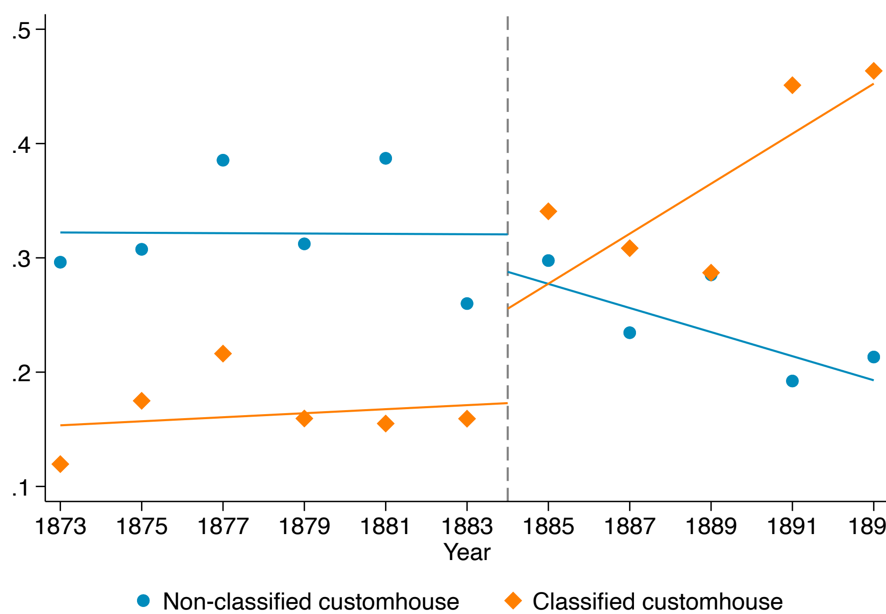
Notes: This figure uses the data on receipts and expenses from the Annual Reports of the Secretary of the Treasury (US Congress, 1874-1893). The figure shows event-study coefficients corresponding to estimating equation 2 in the main body of the paper.

FIGURE 7: SHARE OF EMPLOYEES BELOW THE EXAM CUTOFF

(A) STOCK OF EMPLOYEES

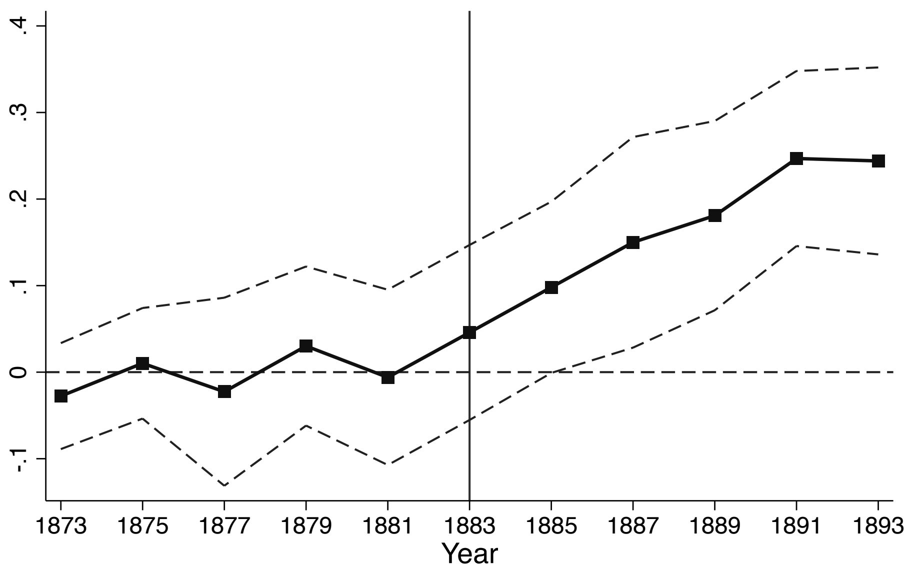


(B) FLOW OF NEWLY-HIRED EMPLOYEES



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 customhouses). Panel (a) focuses on the stock of employees in year t, whereas panel (b) focuses on the flow of newly hired employees in year t.

FIGURE 8: SHARE OF EMPLOYEES BELOW THE EXAM CUTOFF



Notes: The dependent variable is an indicator that takes a value of one if an employee works in a position that pays a salary below $900. The figure shows the estimated coefficients corresponding to an interaction between a "Classified" indicator and year dummies.

Table 1: Summary Statistics

	All Non-Classified			Non-Classified (10+ emp.)			Classified		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Customhouse-level statistics									
Total Expenses (000s)	14.66	8.55	15.16	22.94	18.52	16.68	458.58	239.60	718.29
Receipts (000s)	115.08	23.79	231.35	193.56	63.71	292.53	16172.67	2800.06	36472.89
Employees	14.10	10.00	14.96	21.04	18.00	17.61	331.21	194.00	468.05
# Observations	1520	.	.	800	.	.	220	.	.
# Custom houses	76	.	.	40	.	.	11	.	.
B. Employee-level statistics									
	1113.80	1095.00	793.55	1160.55	1095.00	746.20	1359.25	1277.50	671.38
	0.31	0.00	0.46	0.26	0.00	0.44	0.16	0.00	0.37
	0.47	0.00	0.50	0.46	0.00	0.50	0.34	0.00	0.47
# Observations	12268	.	.	8840	.	.	36683	.	.

Notes: This table shows summary statistics corresponding to the two main data sets that we use throughout the analysis. Panel (a) presents customhouse-level statistics from the "Annual report of the Secretary of the Treasury on the state of the finances" (US Congress, 1874-1893). An observation in this panel corresponds to a customhouse-year. These data cover the 1874-1893 period and are annual. Panel (b) is based on the customhouse personnel records collected from the "Official Registers of the United States" (Department of the Interior, 1871-1893). An observation in this panel corresponds to an employee-year. These data cover the 1871-1893 period and are biennial. Columns (1) to (3) show statistics corresponding to the full set of non-classified customhouses (i.e. those customhouses that were not required to hire through competitive exams after 1883) . Columns (3) to (6) show statistics for non-classified customhouses that had 10 or more employees by 1883. Columns (7) to (9) show statistics for the the classified customhouses (that is, those that required exams to some of their employees after 1883).

