

Civil Service Exams and Organizational Performance: Evidence from the Pendleton Act*

Diana Moreira Santiago Pérez
UC Davis and BREAD UC Davis and NBER

January 20, 2022

Abstract

Competitive exams are a standard method for selecting civil servants. Yet, there is limited evidence on their effectiveness. We digitize personnel and financial data to study the impacts of the 1883 Pendleton Act, which mandated exams for some employees in the largest US customs-collection districts. Although the act improved targeted employees' professional background and reduced turnover, it did not increase cost-effectiveness in customs revenue collection. Moreover, it incentivized hiring in exam-exempted positions, distorting districts' hierarchical structures. These results illustrate how, by triggering countervailing organizational responses, policies that succeed at improving specific organizational aspects might nevertheless fail to improve overall performance.

* dsmoreira@ucdavis.edu, seperez@ucdavis.edu. We thank Luiza Aires, Lisa Pacheco and Mario Remígio for outstanding research assistance, and Enrique Pérez for help with data collection. We have benefited from the comments of Oriana Bandiera, Brian Beach, Assaf Bernstein, Sandra Black, Nicolás Caramp, Katherine Eriksson, James Feigenbaum, Fred Finan, Robert Gibbons, Michela Giorcelli, Walker Hanlon, Daniel Honig, Rick Hornbeck, Chris Meissner, Andrea Pozas-Loyo, Sarah Quincy, Arman Rezaee, Monica Singhal, Michael Ting, Martin Williams, Chenzi Xu and Noam Yutchnmann, as well by seminar participants at *Corporación Andina de Fomento*, CEPR STEG Workshop, Cornell University, University of California - Riverside, NEUDC, NBER Organizational Economics Fall 2020 Conference, Columbia University, University of Chicago Harris School of Public Policy, NYU, and Stanford CASBS. We also benefited from funding from the UC Davis Small Grant in Aid of Research, the Michael Dearing Fellowship in support of Economic History Research, and the Hellman Fellowship program.

1 Introduction

A well-functioning, professionalized bureaucracy is increasingly seen as an important determinant of economic development (Rauch *et al.*, 1995; Finan *et al.*, 2017; Besley *et al.*, 2021). Historically, the standard recipe for attempting to achieve this professionalization has often involved the introduction of *competitive entry exams* for civil service jobs.¹ Although screening applicants based solely on an exam prevents recruiters from using other (potentially useful) information, a promise of this approach is that, by reducing discretion, it would limit political favoritism and ultimately result in the selection of more qualified employees.

However, bureaucracies are living organisms with endogenously chosen formal and informal organizational practices. Hence, even if exams *did improve* workers' selection, their overall effects on organizational performance could still be ambiguous as exams may trigger additional, potentially countervailing (or reinforcing) organizational changes. For instance, politicians might respond to the reduced hiring discretion by creating additional positions not covered by exams, distorting the organization's hierarchical structure. Alternatively, limiting hiring discretion might reduce the incentives to remove employees, further facilitating the accumulation of bureaucratic expertise. Yet, despite this ambiguity (and their widespread use), there is surprisingly limited evidence on whether civil service exams actually improve the overall functioning of the bureaucracy.² More broadly, we know little about how a personnel practice with a well-defined goal (e.g. improving recruits' qualifications) affects a government organization as a whole (Besley *et al.*, 2021).

This paper studies the impacts of civil service exams on organizational performance. To do so, we use evidence from the 1883 Pendleton Civil Service Reform act, a landmark reform in American history which introduced competitive exams for the selection of some federal employees. Our analysis focuses on the consequences of the act for the functioning of the Customs Service, a key government agency that, by the time of the reform, collected more than half of federal revenue. Although we find that the reform improved targeted employees' professional background and reduced their turnover, we show that these changes did not translate into higher cost-effectiveness in customs revenue collection. Moreover, the reform incentivized hiring in exam-exempted positions, provoking distortions in districts' hierarchical structure.

Our setting has three attractive features for a study of the organizational effects of civil service exams. The first feature is that collection districts did not choose whether or not to adopt exam-based hiring. Rather, districts that had 50 or more employees by 1883 were *mandated* to use exams. This is important, as it enables us to deal with the potential endogeneity of reform adoption –for instance, reforms might be more likely to be adopted when bureaucratic performance is already deteriorating. We use this feature of the reform to estimate difference-in-differences models, comparing districts that were mandated to use exams to those that were not, before and after 1883. To validate this design, we show that the outcomes of these two groups of districts evolved similarly

¹Nearly 80% of countries use formal examinations to select some of their public employees (Teorell *et al.*, 2011).

²Besley *et al.* (2021) write that, “even though they are used extensively, there is little evidence on the effects of competitive exams on the selection of civil servants.”

prior to the reform. Moreover, there are no differential changes in districts' outcomes when we estimate the "effects" of placebo reforms that use alternative employee cutoffs.

The second feature helping our analysis is that there is rich information on the personnel and finances of US customs-collection districts. Personnel data include employees' names, position, and salary, whereas financial data include districts' yearly expenses and receipts. We digitized the universe of these records spanning 1871 to 1893. Moreover, we also collected information on employees' professional backgrounds by linking the personnel data to population censuses, using name-based matching techniques ([Abramitzky et al., 2019](#)). These data enable us to observe what happened *inside* these organizations once exams were introduced (including changes in workers' characteristics and hierarchical structure), as well as to observe changes in the cost-effectiveness with which districts performed their main function, the collection of customs revenue.

The third attractive feature is that, much like in many developing countries today ([Colonnelli et al., 2020](#); [Brassiolo et al., 2020](#)), the US federal administration was characterized by widespread political patronage prior to the reform ([Hoogenboom, 1968](#)). Hence, our analysis provides a window into the early stages of a transition towards a professionalized bureaucracy, allowing us to shed light on the (often windy) path to successful organizational change.

We start by showing that the reform led to improvements in targeted employees' professional background: new hires in positions requiring exams were 11 percentage points less likely to report working in an unskilled occupation prior to joining the Customs Service, and 9 percentage points more likely to report working in a professional one. Since exams were aimed at testing practical knowledge relevant to positions in the Customs Service (rather than formal academic training), we interpret these changes in workers' occupational backgrounds as reflecting a likely improvement in targeted employees' actual qualifications for their jobs.³ Indeed, a shortage of workers with a pre-bureaucracy professional occupation might have been a binding constraint in achieving cost-effectiveness: Prior to the reform, there was a strong positive correlation between changes in the share of such employees and changes in districts' revenue.

Next, we document that, despite not granting tenure to workers, the reform nevertheless led to a sizable reduction in employee turnover: the two-year turnover rate was 27% lower among employees in reformed districts.⁴ This overall decline was driven by reduced turnover in years in which the Federal administration changed party hands, and was stronger among employees in positions subject to exams. These findings suggest that, by limiting politicians' discretion to hire, exams also reduced an important incentive to remove employees in the first place –thus lowering turnover and potentially facilitating the accumulation of bureaucratic expertise.

Ten years after the reform, nearly 60% of employees in reformed districts had been appointed through an exam. We next ask if this change led to increased cost-effectiveness in the collection of

³The focus on practical skills was in contrast to other civil service reforms. For instance, [Grindle \(2012\)](#) argues that in the UK exams were designed such that their contents would only be accessible to those with "access to elite educations at Oxford and Cambridge". In contrast, in the US the Civil Service Commission "maintained that a common school education was sufficient to pass examination" ([Hoogenboom, 1959](#)).

⁴We focus on the two-year turnover rate because our personnel records were collected every two years, see Section 3 for further details on the data.

customs revenue. We expect improvements in this regard through three main channels. First, limiting the room for patronage could have reduced the number of employees hired solely to reward political loyalty, thus lowering personnel expenses. Second, by creating a separation between bureaucrats and politicians, the reform could have reduced corruption thereby increasing revenue.⁵ Third, to the extent that the reform increased bureaucratic expertise, workers might have also become better equipped to enforce the customs laws, further increasing revenue. Indeed, the conventional wisdom among both practitioners at the time and modern scholars is that the reform reduced corruption and improved the overall efficiency of the federal bureaucracy (Hoogenboom, 1959; Commission, 1897; Ujhelyi, 2014).⁶

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

In the last part of the paper, we investigate why the improvement in workers' qualifications and reduction in turnover did not lead to higher cost-effectiveness in revenue collection. We first discuss the potential role played by the incomplete reach of the reform. As typical in other reforms introducing exams, the Pendleton Act exempted some positions from such requirement. Specifically, it exempted positions below a salary threshold as well as districts' top managers (the "collectors of customs").⁷ This incomplete reach was likely important for two reasons. First, by exempting employees below a salary threshold, the reform created incentives to inflate the number of workers in low-paid positions. Indeed, we document that the reform caused a near *doubling* in the share of workers in such positions. This shift was likely pernicious for reformed districts' performance, both because it distorted their hierarchical structure but also because, not surprisingly, low-paid employees tended to have weaker professional backgrounds. Second, to the extent that collectors mattered for districts' outcomes (which we show to be the case by estimating "collector fixed effects" and exploiting collectors' deaths while in office as in Jones & Olken, 2005 and Besley *et al.*, 2011), not changing their selection method might have also limited the reform's ability to improve cost-effectiveness.