Table 2: F-tests for event study regression coefficients

(a) Customhouse-level results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean	P-value	Mean	P-value	Mean	P-value	Mean	P-value
log(Receipts/Expenses)	.051	.031	.04	.023	-.042	.4	.003	.056
log(Receipts)	.079	.11	.078	.112	-.04	.37	-.069	.249
log(Expenses)	.028	.112	.039	.111	.002	.55	-.072	.23
Comparison group	All	All	5+	5+	10+	10+	20+	20+
Observations	1740	1740	1580	1580	940	940	640	640

(b) Employee-level results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean	P-value	Mean	P-value	Mean	P-value	Mean	P-value
Annual compensation	27.857	.018	48.85	.017	63.021	.103	58.761	.461
% with compensation <900	.03	.241	.011	.214	-.003	.368	.008	.066
Turnover rate	.009	.308	.001	.423	.015	.527	0	.49
Comparison group	All	All	5+	5+	10+	10+	20+	20+
Observations (comp)	43881	43881	42867	42867	41015	41015	38003	38003
Observations (cutoff)	43941	43941	42925	42925	41070	41070	38046	38046
Observations (turnover)	44563	44563	43422	43422	41455	41455	38294	38294

Notes: This table reports p-values for the hypothesis that all the pre-1883 coefficients on the event-study specification are equal to zero. Each row corresponds to a different outcome variable and each column corresponds to a different estimation sample. Panel (a) focuses on outcomes measured at the customhouse level using data on receipts and expenses from the "Annual Reports of Secretary of the Treasury". Panel (b) focuses on employee-level outcomes using data from the "Official Registers of the United States".

TABLE 3: TURNOVER

	(1)	(2)	(3)
Classified X After	-0.126***	-0.0827*	-0.0569
	(0.0269)	(0.0469)	(0.0358)
Classified X After X Exam		-0.0692*	
		(0.0390)	
Classified X After X Party Turnover			-0.178**
			(0.0704)
Customhouse FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	41435	41435	41435
Mean of dep. var.	0.473	0.473	0.473

Notes: The dependent variable in columns 1 to 3 is an indicator that takes a value of one if employee i in customhouse c who is listed in the Official Register of year t as working in customhouse c is not listed on the Register of year $t+2$ (the Registers were published every two years). $Classified \times After$ takes a value of one for customhouses 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 positions that were non-exempted from examinations. $Classified \times After \times Partyturnover$ 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 customhouses with at least 10 employees by 1883. Standard errors are clustered at the customhouse level.

TABLE 4: PROFESSIONAL BACKGROUND OF CUSTOMHOUSE EMPLOYEES

(A) PROFESSIONAL OCCUPATION IN CENSUS

	(1)	(2)	(3)	(4)
Classified X After	0.0290*	0.0632**	0.0359	0.0905***
	(0.0170)	(0.0264)	(0.0426)	(0.0315)
Customhouse FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	7716	2787	927	1860
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	-0.0356	-0.0814*	-0.0451	-0.110**
	(0.0315)	(0.0447)	(0.0742)	(0.0469)
Customhouse FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	7716	2787	927	1860
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. The sample in column (1) of each of the panels includes the stock of employees in customhouse c in year t. The sample in column (2) focuses instead on newly hired employees in year t (i.e. those employees that are listed on the register of year t but who were not listed in year $t-2$). 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 or technical occupation in the census. Panel (b) focuses on the likelihood than an employee is listed as having an unskilled or no occupation in the census. The sample is restricted to customhouses with at least 10 employees by 1883. Standard errors are clustered at the customhouse level.

TABLE 5: EXPENSES AND REVENUE

	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)
Customhouse 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 natural log of the ratio between the total receipts and expenses. $Classified \times After$ takes a value of one for customhouses that were part of the classified customs service after 1883. Odd columns include Year FE and Customhouse FE. Even columns also include $Region \times TimeFE$. The sample is restricted to customhouses with at least 10 employees by 1883. Standard errors are clustered at the customhouse level.

TABLE 6: PROPORTION OF EMPLOYEES IN POSITIONS EXEMPTED FROM EXAMS

	No Exam		Above cutoff		Below cutoff	
	(1)	(2)	(3)	(4)	(5)	(6)
Classified X After	0.187***	0.260***	0.00798	-0.0231	0.179***	0.283***
	(0.0287)	(0.0310)	(0.0159)	(0.0170)	(0.0291)	(0.0302)
Customhouse FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	45101	15319	45101	15319	45101	15319
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. The odd columns focus on the stock of employees in year t, whereas the even columns focus on the flow of newly hired employees in year t. The dependent variable in columns 1 and 2 is an indicator that takes a value of one if an employee works in a position exempted from exams. Columns (3) to (6) further split exempted positions into those that are above the classification cutoff (columns (3) and (4)), and those that are below the classification cutoff (columns (5) and (6)). $Classified \times After$ takes a value of one for customhouses that were made part of the classified system after 1883. The sample in column 1 includes all customhouse' employees. Standard errors are clustered at the customhouse level.

TABLE 7: COLLECTOR'S DEATH WALD TEST (EXCLUDES TRANSITION)

	(1) J	(2) P-value(chi)	(3) ZJ
log(Expenses)	.8984	.6195	25.1538
log(Receipts)	2.1532	.0004	60.2886
log(Receipts/Expenses)	.862	.6744	24.135

Notes: This table implements the Wald test in Jones & Olken (2005) to assess if collectors mattered for customhouse outcomes.

A Data Appendix

A.1 Linking the Official Registers to the Census

Linking algorithm. In our baseline approach, we link individuals using information on their full names and place of birth. Unlike census-to-census links, we lack precise information on an individual's age that could be use to disambiguate between matches with similar names and places of birth.[71] In addition, we also lack direct information on an individual's gender (other than the information contained in names). As a result of this limitation, we are only able to uniquely identify individuals who are relatively unique based on their combination of place of birth and full name.

Given that we have less identifying information than is available in census-to-census linking (such as that described in (Abramitzky *et al.* , 2019)), we need to adjust our linking approach accordingly. Our linking algorithm has the following steps:

1. Clean names in the Registers and the Census to remove any non-alphabetic characters and account for common misspellings and nicknames (e.g. so that Ben and Benjamin would be considered the same name).

2. For each individual in the Register, search for a potential match in the Census. Potential matches are individuals who:

   (a) Report the same place of birth (states for the US born, country for foreigners). We exclude observations in the Official Registers which lack information on birthplace (about 1.5% of all observations).[72]

   (b) Have a reported age in the census such that they would have been between 18 and 65 years old at the time of the Register (for instance, when linking the 1881 Register to the 1850 Census we only look for individuals aged 0 to 35 in 1850).

   (c) Have a first name and a last name within a Jaro-Winkler distance of c_1, where $c_1 \in [0, 1]$. The Jaro-Winkler distance is a string distance measure such that a value of zero corresponds to two identical strings and a value of one corresponds to two strings with no common characters. We allow for non-identical strings to be considered a match to deal with transcription errors in the Census and for OCR errors in our digitization of the Official Registers. Intuitively, the lower the value of c_1 the more conservative our linking approach (i.e. the lower the number of cases we will match someone to an incorrect individual).