Finally, we discuss three additional potential explanations for the lack of improvements on

⁵Indeed, meritocratic appointments are empirically associated with lower corruption (Rauch & Evans, 2000; Dahlström *et al.*, 2012; Meyer-Sahling *et al.*, 2018).

⁶For instance, Hoogenboom (1959) argues that "service in post offices and customhouses was vastly improved". We provide additional quotes and background information in Section 2.3.

⁷Dahlström *et al.* (2012) refers to the coexistence of patronage and merit hires after civil service reforms as "formal political discretion". Subsequent state and local reforms in the US also led to the coexistence of merit and non-merit employees (Ujhelyi, 2014; Ornaghi, 2016).

costs effectiveness (for which we find more limited evidence). First, we find no evidence that the lack of detectable effects on cost-effectiveness was due to the reform spilling over to the non-reformed districts: proximity to a reformed district does not predict post-reform changes in cost effectiveness. Second, the effect of the reform on cost-effectiveness was limited even over a 20-year horizon. This result is contrary to the hypothesis that, to fully capitalize on the benefits of exams, districts needed to introduce changes that took longer than ten years to implement (for instance, fully replacing the employees who had been hired through the old regime). Third, we consider the possibility that, although employees hired through exams might have been of better "quality", they might have also exerted less effort than patronage hires (for instance, because of a potential misalignment between their preferences and politicians' goals). Although the empirical evidence is inconclusive (as we cannot reject patterns consistent with this explanation), we note that, unlike some modern civil service protections, the Pendleton Act *did not* provide tenure to employees. Hence, the disincentive effects of the reform were likely less prominent than in other contexts.

Our data do not enable us to establish if the introduction of exams led to improvements in performance along margins other than revenue per dollar spent.⁸ For instance, reformed districts might have become faster at clearing imports or may have improved on how closely they followed the tariff laws (which would not necessarily lead to higher revenue).⁹ Although "revenue per dollar spent" does not incorporate all dimensions of performance, it does capture an important aspect of it in the context of an agency whose primary goal was revenue collection. Indeed, this measure was regularly mentioned in government publications and by proponents of civil service reform, who blamed patronage for the US high "cost to collect".¹⁰ Moreover, other scholars studying government agencies in charge of revenue collection have used similar performance measures ([Khan et al., 2016](#); [Xu, 2018](#); [Naritomi, 2019](#)).

Related Literature. This paper makes two main contributions. First, we contribute to the literature on the recruitment and hiring of civil servants ([Dal Bó et al., 2013](#); [Finan et al., 2017](#); [Pepinsky et al., 2017](#); [Best et al., 2017](#); [Deserranno, 2019](#); [Ashraf et al., 2020](#); [Weaver, 2020](#)). Several studies in this literature show some of the potential costs of hiring discretion in the public sector ([Xu, 2018](#); [Brassiolo et al., 2020](#); [Colonnelly et al., 2020](#); [Akhtari et al., 2019](#)).¹¹ Our paper analyzes the impacts of a commonly used (but understudied) tool for limiting such discretion: competitive civil

⁸In this analysis, we take the goal of collecting customs revenue as given. However, increasing such revenue might have been negative for the US economy to the extent that it reduced international trade and increased domestic prices.

⁹Although errors leading to higher import duties would have been more likely to be challenged by importers.

¹⁰The Annual Reports of the Secretary of the Treasury ([US Congress, 1874-1893](#)) include such measure in several of their editions. Claims that the "cost to collect" was unusually high in the US were at the core of reformists' arguments.

¹¹Also focusing on a historical setting, [Xu \(2018\)](#) studies the link between patronage and the performance of British Empire's colonial governors. [Xu \(2018\)](#) finds that being connected with the Secretary of State worsens governor's performance but that this effect disappears after a reform that limited patronage. We depart from [Xu \(2018\)](#) in three main ways. First, we study how a reform affected bureaucrats' *selection*, whereas [Xu \(2018\)](#) focuses on patronage's *incentive* effects. Second, our analysis focuses on the reform's consequences for the overall personnel structure and organization of reformed units, whereas [Xu \(2018\)](#)'s analysis of personnel outcomes focuses on top bureaucrats (that is, colonial governors). Third, the reform in [Xu \(2018\)](#) did not establish exam-based recruitment, but rather that appointments would need to be overseen by an independent board.

service exams.¹² Specifically, we show that such exams can improve employees' qualifications, but that these improvements might not necessarily translate into gains in overall performance –as introducing exams might trigger additional, countervailing organizational responses. These results highlight the value of studying the bureaucracy as a *system*, moving beyond the usual focus on individual-level bureaucratic performance (Besley *et al.*, 2021).¹³

Second, we contribute to the literature on civil service reforms. In the US, state and local reforms reduced incumbents' reelection chances (Folke *et al.*, 2011), reduced political budget cycles (Bostashvili & Ujhelyi, 2019), and improved bureaucratic performance (Ornaghi, 2016; Rauch *et al.*, 1995). Remarkably, however, there is limited evidence on the effects of these reforms on the main objects that they are intended to change: namely, the qualifications and employment stability of bureaucrats –and the existing evidence casts doubts on whether these reforms actually generate these intended changes (Ornaghi, 2016).¹⁴ By contrast, our data allow us to study how these reforms affect *both* the personnel outcomes and the overall performance of reformed units. Moreover, the state and local reforms studied in the literature were heterogeneous in their provisions, with most introducing additional changes beyond competitive exams.¹⁵ Instead, our analysis enables us to isolate the effects of exams, a key ingredient in the "production function" of modern bureaucracies. Finally, we study an important historical and policy context: the Pendleton Act is a landmark reform in US history, and the ability to collect revenue is a key determinant of state capacity.¹⁶

¹² As described above, there is limited evidence on the effects of civil service exams. Notable exceptions are Bertrand *et al.* (2020), Dahis *et al.* (2020) and Estrada (2019). Bertrand *et al.* (2020) and Dahis *et al.* (2020) both document a positive association between civil service exam scores and subsequent job performance, but do not compare employees selected through exams to those selected under alternative criteria. Estrada (2019) studies the consequences of recruiting teachers in Mexico based on a competitive exam rather than at the discretion of teachers' unions. Consistent with our findings that the qualifications of workers improves following the introduction of exams, this study shows improvements in teachers' individual-level performance (as measured by their students' test scores). However, our unit-level outcomes suggests that organizations that had exam-based recruits might have experienced subsequent additional organizational changes besides those in the qualifications of their teachers. A complementary set of papers has studied reforms that instead change the group of individuals making the *still discretionary* hiring decisions (see for instance Muñoz & Prem, 2020).

¹³ Our results are also related to the literature on tax enforcement and administration, see for instance Jensen (2019), Khan *et al.* (2019), Khan *et al.* (2016), Gordon & Li (2009) and Slemrod & Yitzhaki (2002).

¹⁴ Ornaghi (2016) investigates the effects of merit systems on the performance of local police departments. She documents that cities with departments under a merit system had less crime than those under a spoils system, but finds no effects on officers' qualifications or turnover. In the absence of police personnel records, Ornaghi (2016) uses the census to identify officers based on their city of residence and occupation. The greater granularity in our data might rationalize why we find an effect on personnel outcomes while Ornaghi (2016) does not. First, similar to Ornaghi (2016), we find a limited effect on employees' qualifications when focusing on the employee *stock*. However, we document an improvement in the qualifications of employees hired *after* the reform in positions subject to exam (who would not have been easily identifiable in the census). Second, we show that the reduction in turnover was driven by reduced turnover in years in which the federal government changed party hands, with no effects in years without a political transition. By contrast, Ornaghi (2016) does not investigate if the reforms reduced this politically motivated turnover since she does not have data on the outcomes of local elections. Finally, identifying police officers using the census might introduce measurement error when assigning officers to a specific police department. For instance, an officer might reside in a city with a merit mandate but work in a nearby city without such mandate.

¹⁵ For instance, some reforms introduced the notion that workers could only be dismissed if there was a "just cause". See Ornaghi (2016) for a description of the provisions included in different state-level reforms.

¹⁶ Our study provides the first quantitative evaluation of the impacts of the Pendleton Act which, as discussed in Johnson & Libecap (1994a), had not yet been possible due to lack of adequate data. The Pendleton Act has attracted the attention of scholars in economics, history, political science, and public administration (see for instance Hoogenboom

2 Historical Background

2.1 The US Customs Service and Customs Revenue Collection

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

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

Import duties depended on the physical quantities of imported products (for goods subject to specific duties), their total monetary value (for goods subject to ad-valorem duties), and their product category. For most goods subject to specific duties (for instance, a bushel of wheat), accurately computing the amount owed by the importer simply required weighing the shipment and applying the corresponding rate.¹⁷ In some cases, however, establishing the correct product category required a more detailed examination of the good’s physical properties. For instance, determining the tariff on raw sugar required a chemical test so as to determine its saccharine content ([Schmeckebier, 1924](#)).

Goods subject to ad-valorem tariffs required an “appraisal” –based on the price of similar goods commercialized in the exporting country– of their monetary value. Hence, appraisers –the customs officers in charge of this procedure– needed to be aware of current market prices so as to detect cases in which importers attempted to undervalue the contents of a shipment. According to [Parrillo \(2013\)](#), however, employees hired through patronage “were typically ignorant, sometimes unable to read the foreign languages in which invoices were written”, leaving them poorly equipped to assess the accuracy of importers’ invoices.

Finally, after being appraised, goods subject to ad-valorem tariffs also had to be “classified” in order to determine the rate of duty to be levied on them. In cases in which a good was not explicitly listed by the law, the Tariff acts established that its rate of duty had to correspond to that of the “closest” listed product. This, in turn, implied that importers had incentives to have their goods classified in the category entailing the lowest duties. [Schmeckebier \(1924\)](#) provides several examples of the ambiguities in classifying some goods, and of how importers could exploit such

([1968](#)); [Theriault \(2003\)](#); [Johnson & Libecap \(1994a,b\)](#); [Libecap & Johnson \(2007\)](#); [Van Riper \(1976\)](#)).

¹⁷Yet, [Schmeckebier \(1924\)](#) describes a scandal involving a sugar importer who had tampered with the scales so that they would show less weight.

ambiguities to their favor.¹⁸

2.2 US Customs Service Before the Pendleton Act

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

Proponents of civil service reform targeted the Customs Service as a prime example of the dangers of patronage appointments.¹⁹ In 1877, Congress appointed a number of commissions to investigate the functioning of the major collection ports in the country. As part of its investigations, the *Jay Commission* (in charge of investigating the port of New York) compared the expenses and functioning of the Customs Service in the US to those in other countries (Treasury Department, 1877). The Commission found that the “cost to collect one dollar” was much higher in the US than in Germany, England or France, where customs’ employees were appointed through civil service examinations. The Commission attributed this higher cost to overstaffing (which inflated operating expenses), as well as to the “errors and fraud” of patronage employees (which led to yearly revenue losses of up to 40%).²⁰

The *Jay Commission* report includes several examples that suggest high levels of inefficiency and corruption in the New York collection district prior to the Pendleton act (Treasury Department, 1877). The district was overstaffed so that “many of the weighers and foremen rendered little, if any, service to the government”, and “some of the clerks performed no duties at all”. Moreover, widespread corruption reduced customs revenue as “the law against the acceptance of bribes was a dead letter.” Employees’ carelessness further depressed revenue, as “some of the employees didn’t have brains enough to do the work, some were incapacitated by ignorance and some by carelessness and indifference.”²¹ The reports corresponding to the Philadelphia and New Orleans districts (and an earlier report about New York) contain similar examples.²²

¹⁸For instance, he describes the case of the importation of “pickle forks” which could plausibly have been assigned to three different classes (each with a different duty payable).

¹⁹Reformers blamed patronage for revenue losses: “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).

²⁰According to Rogers (1921), “the investigating commission appointed by President Grant in 1871 reported that the loss was probably twenty-five percent. The New York Chamber of Commerce assured Sherman’s commission that it had risen to forty percent in 1877.” Similarly, the 1866 “Report of the Special Commissioner of the Revenue” claimed that \$12 to \$25 million in revenue were annually lost just in the New York district due to “frauds, waste and incompetency”. These losses were mostly driven by the “undervaluation of invoices” (Wells, 1866). For reference, the entire Customs Service collected around 130 US\$ million in 1866.

²¹For instance, they described how “In fact, much of the weighable merchandise was not weighed at all. The Custom-House employees either took their figures from the city weighers or copied off the foreign marks of weight found on the packages.” (Rogers, 1921)

²²The Philadelphia commission asserted that its workforce could “be somewhat reduced without injury to the service.” It also listed examples of corruption leading to lower cost-effectiveness “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

2.3 Civil Service Reform Movement and the Pendleton Act of 1883

While pressure for the adoption of a merit reform had been mounting since the 1860s ([Hoogenboom, 1968](#)), the exact timing of the passing of the Pendleton act is related to two political events. First, in July of 1881, President James A. Garfield was shot by a disappointed office-seeker (Garfield would die by September). This assassination put civil service reform at the center of the political stage, and provided reformists with a powerful example of the evils of the “spoils system.” Only three months after Garfield’s death, Democratic senator George H. Pendleton reintroduced a civil service reform bill. Second, Democrats took control of the House in 1882. Fearing that they would lose the 1884 presidential election, Republicans supported the bill hoping that it would help protect Republican officeholders from politically motivated dismissals. In January of 1883, President Chester A. Arthur signed Pendleton’s bill into law.²³

Appointing Procedure and Exam Characteristics. The act’s main provision was to establish that employees in certain “classified” positions would need to be selected through “open, competitive examinations” ([United States Civil Service Commission, 1883](#)). Specifically, appointing officers willing to fill a classified position were only allowed to select candidates among those with the *top four* scores in an open exam.²⁴

Exams were aimed at testing practical knowledge relevant to an applicant’s future position rather than formal academic training. Applicants to the classified Customs Service could complete one of nine different exams.²⁵ While all the exams tested applicants’ basic literacy and numeracy skills (the exams included the subjects of “orthography”, “copying”, “penmanship” and “arithmetic”), the exact list of subjects depended on the position a candidate was applying to. For instance, whereas applicants to the position of clerk had to complete a bookkeeping examination, applicants to the position of “gauger” had to complete a practical task involving measurement. Panel (a) in Figure [A2](#) shows an example question of the arithmetic exam, whereas Panel (b) shows an example question of the bookkeeping exam.

The act also included provisions intended to ensure the political neutrality of the appointment process. First, exam administration was overseen by a bipartisan “Civil Service Commission”. Second, the exams were graded anonymously by an “examining board”. In the Customs Service, these boards were appointed locally within each classified district. Third, exams could not include questions aimed at eliciting an applicant’s political orientation.

Although it changed the method of appointment for some federal employees, it is important to emphasize that the act *did not* grant tenure to these employees: classified workers remained open to

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”. Finally, [Wells \(1866\)](#) argued in a report published in 1866 that a majority of customs employees in New York “have no special qualifications for their places, and little knowledge of the law under which they discharge their duties.”

²³The law established that the requirement to appoint certain employees through competitive exams would start being enforced six months after the passage of the act.

²⁴This number was further reduced to three in March of 1888 ([Commission, 1886](#), p.128).

²⁵These were the exams for “examiners”, “clerks”, “gaugers”, “weighers”, “samplers”, “inspectors”, “inspectresses”, “night inspectors” and “openers and packers”.

the possibility of removal as administrations changed (Johnson & Libecap, 1994a).²⁶ Later reforms (in particular, the 1912 Lloyd-La Follette Act) further increased the stability of federal government employment by introducing the notion that employees could only be removed for “just causes” (Johnson & Libecap, 1994a).²⁷

Which Collection Districts were Affected? The “classified” (that is, subject to exams) Customs Service was initially restricted to collection districts with *at least 50* employees, and to positions earning \$900 or more within these districts: By 1883, 11 districts met this threshold.²⁸ Despite some of these districts included both a “port of entry” and “ports of delivery”, the reform only applied to employees in the main port (that is, the “port of entry”) within each district.²⁹ Figure 1 shows the location of districts’ ports of entry, distinguishing between those that correspond to a district that would become part of the classified service by 1883 and those that would not.³⁰

From 1883 to 1893, nearly 22,000 applicants completed an exam to join the Customs Service, out of which about 2,800 had been appointed by 1893 (United States Civil Service Commission, 1893, p.240). By 1893, nearly 60% of the workforce in classified collection districts had been appointed through examinations.³¹

Historical Evidence on the Effects of the Reform. The conventional wisdom is that, although initially covering only a limited number of federal jobs, the reform improved the efficiency of the federal bureaucracy in general and of the Customs Service in particular.³² For instance, Hoogen-

²⁶“The power to remove for even the most partisan and selfish reasons remains unchanged” (United States Civil Service Commission, 1883). The only exception is that employees could no longer be removed for refusing to perform a political service or paying a political assessment, although this provision of the law applied to all positions, not just the classified ones.

²⁷The reform did not regulate how promotion to a higher-paying occupation within the classified Customs Service would be determined. However, it established that employees could not be promoted from a position outside the classified service into a position subject to exams (see “Customs Rule V” in United States Civil Service Commission (1893)). Similarly, the reform did not allow for transferring employees across districts unless the transfer occurred from one classified district to another.

²⁸These were New York, NY; Boston, MA; Philadelphia, PA; San Francisco, CA; Baltimore, MD; New Orleans, LA; Chicago, IL; Burlington, VT; Portland, ME; Detroit, MI and Port Huron, MI. Districts whose number of employees fell below 50 after the reform was implemented had to nevertheless remain within the classified service. The law further divided the classified civil service into the “Classified Departmental Service” for federal employees in DC, and the “Classified Postal Service” for postal workers.

²⁹As described in Commission (1886), the classified Customs Service included the offices of the “Collectors of customs, naval officers, surveyors, and appraisers at the *ports* of Boston, New York, Philadelphia, Baltimore, New Orleans, and San Francisco; collector of customs and appraiser at Chicago, and collector of customs at Portland, Me., Burlington, Detroit, and Port Huron.” Ports of delivery in which the surveyor acted as a collector –which were allowed to collect revenue– functioned in practical terms as a separate district. In our baseline specification, these “ports of delivery” are potentially included in the control group. In practice, however, all of them except Cincinnati and Saint Louis (that belong to the New Orleans’ collection district) are dropped once we restrict the sample to ports with at least 10 employees. All of our results are similar if we: (1) exclude these ports of delivery from the control group (Table B8), or (2) treat them as part of the classified Customs Service (Table B7). This is not surprising because ports of delivery accounted for only 3% of total employment and 2% of total annual customs receipts.