   (d) There is no other potential link with a first name and a last name within a Jaro-Winkler distance of c_2, where $c_2 \in [c_1, 1]$. That is, we impose that, if the closest individual is within a Jaro-Winkler distance of c_1, the second closest potential match needs to be at a

[71] We can however use the fact that individuals are not expected to work at very young and very old ages, which helps with disambiguation in some cases. Specifically, we assume that individuals working in the federal government are between the ages of 18 and 65.

[72] Importantly, there is no correlation between the likelihood of a missing birthplace and the reform.

distance of at least c_2 with $c_2 \geq c_1$. For a given value of c_1, a higher value of c_2 represents a more conservative choice.

In our baseline analysis, we chose $c_1 = 0.07$ and $c_2 = 0.07$. In other words, we deem an observation as a match provided that it is the *unique* observation within a Jaro-Winkler distance of 0.07 with respect to both first and last names. For reference, the Jaro-Winkler distance between "Smith" and "Smiht" is 0.046. However, Figure A7 shows that our results on the likelihood that an employee would have had a professional occupation prior to joining the customs service (our only result that relies on the linked data) are very similar when we implement alternative cutoffs (including just using exact matches).

Figure A5 shows the proportion of individuals that we match to at least one census (and to at least 2, 3 and 4, respectively) by register year when using our baseline choice of parameters. In this figure, we focus on matches to censuses conducted prior to each register year (that is, when we focus on the 1871 Register we ask whether we are able link it to the 1850, 1860 and 1870 censuses) . On average, we are able to find at least one match for about 20% of the customhouse employees. We expect a lower proportion of individuals in later register years to be matched to at least one adult observation, as the last census we include is 1880 and some employees would have been less than 18 years old by this year (particularly those employed in later years). Overall, these matching rates are similar to those in other studies using historical data (Abramitzky *et al.* , 2019).

Representativeness of linked data. In our main analysis using linked data, we assess how the characteristics of customhouse employees changed with the passing of the Pendleton Act. Our sample in this analysis includes *only* employees of the US customs service who were successfully linked to at least one observation in the census. Specifically, we compare the characteristics of bureaucrats in classified customhouses to those in non-classified customhouses, before and after the implementation of the reforms. Hence, for our analysis to be biased by selection it would need to be the case that selection into linkage changed *differentially* for individuals in classified customhouses after the reforms. This is unlikely because our linking procedure is exactly the same throughout all sample years and across customhouses.

To further alleviate this concern, we estimate our main difference-in-differences specification but using as outcome variables: (1) the total number of censuses to which we link an employee, or (2) and indicator that takes a value of one if the employee is linked to at least one observation in the census. Figure A6 shows that, while employees in non-classified customhouses are more likely to be matched throughout the period of analysis, there is no evidence that such difference became larger or smaller after the reform. Indeed, Table A1 in the Online Appendix shows that there is no correlation between the reform and the likelihood of matching an individual to a census. Hence, it is very unlikely that the change in the professional background of employees that we document is due to biases in linking.

Finally, Table A2 shows that our main results on employee turnover and the likelihood that an employee would work in an exempted position are very similar when we estimate them on the smaller linked sample we use for our results on occupational outcomes.

Table A1: No Effects of the Reform on Probability of Matching

	(1) At least 1 match	(2) N. of matches
Classified X After	-0.0111	-0.00417
	(0.0202)	(0.0283)
Customhouse FE	Yes	Yes
Year FE	Yes	Yes
Observations	45523	45523

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. The dependent variable in column (1) is an indicator that takes a value of one if a customhouse employee is successfully matched to an observation in the census. The dependent variables in column (2) is instead the total number of censuses to which a customhouse employee is matched to.

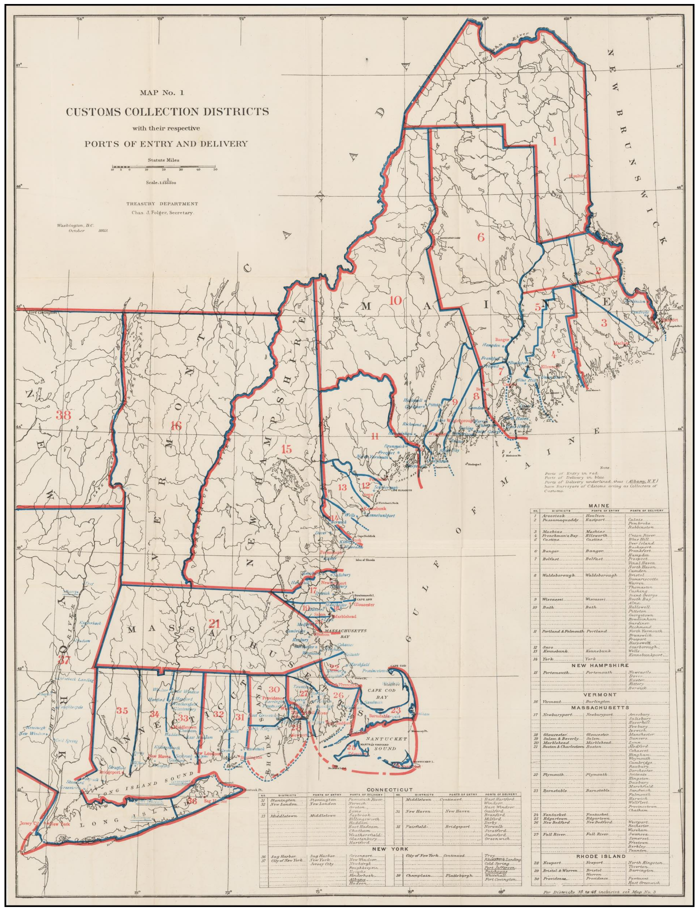
Table A2: All Personnel Outcomes, Linked Sample

	Turnover		Exempted Occ.	
	(1)	(2)	(3)	(4)
Classified X After	-0.126***	-0.115***	0.187***	0.181***
	(0.0269)	(0.0395)	(0.0287)	(0.0355)
Customhouse FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	41435	7658	45101	8279
Sample	All	Linked	All	Linked

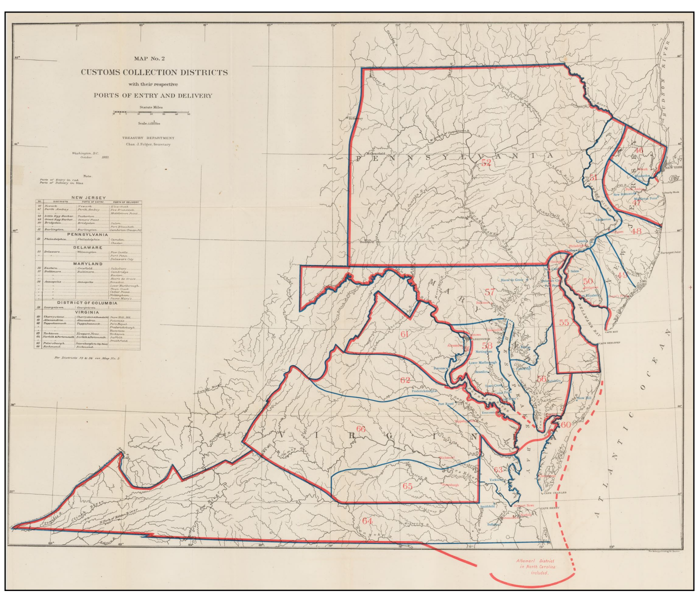
Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table shows the robustness of our personnel results that do not rely on linked register-to-census data to using the linked sample that we use in our results on occupational background. Standard errors are clustered at the customhouse level.