³⁰Figure A1 shows a detailed map of these districts, together with a list of their associated ports of entry.

³¹Initially the act covered around 60% of all positions in the Customs Service: 75% of Customs Service employees worked in a “classified” district, and within these districts 80% of employees worked in a non-exempted position. As we explain in footnote ³², the reform was initially much less comprehensive for the federal government as a whole.

³²The limited initial coverage of the Pendleton Act is highlighted in the historical literature (see, for instance, Hoogenboom (1959)). Initially, the act covered approximately 10% of all federal positions (although a higher share within the Customs Service as described in footnote ³¹). A large share of the exempted positions corresponded to employees in the

boom (1959) argues that “service in post offices and customhouses was vastly improved”. Similarly, Johnson & Libecap (1994b) describe how the reform “improved the performance of federal workers in the positions that were covered by the law” (although they acknowledge that these claims are based on limited quantitative data). The annual reports of the Civil Service Commission (arguably, an interested party) contain several accounts of customs collectors describing improvements in the functioning of their agencies (see pages 32-39 in [Commission \(1884\)](#) and page 38 in [Commission \(1885\)](#)). Indeed, in its 15th report, the Commission claimed that the reform had caused a 25% decline in the cost of collecting customs revenue, although the basis of this calculation is unclear ([Commission, 1897](#)).³³

3 Data

Our analysis combines newly digitized customs-collection districts’ personnel and financial data. In this section, we describe how we collected the data and provide summary statistics.

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

We gathered additional information on the professional background of these employees by linking the Registers to US population censuses. Specifically, we used name-based matching techniques ([Abamitzky et al., 2019](#)) to link each of the 1871-1893 Registers to the 1850, 1860, 1870 and 1880 population censuses.³⁴ Through this procedure, we obtained information on employees’ occupations *prior* to their employment in the federal government.³⁵ While we provide further details on the linking strategy and sensitivity checks in Online Appendix Section A.1, we note here that: (1) the reform does not affect the likelihood of finding an individual in the census ([Table A1](#)), and (2) the results that *do not* require the linked data are very similar when estimated in this linked sample ([Table A2](#)).

Postal Service. The act also authorized the President to extend the coverage of the classified service through executive order.

³³The quote is as follows: “Officials in charge of collecting the customs duties of the Government have emphatically stated that there has been a saving of about one-fourth in the cost of gathering this part of the public revenue.” ([Commission, 1897](#)). First, as we show below, there is no evidence of a reduction in expenses in our difference-in-differences analysis. Moreover, a simple comparison of total expenses in 1897 and 1883 yields an actual nominal *increase* of 8%, which corresponds to a higher increase in real terms since the US price level actually declined in this period. Indeed, Fish (1905)’s analysis of the Pendleton Act states (regarding the savings claimed by the Civil Service Commission) that “such definite statements, however, lack a firm basis.”

³⁴The census of 1850 is the first to list persons individually, and there are no surviving individual-level records for 1890.

³⁵A challenge when linking US historical data is that women have historically changed their last name upon marriage, making it hard to track them across different sources. This issue is less problematic in our context since very few married women worked in 19th century US. Moreover, unlike the “Departmental Service” in DC ([Ziparo, 2017](#)), the Customs Service employed extremely few women ([United States Civil Service Commission, 1883](#)).

Financial Records. We collected yearly data on the annual receipts and expenses of each collection district from 1874 to 1893.³⁶ These data come from the *Annual Report of the Secretary of the Treasury on the State of the Finances* ([US Congress, 1874-1893](#)). This report was published annually and includes detailed information on the revenues and expenditures of the different branches of the federal government, including the Customs Service.

Our expenses measure corresponds to the amount reported as “expenses for collecting the revenue from customs” ([US Congress, 1874-1893](#)). Our baseline measure of total receipts corresponds to the sum of receipts from “customs”, “fines, penalties and forfeitures”, “emolument fees”, “services of United States officers”, “labor, drayage and storage” and “weighing fees”, although we find very similar results if we just focus on total receipts from customs.³⁷ Finally, our baseline measure of cost-effectiveness is the ratio between total receipts and total expenses.

Summary Statistics. Table B1 shows some basic information on the personnel structure and finances of customs collection districts. Panel (a) focuses on statistics computed at the collection-district 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 districts (in particular New York) accounting for most revenue and expenses.³⁸ Similarly, the distribution of the number of employees is skewed; the median non-classified district had about 10 employees, whereas the median classified district had close to 200.

4 Empirical Strategy

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

$$y_{ct} = \alpha_c + \alpha_t + \beta \text{Classified}_c \times \text{After}_t + \gamma X_{ct} + \epsilon_{ct} \quad (1)$$

where y_{ct} is an outcome of district c in year t , α_c are district fixed effects, and α_t are year fixed effects. When we focus on personnel outcomes, we estimate an analogous equation but using employees as the unit of analysis.³⁹ Our interaction of interest is $\text{Classified}_c \times \text{After}_t$, which takes a value of one in the post-reform period (that is, after 1883) for districts employing at least 50 employees by 1883.⁴⁰ For some specifications, X_{ct} includes census region-year fixed effects as well

³⁶In most cases, these data are reported at the collection-district level. The one exception is that we have separate information on the receipts and expenses for those “ports of delivery” in which the surveyor acted as a collector.

³⁷Receipts from customs accounted for more than 95% of total receipts in our baseline sample of districts.

³⁸In a robustness check, we exclude the New York district from the analysis and find similar results.

³⁹Our personnel results are nevertheless similar if we use data collapsed at the district-year level (Table B6).

⁴⁰The list of “classified” districts remained the same throughout our baseline analysis period (up to 1894). In 1894, the classification was extended to include ports having as many as 20 employees. However, the cutoff was further reduced in May of 1896 to districts having as many as five employees ([Commission, 1897](#)). After this extension, only 63 employees in the entire Customs Service remained outside of the classified service. Hence, these later expansions are not useful to study the effects of the reform since, by 1896, nearly all Customs Service’s employees were within the classified civil service.

as interactions between districts' number of employees in 1883 and year dummies. Throughout the paper, we cluster standard errors at the district level.

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

$$y_{ct} = \alpha_c + \alpha_t + \sum_{t=1875}^{1893} \beta_t \text{Classified}_c \times \alpha_t + \gamma X_{ct} + \epsilon_{ct} \quad (2)$$

where the β_t coefficients capture the differential evolution in outcomes in the classified and non-classified districts during our sample period. When estimating event-study specifications using the district-level data, the omitted category is 1874, the first year for which we have data.

Challenges to Identification and Validity of Research Strategy. A first concern with our strategy is that the 50 employees cutoff might have been chosen so as to include or exclude certain districts. Similarly, districts might have manipulated their number of employees in anticipation of the reform. There are three pieces of evidence that suggest that these possibilities are unlikely. First, there are no districts that, based on their number of employees in 1879 (prior to the introduction of the bill), would have been part of the classified service but which downsized their workforce in order to remain exempted from exams. Moreover, there is no evidence of districts manipulating the *growth* in their number of employees: there are no districts that ended up below 50 employees by 1883 but that would have been above this cutoff had their employee numbers continued growing at the 1871-1879 rate after 1879 (Panel (a) of Figure 2). Finally, Panel (b) of Figure 2 shows no evidence of districts bunching right below the 50 employees cutoff in 1883.

A second concern is that the outcomes in smaller districts could have been on a different trend relative to those in the larger districts. We present a number of tests for the common trends assumption. First, Table 1 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. Each row corresponds to one of our main outcomes, and each column corresponds to a different comparison group (based on a district's number of employees in 1883). Starting from districts with ten employees or more (column 3), we do not reject the hypothesis that all pre-reform coefficients are zero for all outcomes. With this in mind, we use districts with ten employees or more by 1883 as the main comparison group in our analysis. We note, however, that our results are similar when using alternative control groups with fewer or more employees (see Table B2). In Section 5, we also present event-study figures for each of the outcomes, providing graphical support for the common trends assumption.

Finally, to further validate our empirical strategy, we estimate our baseline difference-in-differences specification using alternative "placebo" cutoffs to determine whether a district would have been classified, and find no effects of these placebo reforms (see Table B3). We discuss a number of additional specification and robustness checks at the end of Section 5.1.

5 Results

5.1 Personnel Outcomes

A standard argument in favor of limiting political influence over the bureaucracy is that doing so will lead to efficiency gains along three main dimensions: (1) improving employee qualifications by prioritizing skills over political connections; (2) reducing employee turnover, thus facilitating the accumulation of bureaucratic expertise; and (3) reducing patronage-related "excess" hiring, thus deflating expenses. In this subsection, we investigate whether reformed districts improved along the first two dimensions. We discuss if the reform reduced "excess hiring" in Section 6.

Employees' Professional Background. Classified districts had to hire some of their employees through open and competitive exams. We expect this change to lead to an improvement in employees' professional and educational background if, in the pre-reform period, districts traded-off expertise for political loyalty when screening employees.⁴¹ Alternatively, if hiring individuals with inadequate qualifications was costly for appointing officers, they might have placed a heavy emphasis on expertise even when hiring was discretionary (Brierley, 2019). Finally, the reform – by potentially increasing bureaucrats' job stability – might have increased the appeal of a career in government, thus leading to an improvement in the applicant pool.