FIGURE A1: CUSTOMS COLLECTION DISTRICTS IN 1883

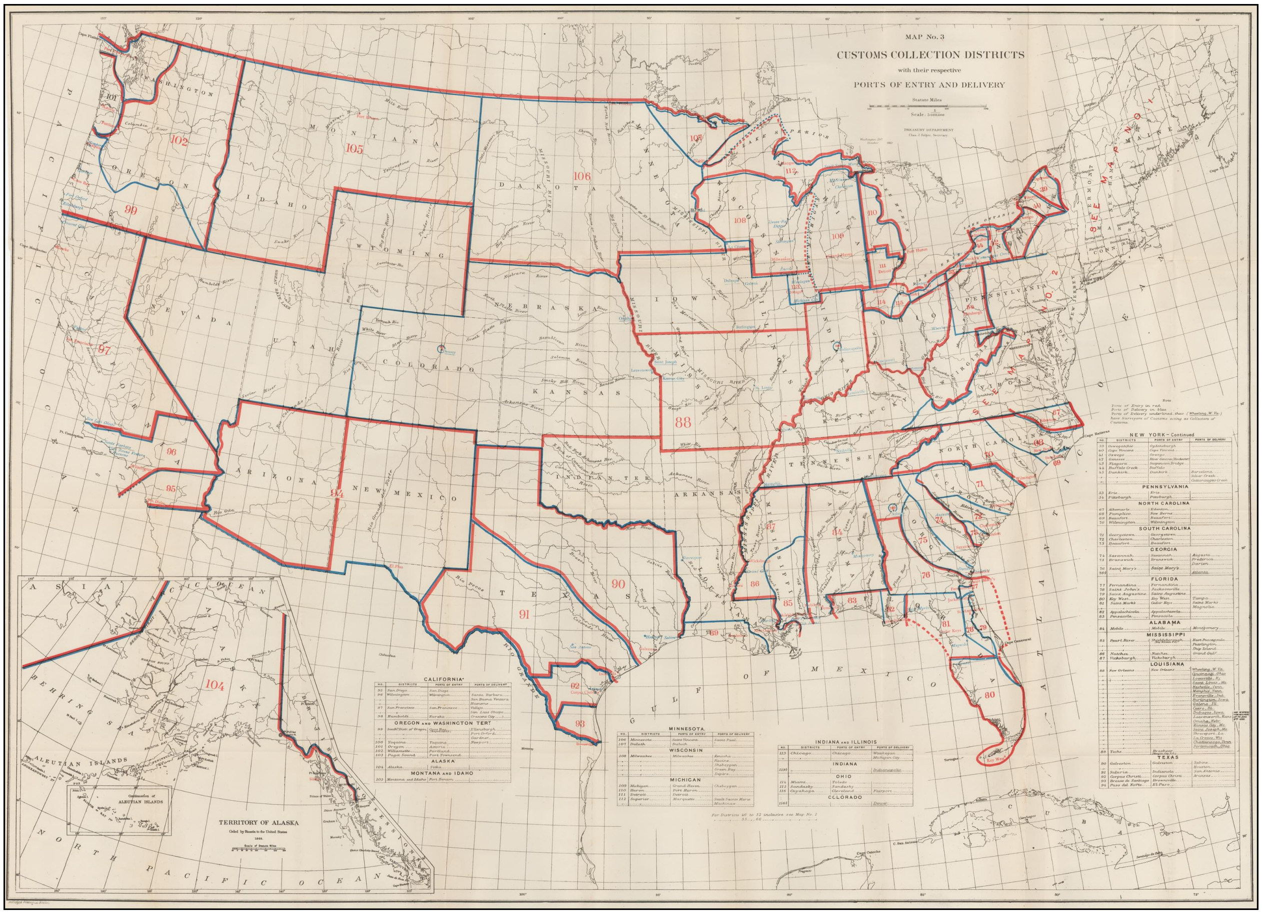
(A) NEW ENGLAND



(B) MIDDLE ATLANTIC

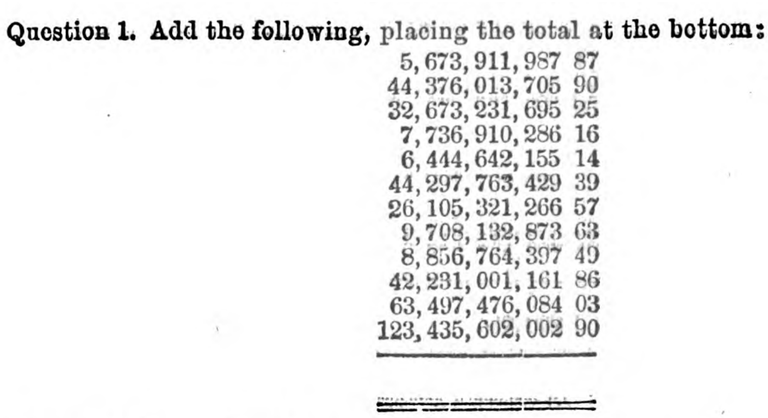


(C) REST OF THE COUNTRY



Notes: This figure shows the customs collection districts in 1883. Source: US Congress (1874-1893).

FIGURE A2: EXAMPLE QUESTION, ARITHMETICS EXAM



Question 1. Add the following, placing the total at the bottom:

5, 673, 911, 987 87
44, 376, 013, 705 90
32, 673, 231, 695 25
7, 736, 910, 286 16
6, 444, 642, 155 14
44, 297, 763, 429 39
26, 105, 321, 266 57
9, 708, 132, 873 63
8, 856, 764, 397 49
42, 231, 001, 161 86
63, 497, 476, 084 03
123, 435, 602, 002 90

Notes: This figure shows an example question for individuals who applied for the position of clerk in the classified customs service.