To test if the reform led to improvements in employees' qualifications, we use the data linking Customs Service personnel records to population censuses. Specifically, we focus on workers' occupations *prior* to joining the Customs Service.⁴² Note that, since we only observe previous occupations for those employees that we can successfully link to an adult observation in the census, the sample size is smaller than in other exercises using the personnel records. We estimate:

$$y_{ict} = \alpha_c + \alpha_t + \beta \text{Classified}_c \times \text{After}_t + \epsilon_{ict} \quad (3)$$

where y_{ict} is a characteristic of employee i who worked in district c in year t . In cases in which we link an employee to more than one census, we focus on the most recent census year (among those collected prior to Register year t). In Section B.1, we estimate specifications that also control for census region-year fixed effects as well as interactions between districts' number of employees in 1883 and year dummies.

In Panel (a) of Table 2, we focus on the probability that an employee was listed in the census as having a professional occupation (such as lawyers, or accountants).⁴³ Column 1 focuses on the *stock* of employees, which reflects a combination of employees hired before and after the reform, in exempted and non-exempted positions. In this case, we observe a small increase in the likelihood

⁴¹For instance, Colonnelli *et al.* (2020) and Oliveros & Schuster (2018) find that politically-connected bureaucrats tend to be less qualified.

⁴²Unfortunately, US censuses prior to 1940 do not include information on years of schooling. While the census does include information on literacy, it is a very coarse measure of human capital in this context: the literacy rate was above 90% among individuals age 20 to 50 in the 1880 census (the last pre-reform census).

⁴³Professional occupations are those with a value of less than 100 in the 1950 Census Bureau occupational classification system.

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

In Column 2, we instead focus on the *flow* of newly hired employees. Here, we see a larger increase of 6.4 percentage points in the likelihood of previously holding a professional occupation, which is significant at the 5% level. In Columns 3 and 4, we continue to focus on new hires but further distinguish employees based on whether or not they were hired in positions that required an exam. Consistent with the exams helping select more qualified employees, we find the strongest effects among newly hired employees in positions that were subject to exams (Column 4). Specifically, newly-appointed employees in non-exempted positions were 9 percentage points more likely to have held a professional occupation – a very large increase as less than 10% of newly hired employees held a professional occupation in the pre-reform period. In contrast, we find a small and statistically insignificant effect among those hired in exempted positions (Column 3).

Panel (b) shows that the increase in the likelihood of hiring employees with a previous professional occupation was driven by a decline in the likelihood of selecting employees who in the census listed either none or an unskilled occupation.⁴⁴ Similar to the pattern in Panel (a), we find a relatively small decline in the share of such workers when focusing on the employee stock (Column 1), but a larger decline when focusing on new hires in non-exempted positions (Column 4). Overall, these findings show that, after the reform, classified districts replaced workers with occupations requiring low educational attainment (unskilled occupations) with workers with occupations requiring higher educational attainment (professional occupations).⁴⁵

Although we find improvements in the observed occupational background of workers, screening applicants through exams may have shaped recruits' occupational profile without necessarily improving their actual qualifications for the job. For instance, exams might have favored younger workers who had more recent experiences with test-taking and who might have also have been more likely to hold a professional occupation due to secular changes in educational attainment. However, columns 1 to 4 in Table B4 show that the increases in the likelihood that employees would have previously held a professional occupation were not driven by changes in the age composition of the workforce, nor by changes in their geographic origins or racial mix. In this table, we estimate our baseline difference-in-differences specification (equation 3) while including age, birthplace (state for natives and country for foreigners), or race fixed effects. Similarly, Columns 5 to 8 in Table B4 show that the decline in the likelihood of holding an unskilled or no occupation is similar regardless of whether or not we add these fixed effects.

Employee Turnover. As discussed in Section 2, the reform did not provide tenure to civil servants but rather just changed the selection method for employees in certain classified positions. Yet, limiting politicians' discretion to hire might have removed an important incentive to remove

⁴⁴Unskilled occupations are those with a value of 700 or more in the 1950 Census Bureau occupational classification system (for instance, laborers and janitors).

⁴⁵The 1940 census is the first US census to report years of schooling. There were substantial differences in educational attainment between professional and unskilled workers: Male workers aged 30 to 60 with an unskilled occupation completed an average of 8 years of schooling in 1940, whereas those with a professional occupation completed an average of 16.

employees in the first place, thus potentially leading to lower turnover.⁴⁶ To assess whether this was the case, we estimate:

$$Turnover_{ict} = \alpha_c + \alpha_t + \beta Classified_c \times After_t + \epsilon_{ict} \quad (4)$$

where $Turnover_{ict}$ is an indicator that takes a value of one if employee i in district c who was listed in year's $t - 2$ Register was no longer listed in year's t (Registers were collected every two years).⁴⁷

Table 3 shows that the average employee was less likely to leave (either through removal or resignation) after the reform. The reduction in turnover is sizable: employees in reformed districts were 12.6 percentage points less likely to be out of their job by the next Register, relative to a control group mean of 47%.⁴⁸ This finding implies that employees in reformed districts had a longer time horizon over which to accumulate bureaucratic expertise.

If the reduced turnover that we document was due to the reform, we should observe the strongest responses among employees in positions not exempted from exams. Column 2 presents evidence consistent with this prediction. In this column, we report the triple interaction between working in a position subject to exams and working in a classified district in the post-reform period (the specification also includes all the relevant two-way interactions). This interaction is negative and statistically significant: there is an additional 7 percentage points decline in turnover among individuals in these positions. Interestingly, however, there is some evidence of a decline in turnover also among workers in *exempted* positions.⁴⁹

Similarly, if introducing exams reduced turnover by limiting political favoritism in the allocation of jobs, we should also see the largest declines in turnover in years in which there was a party transition at the federal level. To test this hypothesis, we interact the $Classified_c$ indicator with a variable that takes a value of one in 1887 (the first Register year after the presidency went from Chester Arthur, a Republican, to Grover Cleveland, a Democrat) and in 1891 (the first year after it went from Cleveland (D) to Benjamin Harrison (R)). Column 3 of Table 3 shows that the declines in turnover are indeed more pronounced in these years: the interaction term is negative and close to 18 percentage points. In contrast, the main effect is close to zero and statistically insignificant, suggesting that the reform did not significantly affect turnover in years with no party transitions. Figures 3 and 4 show a similar pattern (that is, a large negative gap between the treatment and the control groups in 1887 and 1891) both in the raw data or when we implement an event-study regression.

⁴⁶Indeed, the reports of the Civil Service Commission anticipated this possibility: "when a removal cannot be followed by the appointment of a favorite pressing for the vacancy to be made, most of the temptations to make unjustifiable removals are themselves removed."

⁴⁷Since our personnel data starts in 1871, 1873 is the first year in which we observe if employees had left their job.

⁴⁸For comparison, the *annual* turnover rate in the US federal government was 16% in 2019 <https://www.bls.gov/news.release/jolts.t16.htm>.

⁴⁹The Civil Service Commission reports mention that some collectors avoided removing workers even if they worked in positions exempted from exams, presumably due to concerns about employee morale: "The only exception to this rule is that afforded by Collector Saltonstall, at Boston, who apparently refused to make removals in the excepted and unclassified places of his office for reasons which would not warrant the removal of men from the classified places." ([United States Civil Service Commission, 1890](#))

Overall, these results suggest that the reform led to a professionalization of classified districts' personnel: employees had stronger professional backgrounds to begin with, and this initial advantage was likely amplified as workers had a longer time horizon over which to accumulate bureaucratic expertise.

Robustness of Personnel Results. Online Appendix Section B.1 shows that our results on personnel outcomes are similar when we: (1) implement a randomization inference approach (Figure B1), (2) exclude one classified district at a time (Figure B2), (3) use alternative control groups with collection districts with fewer or more than ten employees (Table B2), (4) control for census region-year fixed effects and/or the interaction between districts' number of employees in 1883 and year dummies (Table B5), or (5) estimate the effects of the reform using the district rather than the employee as the unit of analysis (Table B6). Finally, we also show that there are no effects on personnel outcomes when we use placebo cutoffs for the minimum number of employees instead of the actual cutoff of 50 (Table B3).

5.2 Cost-Effectiveness in Revenue Collection

As discussed in Section 2, proponents of civil service reform argued that the high "cost to collect" in the US Customs Service was to a large extent explained by its lack of a workforce hired on the basis of merit. In this subsection, we use the data on districts' receipts and expenses to ask if the changes in personnel outcomes that we documented above translated into higher cost-effectiveness in customs revenue collection.

Figure 5 provides preliminary evidence suggesting that the reform did not lead to higher cost-effectiveness. The figure shows average log expenses (Panel (a)), log receipts (Panel (b)) and log receipts over expenses (Panel (c)) in classified and non-classified districts, from 1874 to 1893. Prior to the reform, both expenses and receipts evolved in a parallel fashion in both groups of districts. After the reform, receipts and expenses continued to evolve similarly in both groups of districts, with no evidence of a differential change in the classified ones.