FIGURE A3: EXAMPLE PAGE, OFFICIAL REGISTERS OF THE UNITED STATES

1, 1883.] TREASURY DEPARTMENT. 201

Customs Service.

Name.	Office.	Where born.	Whence appointed.	Where employed.	Compensation.
Joseph Jewett	Clerk	Massachusetts	New York	New York	$1, 600 00
George H. Keim	do	New York	do	do	1, 600 00
Berrien Keyser	do	do	do	do	1, 600 00
Louis Oppenheim	do	do	do	do	1, 600 00
Samuel P. Putnam	do	New Hampshire.	do	do	1, 600 00
James H. Thayer	do	Massachusetts.	do	do	1, 600 00
Theodore D. Wilson	do	Pennsylvania	do	do	1, 600 00
Edward E. Worl	do	do	do	do	1, 600 00
George W. Marston	do	New Hampshire.	do	do	1, 550 00
Michael Carey	do	Ireland	do	do	1, 400 00
Herman G. Carter	do	New York	do	do	1, 400 00
Calvin C. Church	do	do	do	do	1, 400 00
William B. Crawford	do	do	do	do	1, 400 00
Alfred Eaton	do	do	do	do	1, 400 00
Stephen B. Gregory	do	do	do	do	1, 400 00
Charles B. Jenney*	do	do	do	do	1, 400 00
Oliver W. Marvin	do	do	do	do	1, 400 00
John H. Walsh	do	do	do	do	1, 400 00
John Welch, jr	do	Massachusetts	do	do	1, 400 00
Thomas S. Woodcock	do	New York	do	do	1, 400 00
George P. Babcock	do	Connecticut	do	do	1, 200 00
Theodore Babcock, jr	do	New York	do	do	1, 200 00
John J. Barnicle	do	do	do	do	1, 200 00
Thomas H. Bryden	do	do	do	do	1, 200 00
Ogden D. Budd	do	do	do	do	1, 200 00
Samuel G. Burns*	do	do	do	do	1, 200 00
Frederick S. Cooke*	do	do	do	do	1, 200 00
George W. Cooney*	do	do	do	do	1, 200 00
Anthony Gross	do	Austria	do	do	1, 200 00
Edward H. Jones	do	England	do	do	1, 200 00
George Kleine	do	New York	do	do	1, 200 00
James B. Martine*	do	North Carolina	do	do	1, 200 00
John O'Shea	do	New York	do	do	1, 200 00
Charles E. Parsons	do	do	do	do	1, 200 00
James M. Smith	do	do	do	do	1, 200 00
Lewis A. Strahan*	do	do	do	do	1, 200 00
William P. Thomson	do	do	New Jersey	do	1, 200 00
Benson Van Voast*	do	do	New York	do	1, 200 00
Leonard Wightman	do	do	do	do	1, 200 00
Stephen B. Goszler	Clerk and messenger	Dist. Columbia	New Jersey	do	1, 000 00

Notes: This figure shows an example page of the "Official Registers of the United States" corresponding to employees of the New York customhouse in 1883.

FIGURE A4: EXAMPLE OF A COLLECTOR WHO DIED WHILE IN OFFICE

DEC. 19, 1889.] EXECUTIVE JOURNAL. 245

To the Senate of the United States:

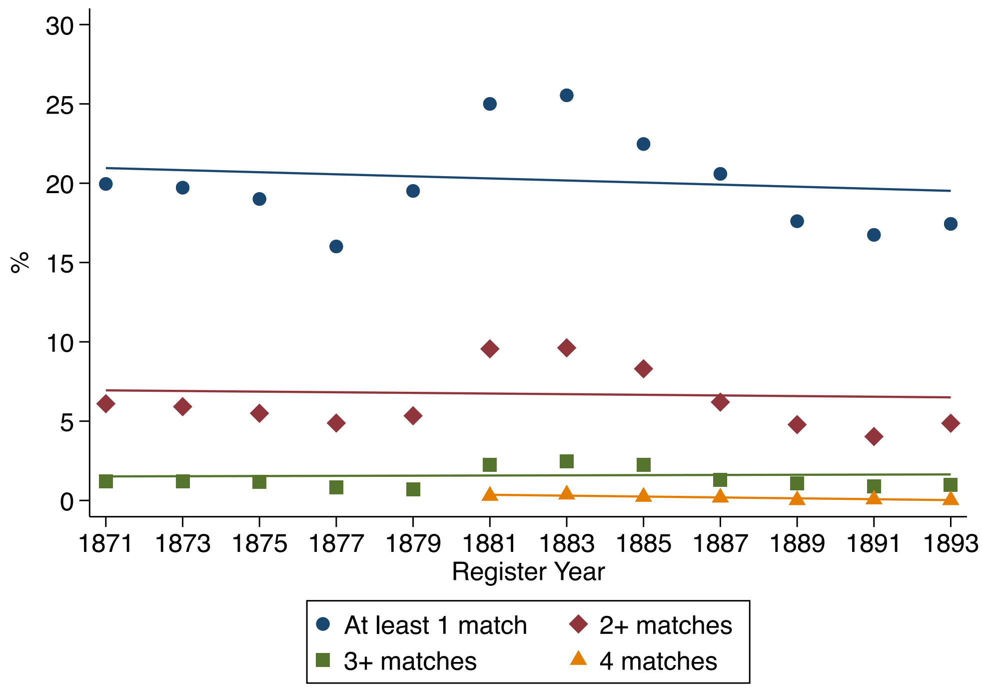
I nominate T. Jefferson Jarrett, of Virginia, to be collector of customs for the district of Petersburgh, in the State of Virginia, to succeed Peter F. Cogbill, deceased.

Mr. Jarrett was temporarily commissioned during the recess of the Senate, June 13, 1889.

BENJ. HARRISON.