Table 4 confirms the pattern of limited effects of the reform on expenses and receipts. This table shows the results of estimating equation 1 using the same outcome variables as in Figure 5. In the even columns, we add region-year fixed effects to account for potential differential trends in economic activity across broad US regions (which would have influenced regional trade levels and hence customs activity). Columns 1 and 2 show no statistically significant effect of the reform on expenses: the effect size is very close to zero and statistically insignificant when we do not include region-time fixed effects, and negative but also statistically insignificant when we include them. These estimates allow us to rule out a reduction in expenses that was larger than 17%.

In Columns 3 and 4, we show that the reform did not lead to a statistically significant increase in total revenue. Specifically, the effect on revenue is small (0.025) and insignificant when not including region-time fixed effects, and actually *negative* (and also statistically insignificant) when including them. These point estimates enable us to rule out an increase in total revenue that was higher than 21%.

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

Figure 6 shows event-study estimates of the effect of the reform on each of these outcomes. The event-study results confirm the pattern of limited effects documented above: regardless of the outcome we consider, we find that none of the pre- or post-reform event-study coefficients are statistically different from zero. Indeed, the event-study coefficients do not show a pattern that would be consistent with a systematic improvement (nor with a deterioration) over time: they are sometimes positive and sometimes negative, both in the pre- and post-reform periods. Consistent with this result, Table B9 shows that we find similar effects if we estimate our baseline difference-in-differences model while excluding the first five years of post-reform data (so as to allow for a lag in the effects to materialize). Overall, we conclude that the reform did not appear to have meaningfully affected districts' financial performance.

6 Mechanisms

Employees in classified districts had stronger professional backgrounds and stayed longer in their jobs. Yet, this professionalization did not translate into measurable improvements in the cost-effectiveness of customs revenue collection. This is despite the fact that, prior to the reform, there was a strong positive correlation between changes in districts' share of employees with a professional occupation and changes in receipts collected (Table B10).⁵⁰ In this section, we discuss possible reasons for this lack of improvement. While we find evidence that the incomplete reach of the reform played a role, the support for other explanations is more limited.

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

Distortions in Districts' Hierarchical Structure. In classified districts, employees paid at a rate of less than \$900 a year were exempted from examinations and could be hired through traditional patronage methods. To the extent that appointing officers wanted to retain their ability to distribute jobs on the basis of patronage, we should observe a post-reform increase in the proportion of workers below this salary cutoff.⁵¹ 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

⁵⁰In this table, we use the pre-reform data to estimate a regression in which the dependent variables are districts' (1) log expenses, (2) log receipts and (3) log receipts over expenses, and the independent variable of interest is the share of workers with a professional occupation. The regressions include district and year fixed effects.

⁵¹Retaining hiring discretion might have also been useful for well-intentioned appointing officers whose ability to hire the best people was constrained by the exam requirement. However, the fact that the reform reduced turnover *only* in years in which the Federal administration changed party hands suggests the importance of partisanship (rather than employees' qualifications) in driving hiring and firing decisions.

room for this kind of distortion.⁵²

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 districts, before and after the reform. Panel (a) focuses on the stock of employees, whereas Panel (b) focuses on new hires. From this figure, it is evident that there was a sharp increase in the proportion of the employees below the \$900 threshold after 1883, and that this increase was particularly pronounced among new hires.

Table 5 confirms that the reform led to a sharp increase in the share of workers in exempted positions. Specifically, employees in reformed districts were 18 percentage points more likely to work in an exempted position (26 percentage points among new hires). For comparison, in the pre-reform period less than 15% of the employees in classified districts worked in positions that would continue not requiring exams after 1883.⁵³ Columns 3 to 6 further split exempted positions into those reserved for political appointees (such as the collector) and their direct assistants, and those exempted due to the \$900 salary cutoff. The reform permitted only a limited number of political appointees, but did not limit the number of employees that could earn less than \$900.⁵⁴ Indeed, the increase in the overall share of employees in exempted positions came exclusively from an increase in the proportion of employees below the \$900 cutoff, with no change in the proportion of workers in other exempted positions. Figure 8 presents event-study estimates confirming the finding of an increased proportion of employees earning below \$900 after 1883.⁵⁵

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

⁵²"Turning to the custom-houses, the Commission is able to present much less satisfactory tables. The classification of the Customs Service has always been very imperfect. It has been classified by salary rather than by employment, and has been possible to take the employees out of the classified grades by lowering their salaries or by changing their designations. 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 to examination. At the time the rule concerning excepted plans was changed in March, 1888, there were only three."(United States Civil Service Commission, 1890).

⁵³Since the reform reduced turnover of employees in non-exempted positions, the increase in the proportion of new hires in exempted positions could be a mechanical consequence of this reduction. However, note that we also observe an increase in the stock of employees in exempted positions. Similarly, the fact that there is an increase in the proportion of workers making less than \$900 among *new hires* suggests that the effects do not simply capture a reduction in the salaries of existing employees.

⁵⁴Besides the collector, the following positions were filled by political appointees and their staff: deputy collectors; cashier of the collector; assistant cashier of the collector; auditor of the collector; chief acting disbursing officer; deputy naval officers; deputy surveyors; one private secretary or one confidential clerk.

⁵⁵Online Appendix Section B.1 shows the robustness of this result to a number of alternative specifications similar to those described at the end of Section 5.1.

did not change this amount (Schmeckebier, 1924).⁵⁶ Hence, districts could reallocate the savings from reduced patronage appointments in the positions now subject to exams toward patronage appointments in the still exempted positions (rather than toward non-personnel expenses).⁵⁷

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

Finally, Table B13 provides direct evidence showing that the increase in the proportion of employees in low-paid positions dampened the effects of the reform on personnel outcomes. In this table, we focus on the effects of the reform on turnover (Columns 1 and 2) and on the likelihood that an employee would have held a professional occupation (Columns 3 and 4). In the even columns, we add *position* fixed effects, where a position is defined as the combination of an occupation and a salary (for instance, a \$1200 clerk). When we add position fixed effects (thus fixing the composition of workers across positions), we find a larger reduction in turnover as well as a larger increase in the likelihood that an employee would have held a professional occupation (although the point estimates are not significantly different from each other). These findings provide suggestive evidence that the improvements in personnel outcomes would have been stronger had districts maintained the same hierarchical structure as in the pre-reform period.

The Role of Collectors. Each collection district was administered by a “collector of customs”. Collectors had significant prerogatives, including the ability to appoint and remove employees.⁵⁸ However, the reform did not introduce any changes with respect to how collectors would be selected: until the abolition of this position in 1965, collectors continued to be political appointees, nominated by the President and confirmed by the Senate.

This continuity could be important in explaining the lack of improvements in cost-effectiveness. First, the fact that collectors continued to be political appointees likely facilitated the distortions in personnel structure documented above. Second, if a district’s performance depended on the characteristics of its top official (as argued in the historical literature, see Parrillo, 2013; Rao, 2016), then the lack of change in how this person was selected might have limited the reform’s ability to improve cost-effectiveness. Indeed, US state and local civil service reforms that maintained political discretion to select top officials have been associated with more limited improvements than

⁵⁶This appropriation was set at \$5,500,000 in 1871. Districts were also allowed to keep the receipts that they collected from “fines, penalties and forfeitures” and “labor, drayage and storage” for their operating expenses (Schmeckebier, 1924).

⁵⁷Due to the higher average wage of merit hires (as compared to patronage appointments), each merit hire could be replaced by more than one low-paid non-merit hire without increasing total personnel expenses. Indeed, Column 1 in Table B11 shows that the point estimate on the total number of employees is positive (although statistically not significant).

⁵⁸Collectors’ functions included employing “proper persons as weighers, gaugers, measurers and inspectors at the several ports within his district”.

those which also reduced discretion to select such officials (Ornaghi, 2016; Ujhelyi, 2014).

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

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

$$y_{lct} = \alpha_l + \alpha_c + \alpha_t + \epsilon_{lct} \quad (5)$$

where y_{lct} is an outcome of district c in year t under the leadership of collector l . Testing the hypothesis that collectors do not matter is equivalent to testing whether the collector fixed effects α_l are all equal to zero.⁵⁹ Note that identification of α_l is possible since the data include collectors who served only for a subset of our sample years. Indeed, Panel (a) of Table 6 shows that the collector-fixed effects are highly jointly significant, regardless of the outcome (receipts, expenses or receipts over expenses) we consider. This finding suggests that collectors mattered for districts' financial outcomes.⁶⁰

However, a concern with this test is that collectors' transitions might have been endogenous to districts' outcomes. For instance, collectors might have been more likely to be replaced in districts in which performance was already deteriorating. To assess this possibility, we collected data from the *Journal of the Executive Proceedings of the United States Senate* (Senate, 1875) on each nomination to the position of collector of customs. Importantly, these data include the reason why a new collector had to be nominated (death, removal, resignation or end of term of the previous collector). Table B14 categorizes all transitions into each of these groups, whereas Figure B5 shows the number of such transitions per year. The most common reason for transitions (38% of cases) is term expiration. In particular, it was very likely for collectors to not be renewed if their term expired after the presidency had changed party hands. The second most common motive (32%) are cases in which the collector was removed or suspended. Similar to the case of term expiration, most removals also occurred in years in which the presidency changed party hands. This pattern suggests that transitions were mostly driven by political considerations rather than by performance.⁶¹

⁵⁹In this analysis, we exclude year-districts in which the collector changed and hence there were multiple collectors in a single year; the "collector fixed effects" are not well defined in that case.

⁶⁰These results are related to the literature estimating "leader effects" in other contexts such as CEOs in the private (Bertrand & Schoar, 2003) and public (Janke *et al.*, 2019) sectors, or sports coaches (Berry & Fowler, 2021). An important difference between our findings and those in this literature is that we do not observe a given collector in multiple districts. Hence, we are closer to Besley *et al.* (2011) who identify leader fixed effects out of leaders who stay in power for a subset of years in a given country.

⁶¹Indeed, Table B15 shows that there is no correlation between past performance in revenue collection (or expenses) and the likelihood that a collector would leave or be removed in the following year. In this table, we estimate:

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

where $Transition_{ct}$ takes a value of one if there was a change in the identity of the collector heading district c in year t , and y_{ct-1} are the total receipts, total expenses or total receipts over total expenses of district c in year $t - 1$.

Nonetheless, to further deal with the concern of endogenous collectors' transitions, we implement a second test in which we use deaths of collectors while in office as a shock to leadership (as in [Jones & Olken \(2005\)](#)).⁶² This method compares average districts' outcomes in the T years pre and post a collector's death. The subindex z represents a particular death:

$$\overline{PRE}_z = \frac{1}{T} \sum_t^T y_{zt}^{PRE}$$

$$\overline{POST}_z = \frac{1}{T} \sum_t^T y_{zt}^{POST}$$

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

$$\overline{POST}_z - \overline{PRE}_z \sim N(0, \frac{2\sigma_{\epsilon i}^2}{T})$$

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

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

For each collector's death z , there is a separate set of PRE_z and $POST_z$ dummies: PRE_z is a dummy equal to one in the T years prior to collector z 's death in that collector's district, whereas $POST_z$ is a dummy equal to one in the T years after.⁶³ We follow [Jones & Olken \(2005\)](#) with respect to two choices. First, to exclude any immediate disruption caused by collectors' deaths, we *do not* include the year of the death itself in either the PRE nor the $POST$ dummies. Second, we include all districts in the sample (even those without collectors' deaths). Doing so helps us estimate the other parameters in the regression (for instance, the year fixed effects). Finally, since the death of a collector while in office is a relatively rare event, we include all districts (regardless of their number of employees in 1883) and consider a longer time period than in our baseline analysis (1874-1903 rather than 1874-1893).⁶⁴ Our sample includes a total of 33 deaths over a 30-year period.⁶⁵

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

⁶²Figure A4 shows an example case in which a new collector ("T. Jefferson Jarrett") is nominated to replace a deceased collector ("Peter F. Cogbill") in the Petersburgh, VA district.

⁶³In our baseline analysis, we use $T = 3$. We use a shorter time window than in [Jones & Olken \(2005\)](#) because our sample spans a shorter time period. We also include deaths that are close to the limits of the time window we study for which we observe less than three years before and after the death. In this case, the PRE and $POST$ dummies are equal to one only for the minimum number of periods that we observe before and after. For example, if a death occurred in 1875, each dummy takes a value of one for only one observation before and after. There are only two deaths that fall under that category and the J-statistic continues to be statistically significant at the 5% level for expenses if we exclude those observations.

⁶⁴In Section 6.4, we use these data to investigate longer-term consequences of the reform.

⁶⁵As a comparison [Jones & Olken \(2005\)](#) had 57 deaths in a 40-year period.

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

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

Panel (b) of Table 6 presents the Wald tests for each of our measures of expenses and receipts. We reject the null hypothesis that collectors do not matter for expenses (p-value: 0.006). The J-statistic is 1.76, implying that the variance of expenses is 76 percent higher around collectors' deaths than what it would normally be. Similarly, the variance of receipts is also higher around collectors' deaths, although the evidence that collectors matter is weaker statistically (p-value: 0.14). Finally, our main measure of performance (receipts over expenses) also exhibits excess variability around collectors' deaths (p-value: 0.085).

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

6.2 Was There a Trade-off Between “Expertise” and “Effort”?

We next consider the possibility that, although employees hired through exams might have been better qualified, they might have also exerted less effort than patronage hires. We first note that, in practice, we expect this channel to be less relevant in this context than in other civil service reforms: As discussed in Section 2, the reform *did not* eliminate the possibility that employees would be removed when administrations changed (or at any time). Indeed, employees remained to be employed “at-will”, implying that they could be removed without having to establish a just cause. Hence, the reform did not eliminate the incentive to perform to avoid being terminated.

However, merit hires might have exercised less effort as a result of a potential misalignment between their preferences and politicians' goals.⁶⁶ Empirically testing if this was the case is challenging since we do not have individual-level measures of work effort. Nonetheless, assuming there were fewer incentives to perform, reduced effort should be most noticeable where employees enjoyed greater discretion. Moreover, we should also observe the strongest responses where the misalignment between the preferences of bureaucrats and politicians was potentially more costly for the latter.

With this in mind, we investigate whether the reform led to changes in the total revenue collected from “fines, penalties and forfeitures” and “labor, drayage and storage”.⁶⁷ There are two reasons why focusing on such revenue is a useful test of this hypothesis. First, the collection of

⁶⁶Existing evidence from the British empire actually suggests the opposite effect: bureaucrats exhibited *lower* performance when they were socially connected with their boss (Xu, 2018). Also, note that the reform preserved some residual discretion for selecting the most aligned employee among the top three or four highest-scoring candidates).

⁶⁷“Drayage” fees were collected by customs officers as charges for the transfers of packages from a wharf to appraisers' stores. “Storage” fees were collected in compensation for goods for which no import or export duties had been paid and which had to be stored prior to entering the country.

fines was likely more discretionary than the collection of other sources of customs revenue. For instance, [Parrillo \(2013\)](#) describes how, during the 1860s, a change in the monetary incentives for imposing fines led to a nearly *seven-fold* increase in revenue from this source. Second, districts were allowed to keep the revenue that they collected from both of these sources to finance their operating expenses ([Schmeckebier, 1924](#)). Hence, politically aligned bureaucrats would have likely put a higher effort in collecting such revenue.

Table [B16](#) shows that the reform did not lead to statistically significant decreases in the amount of revenue collected from these sources. Moreover, although the point estimates are in all cases negative (which would be consistent with the hypothesis of reduced bureaucratic responsiveness), there is limited graphical evidence that would suggest a more pronounced relative decline in classified districts after 1883 (Figures [B3](#) and [B4](#)). Overall, we find limited evidence supporting the hypothesis that the reform led to reduced bureaucratic effort.

6.3 Spillovers to Non-Classified Districts

The reform might have affected districts' outcomes by causing a diversion of trade away (or into) the classified districts. On the one hand, if the reform reduced "collusive" corruption, importers willing to engage in corruption so as to reduce their tariff costs might have avoided the classified districts. On the other hand, if the reform reduced "coercive" corruption (or increased efficiency and speed in processing imports), importers might have diverted their business into the classified districts ([Sequeira & Djankov, 2014](#)).⁶⁸ By a similar logic, the reform might also have affected personnel outcomes in non-classified districts. For instance, if workers who could not be hired for patronage jobs in classified districts were instead hired by the non-classified ones.

To test this possibility, we analyze whether proximity to a classified district's port of entry led to different outcomes in the pre- and post-reform periods. Intuitively, if the reform led to a diversion of importers away from classified districts, we should observe that being close to a classified district leads to *higher* receipts in the post-reform period. Alternatively, if the reform made classified districts more attractive for importers, then proximity to a classified district should lead to *lower* receipts after 1883.

Figure [B6](#) shows limited evidence that the reform led to a diversion of trade away or into the classified districts. This figure (which uses the sample of non-classified districts) shows the association between a district's distance to the closest classified district and its total revenue, before and after the reform. The figure shows that there is little correlation between proximity to a classified district 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 such spillovers, we estimate:

⁶⁸ [Sequeira & Djankov \(2014\)](#) define "collusive" corruption as that which "occurs when public officials and private agents collude to share rents generated by the illicit transaction, thus reducing firm-level trade costs" and "coercive" corruption as that which "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."

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

where y_{ct} is an outcome of district c in year t , and $Distance\ to\ Nearest\ Classified_c$ is the physical distance of district c to the closest classified district's port of entry.

Table B17 shows no evidence that would suggest spillovers of the reform toward non-classified districts: proximity to a classified district does not predict either decreases or increases in receipts after the reform. Similarly, Table B18 shows no spillover effects with respect to our main personnel outcomes.⁶⁹

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). To test this possibility, we use data on receipts and expenses that span a longer time period (from 1874 to 1903 rather than from 1874 to 1893).⁷⁰ Yet, using this extended sample we continue to find no evidence of an improvement in cost-effectiveness; if anything, the point estimates become smaller than in our baseline sample (Table B20).

7 Conclusions

Despite several attempts at reform, a well-functioning civil service remains an elusive goal for many developing countries (Schuster, 2017).⁷¹ The historical experience of the US in its transition toward a professionalized bureaucracy offers a window into some of the challenges involved in these reforms. The US experience is particularly relevant from a development perspective, as it illustrates these challenges in a context where party patronage was “fully embedded in political reality” (Grindle, 2012).