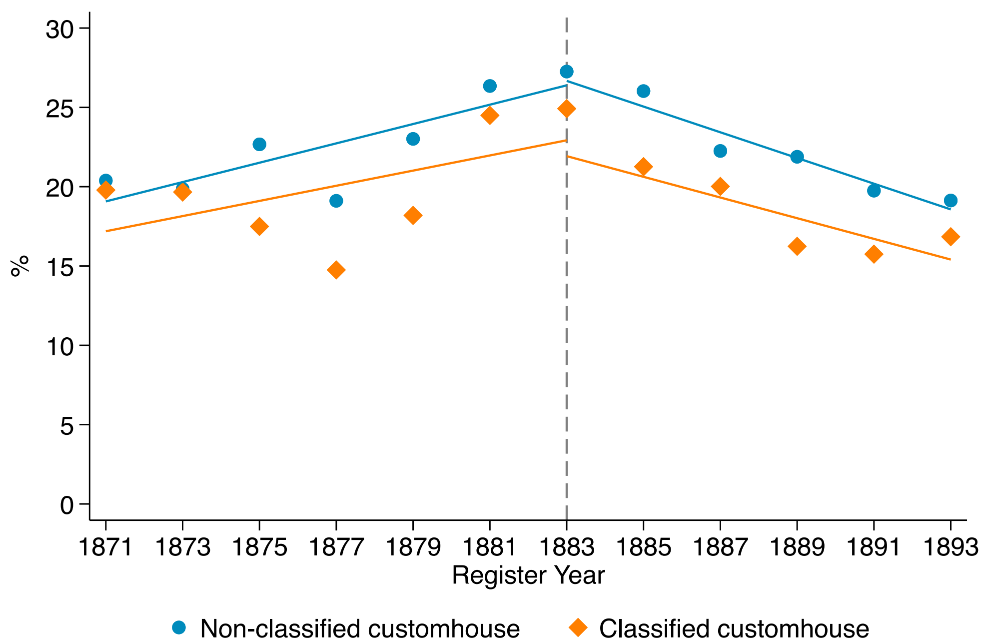
EXECUTIVE MANSION, *December 19, 1889.*

FIGURE A5: MATCH RATE BY REGISTER YEAR



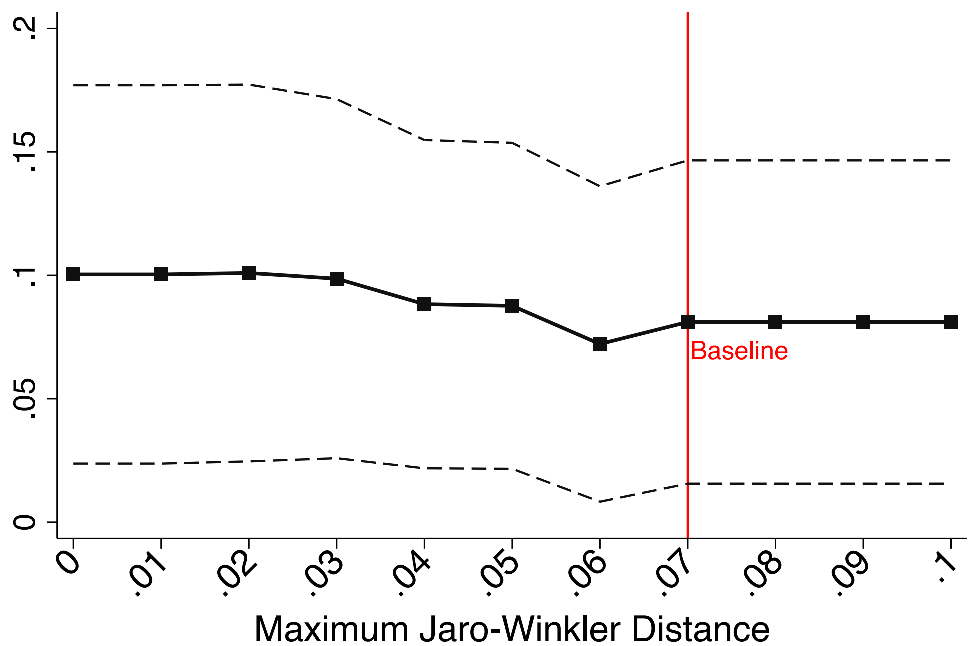
Notes: This figure shows the percent of customhouse employees that are matched to at least 1, 2, 3 or 4 censuses, by register year.

Figure A6: Match Rate by Register Year and Classification Status



Notes: This figure shows the percent of customhouse employees that are matched to at least one observation in the census, by register year and depending on whether the individual worker in a classified or a non-classified customhouse.

FIGURE A7: ROBUSTNESS TO ALTERNATIVE JARO WINKLER CUTOFFS



Notes: This figure shows the estimated effect of the reform on the likelihood that a customhouse employee would have held a professional occupation (y-axis), as a function of the maximum Jaro Winkler distance above which an observation would no longer be considered a match (x-axis). Lower values of the Jaro Winkler distance represent a more conservative match.

B Additional Results

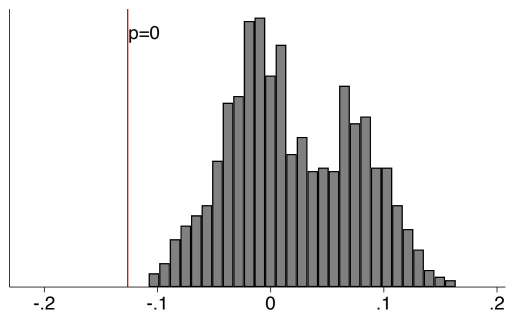
Table B1: Expenses and Revenue, Excluding First Five Post-Reform Years

	log(Expenses)		log(Receipts)		log(Receipts/Expenses)	
	(1)	(2)	(3)	(4)	(5)	(6)
Classified X After	-0.0150	-0.115	0.0231	-0.187	0.0381	-0.0715
	(0.111)	(0.0866)	(0.244)	(0.229)	(0.171)	(0.173)
Customhouse 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	705	705	705	705	705	705

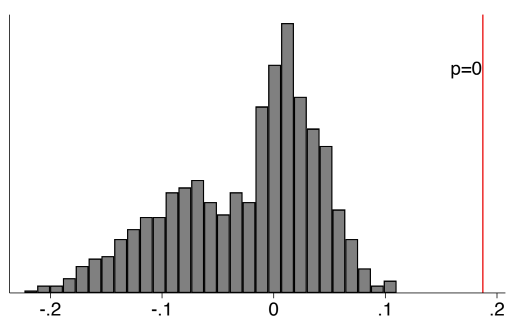
Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. In this table, we exclude the first five post-reform years from the sample. The dependent variable in columns 1 and 2 is the log of total expenses, in columns 3 and 4 it is the log of total receipts, and in column 5 and 6 it is the natural log of the ratio between the total receipts and expenses. $Classified \times After$ takes a value of one for customhouses that were part of the classified customs service after 1883. Odd columns include Year FE and Customhouse FE. Even columns also include $Region \times TimeFE$. The sample is restricted to customhouses with at least 10 employees by 1883. Standard errors are clustered at the customhouse level.

FIGURE B1: PERSONNEL PRACTICES, RANDOMIZATION INFERENCE

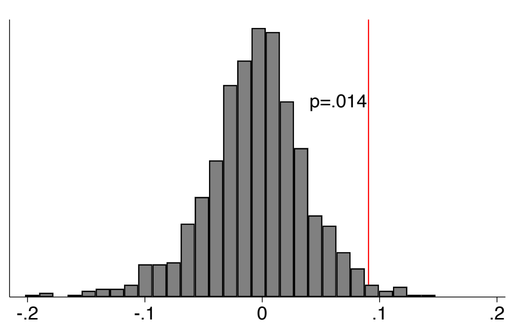
(A) TURNOVER



(B) EXEMPTED OCCUPATION



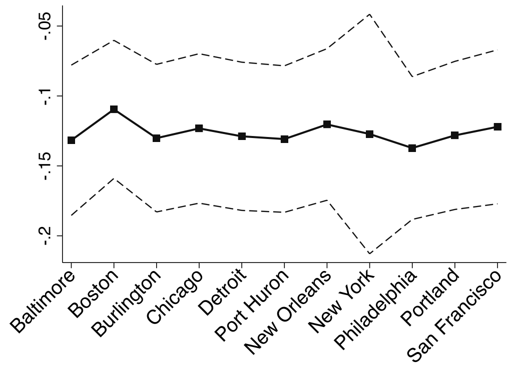
(C) PROFESSIONAL OCCUPATION



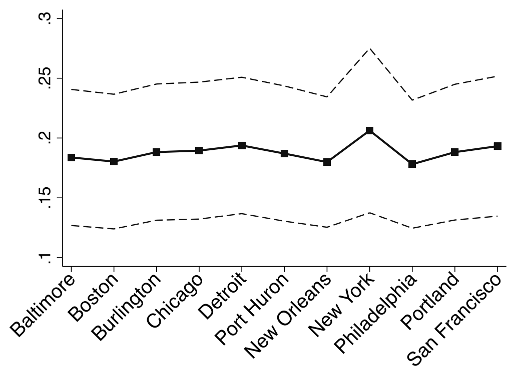
Notes: These figures show the empirical distribution of estimated effects when we implement a randomization inference approach. In this exercise, we randomly select eleven customhouses as bein classified and estimate the "effects" of the reform using our baseline differences-in-differences setup. We repeat this exercise 1,000 times and plot the empirical distribution of estimated effects. The vertical red line corresponds to our estimated effect when we use the actual set of customhouses that were added to the classified civil service. The specification and outcome in Panel (a) correspond to those in column 1 of Table 3. The specification and outcome in Panel (b) correspond to those in column 1 of Table 6. The specification and outcome in Panel (c) correspond to those in Panel (a), column (4) of Table 4.

FIGURE B2: PERSONNEL PRACTICES, EXCLUDING ONE CUSTOMHOUSE AT A TIME

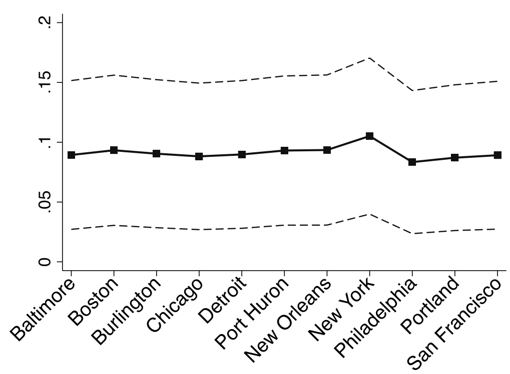
(A) TURNOVER



(B) EXEMPTED OCCUPATION

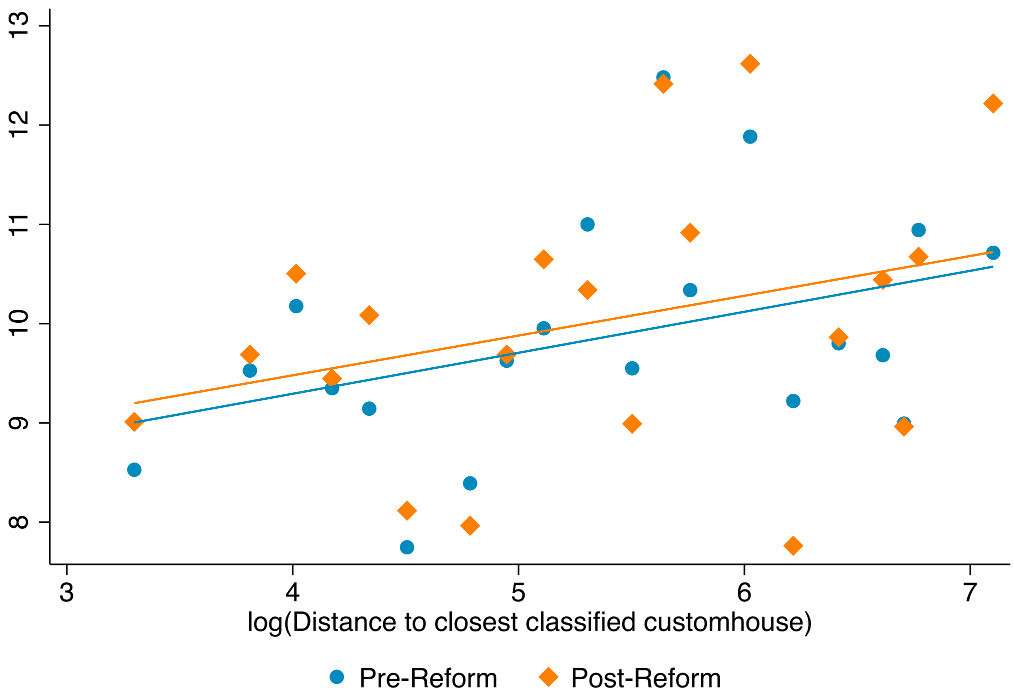


(C) PROFESSIONAL OCCUPATION



Notes: These figures show the sensitivity of the personnel results to excluding one classified customhouse at at time. The y-axis shows our baseline difference-in-difference estimates around a 95% confidence interval when estimated excluding each of the classified customhouses indicated in the x-axis. The specification and outcome in Panel (a) correspond to those in column 1 of Table 3. The specification and outcome in Panel (b) correspond to those in column 1 of Table 6. The specification and outcome in Panel (c) correspond to those in Panel (a), column (4) of Table 4.

FIGURE B3: PROXIMITY TO CLASSIFIED CUSTOMHOUSES AND CUSTOMHOUSES' REVENUE



Notes: This figure shows the correlation between the distance to the nearest classified customhouse and total receipts, before and after 1883. The sample is restricted to the non-classified customhouses.

Table B2: Placebo, Personnel Practices

(a) 20+ employees

	(1) Turnover	(2) Exempted Occ.	(3) Professional Occ.
Placebo Classified X After	0.0279	0.0267	0.00824
	(0.0349)	(0.0528)	(0.0759)
Customhouse FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	8638	9158	482

(b) 30+ employees

	(1) Turnover	(2) Exempted Occ.	(3) Professional Occ.
Placebo Classified X After	-0.00297	0.0594	-0.0687
	(0.0301)	(0.0490)	(0.0596)
Customhouse FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	8638	9158	482

(c) 40+ employees

	(1) Turnover	(2) Exempted Occ.	(3) Professional Occ.
Placebo Classified X After	-0.0134	-0.00227	0.0226
	(0.0233)	(0.0603)	(0.0834)
Customhouse FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	8638	9158	482

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. Standard errors are clustered at the customhouse level.

TABLE B3: ROBUSTNESS TO USING ALTERNATIVE CONTROL GROUPS

	Turnover			Exempted Occ.			Professional Occ.		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Classified X After	-0.105***	-0.126***	-0.138***	0.176***	0.187***	0.180***	0.0868***	0.0905***	0.0914**
	(0.0251)	(0.0269)	(0.0285)	(0.0256)	(0.0287)	(0.0342)	(0.0310)	(0.0315)	(0.0358)
Customhouse FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	44618	41435	38274	48264	45101	41786	1953	1860	1712
Comparison group	0+	10+	20+	0+	10+	20+	0+	10+	20+

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table shows the robustness of our personnel results to using alternative control groups. In columns 1, 4 and 7, the control group is comprised of all non-reformed customhouses (regardless of their number of employees in 1883). Columns 2, 5 and 8 correspond to our baseline sample (using customhouses with 10+ employees in 1883 as the control group). In columns 3, 6 and 9, the control group includes customhouses with 20+ employees by 1883. Standard errors are clustered at the customhouse level.

TABLE B4: CONTROLLING FOR REGION X TIME EFFECTS

	Turnover			Exempted Occ.			Professional Occ.		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Classified X After	-0.120**	-0.121***	-0.117**	0.207***	0.194***	0.218***	0.110***	0.0859***	0.111***
	(0.0479)	(0.0294)	(0.0506)	(0.0406)	(0.0263)	(0.0320)	(0.0346)	(0.0299)	(0.0349)
Customhouse FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1883 Employees X Year	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes
Region X Year	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	41435	41435	41435	45101	45101	45101	1860	1860	1860

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table shows the robustness of our personnel results to controlling for: (1) interactions between the number of employees in 1883 and year dummies, and (2) interactions between census region and year dummies. Standard errors are clustered at the customhouse level.

TABLE B5: NUMBER OF EMPLOYEES

	Total	Subject to Exam	Non Subject to Exam
	(1)	(2)	(3)
Classified X After	0.0852	-0.206	0.659***
	(0.0911)	(0.130)	(0.158)
Customhouse FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	588	572	587

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. An observation corresponds to customhouse-year. The dependent variable in column 1 is the log number of employees in customhouse c in year t. The dependent variable in column 2 is the log number of employees in non-exempted positions and in column 3 is the log number of employees in exempted positions. $Classified \times After$ takes a value of one for customhouses that were made part of the classified system after 1883. Standard errors are clustered at the customhouse level.

TABLE B6: EMPLOYEES PAID BELOW THE EXAM CUTOFF HAD WEAKER PROFESSIONAL BACKGROUNDS

	(1) Professional Occ.	(2) Unskilled	(3) log(Occscore)	(4) Literate
Below Exam Cutoff	-0.0273**	0.0965***	-0.0598	-0.0203
	(0.0108)	(0.0187)	(0.0806)	(0.0126)
Customhouse FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	2752	2752	2752	2752

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. "Below Exam Cutoff" is an indicator that takes a value of one if an employee made less than $900a year. Standard errors are clustered at the customhouse level.

TABLE B7: TURNOVER AND RECEIPTS

	log(Receipts)		log(Fines)		log(Expenses)	
	(1)	(2)	(3)	(4)	(5)	(6)
Share of New Hires	0.144		0.675**		-0.0498	
	(0.162)		(0.281)		(0.0717)	
Died in Office		0.469*		0.596		-0.179
		(0.253)		(0.497)		(0.132)
Customhouse FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	810	810	810	810	810	810

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. An observation corresponds to a customhouse-year."Share of New Hires" is the proportion of a customhouse employees that show up for the first time in year t register. "Died in Office" is an indicator takes a value of one if a collector died while in office. Standard errors are clustered at the customhouse level.

Table B8: Spillovers to non-classified customhouses

	(1)	(2)	(3)	(4)
Distance Closest Classfied X After	-0.0122	-0.0157	0.152	-0.00241
	(0.126)	(0.137)	(0.143)	(0.215)
Customhouse FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1520	1360	720	420
Comparison group	0+	5+	10+	20+

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. An observation corresponds to a customhouse-year. This table shows the correlation between distance to the closest classified customhouse in the post-reform period and a customhouse total receipts.

Table B9: Spillovers, Personnel Practices

	(1) Turnover	(2) Exempted Occ.	(3) Professional Occ.
Distance to Closest Classified X After	-0.0257	0.0243	0.00696
	(0.0177)	(0.0257)	(0.0306)
Customhouse FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	7921	8417	454

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. An observation corresponds to an employee-year.This table shows the correlation between distance to the closest classified customhouse in the post-reform period and personnel outcomes. The sample is restricted to customhouses that were not subject to the reform.

Table B10: Spillovers to non-classified customhouses: excluding non-classified customhouses in close proximity to classified customhouses

	(1)	(2)	(3)	(4)
Classified X After	0.0251	0.00703	-0.00443	0.0283
	(0.185)	(0.194)	(0.206)	(0.208)
Customhouse FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	940	880	820	700
Comparison group	All	50+ miles	100+ miles	200+ miles

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table shows the estimated effects of the reform on log receipts when we restrict the control group to customhouses who are at least 50, 100 or 200 miles away from a classified customhouses.

TABLE B11: LONGER-TERM EFFECTS

	1874-1893		1874-1904	
	(1)	(2)	(3)	(4)
Classified X After	0.0358	-0.0543	0.0113	-0.0689
	(0.137)	(0.138)	(0.164)	(0.169)
Region X Time FE	No	Yes	No	Yes
Comparison group	10+	10+	10+	10+
Observations	940	940	1426	1426

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. The dependent variable is the natural log of the ratio between total receipts and expenses. $Classified \times After$ takes a value of one for customhouses that were part of the classified customs service starting in 1883. Odd columns include year fixed effects and customhouse fixed effects. Even columns also include $Region \times TimeFE$. Columns (1) and (2) report the results for the dependent variable in the first 10 years of the reform (up to 1893), while columns (3) and (4) report the results for the dependent variable in the first 20 years (up to 1904). Standard errors are clustered at the customhouse level.