This paper focused on the Pendleton Act, a milestone reform in the history of the US civil service. Specifically, our analysis studied the consequences of the act for the functioning of the Cus-

⁶⁹As an alternative approach to assess the role of spillovers, we estimate our baseline difference-in-differences model while excluding from the sample those non-classified districts that were in close proximity to a classified district (and hence were more likely to be affected by such spillovers). Table B19 shows that excluding districts within a radius of 50, 100 or 200 miles from a classified district does not affect our conclusion of the reform having limited effects on cost-effectiveness.

⁷⁰We do not use this longer time period in our main analysis because, as discussed in Section 2, in 1896 the classified Customs Service was further expanded to include districts with five employees or more, after which only 63 employees in the entire US Customs Service remained outside of the classified service. Hence, when using this expanded sample we lack a clearly defined “control group” after 1896.

⁷¹Schuster (2017) shows that there is little association between the existence of a law mandating merit-based hiring in the public sector and survey responses regarding whether hiring in the public sector is actually meritocratic. Moreover, despite substantial heterogeneity in bureaucratic performance across countries, “meritocratic” civil services are, at least in theory, the norm: out of 117 countries with available data, 94 had laws establishing merit-based hiring in the public sector by 2015.

toms Service, a key government agency in charge of the collection of customs revenue –the main source of federal revenue in late 19th century US. Our empirical strategy exploited the fact that the reform mandated exam-based hiring only in collection districts with 50 or more employees, enabling us to compare districts above and below such threshold before and after the reform. We find that the reform improved targeted employees' professional background and reduced turnover. However, there is limited evidence that it increased cost-effectiveness in the Customs Service's main function, the collection of customs revenue. Moreover, the reform induced distortions in districts' hierarchical structure, as districts responded to the reduced hiring discretion by creating additional positions not covered by exams.

What are the broader implications of these findings? As governments try to improve state capacity, the ability to select qualified candidates remains a central concern. To that end, several governments have passed reforms mandating employee recruitment through competitive exams. Our results show that even a parsimonious reform that just changed employees' selection method without granting them tenure improved workers' qualifications and reduced turnover. However, our findings also provide a "cautionary tale" on the effectiveness and design of these reforms. First, even if it improves workers' qualifications, using exams might not by itself lead to rapid improvements in government effectiveness. More broadly, these results illustrate how, by potentially triggering countervailing organizational responses, policies that succeed at improving specific aspects of an organization might nevertheless fail to improve overall organizational performance. Although the distortions in hierarchical structure induced by the reform were perhaps not intended, they were within the "rules of the game". Hence, actors who lost with the reform or disagreed with its goals were able to take actions against its "spirit" without actually disobeying its rules –parallel to how "gaming-the-system" can become an integral part of organizations that implement pay-for-performance contracts (Kerr, 1975; Holmstrom & Milgrom, 1991). Either reformists should internalize the (unintended) organizational changes that reforms might promote or, to the extent possible, limiting the room for such actions should be built-in into reform design.

References

- ABRAMITZKY, RAN, BOUSTAN, LEAH PLATT, ERIKSSON, KATHERINE, FEIGENBAUM, JAMES J., & PÉREZ, SANTIAGO. 2019. *Automated Linking of Historical Data*. Tech. rept. NBER WP 25825.
- AKHTARI, MITRA, MOREIRA, DIANA, & TRUCCO, LAURA. 2019. Political Turnover, Bureaucratic Turnover and The Quality of Public Services.
- 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.
- BERRY, CHRISTOPHER R, & FOWLER, ANTHONY. 2021. Leadership or luck? Randomization inference for leader effects in politics, business, and sports. *Science advances*, **7**(4), eabe3404.
- BERTRAND, MARIANNE, & SCHOAR, ANTOINETTE. 2003. Managing with style: The effect of managers on firm policies. *The Quarterly journal of economics*, **118**(4), 1169–1208.
- BERTRAND, MARIANNE, BURGESS, ROBIN, CHAWLA, ARUNISH, & XU, GUO. 2020. The glittering prizes: Career incentives and bureaucrat performance. *The Review of Economic Studies*, **87**(2), 626–655.
- BESLEY, TIMOTHY, MONTALVO, JOSE G, & REYNAL-QUEROL, MARTA. 2011. Do educated leaders matter? *The Economic Journal*, **121**(554), F205–227.
- BESLEY, TIMOTHY J, BURGESS, ROBIN, KHAN, ADNAN, & XU, GUO. 2021. Bureaucracy and Development.
- 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*.
- COLONNELLI, EMANUELE, TESO, EDOARDO, & PREM, MOUNU. 2020. Patronage and Selection in Public Sector Organizations. *American Economic Review*, **110**(10), 3071–99.
- 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.
- DAHIS, RICARDO, SCHIAVON, LAURA, & SCOT, THIAGO. 2020. Selecting Top Bureaucrats: Admission Exams and Performance in Brazil.
- 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.
- GORDON, ROGER, & LI, WEI. 2009. Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation. *Journal of Public Economics*, 93(7-8), 855–66.
- GRINDLE, MERILEE S. 2012. *Jobs for the Boys*. Harvard University Press.
- HOLMSTROM, BENGT, & MILGROM, PAUL. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, 7, 24.

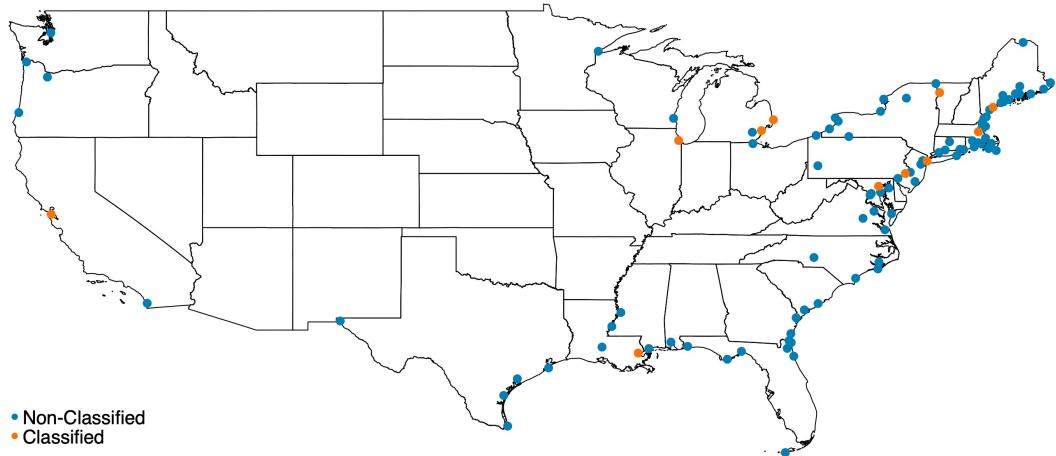
- HOOGENBOOM, ARI. 1959. The Pendleton Act and the civil service. *The American Historical Review*, **64**(2), 301–318.
- HOOGENBOOM, ARI ARTHUR. 1968. *Outlawing the spoils: a history of the civil service reform movement, 1865–1883*. Vol. 50. University of Illinois Press.
- JANKE, KATHARINA, PROPPER, CAROL, & SADUN, RAFFAELLA. 2019. *The impact of ceos in the public sector: Evidence from the english nhs*. Tech. rept. National Bureau of Economic Research.
- JENSEN, ANDERS. 2019. *Employment structure and the rise of the modern tax system*. Tech. rept. National Bureau of Economic Research.
- 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.
- KERR, STEVEN. 1975. On the folly of rewarding A, while hoping for B. *Academy of Management journal*, **18**(4), 769–783.
- 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.
- KHAN, ADNAN Q., KHWAJA, ASIM IJAZ, & OLKEN, BENJAMIN A. 2019. Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings. *American Economic Review*, **109**, 237–270.
- 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, 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.
- MUÑOZ, PABLO, & PREM, MOUNU. 2020. Managers' Productivity and Labor Market: Evidence from School Principals. *Documentos de Trabajo*.
- NARITOMI, JOANA. 2019. Consumers as tax auditors. *American Economic Review*, **109**(9), 3031–72.
- 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.
- PEPINSKY, THOMAS B., PIERSKALLA, JAN H., & SACKS, AUDREY. 2017. Bureaucracy and Service Delivery. *Annual Review of Political Science*, **20**(1), 249–268.
- 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.
- SLEMROD, JOEL, & YITZHAKI, SHLOMO. 2002. Tax Avoidance, Evasion, and Administration. *Pages 1423–70 of: Handbook of Public Economics*, vol. 3. Elsevier.
- 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.
- TEORELL, JAN, DAHLSTRÖM, CARL, & DAHLBERG, STEFAN. 2011. The QoG expert survey dataset. Available at SSRN 3569575.
- THERIAULT, SEAN M. 2003. Patronage, the Pendleton Act, and the Power of the People. *The Journal of Politics*, **65**(1), 50–68.
- TREASURY 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.

- 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.
- WEAVER, JEFFREY. 2020. Jobs for sale: Corruption and misallocation in hiring. Available at SSRN 3590721.
- WELLS, DAVID AMES. 1866. *Report of the Special Commissioner of the Revenue*. Making of America.
- XU, GUO. 2018. The costs of patronage: Evidence from the british empire. *American Economic Review*, **108**(11), 3170–98.
- ZIPARO, JESSICA. 2017. *This Grand Experiment: When Women Entered the Federal Workforce in Civil War-Era Washington*. UNC Press Books.

Figures

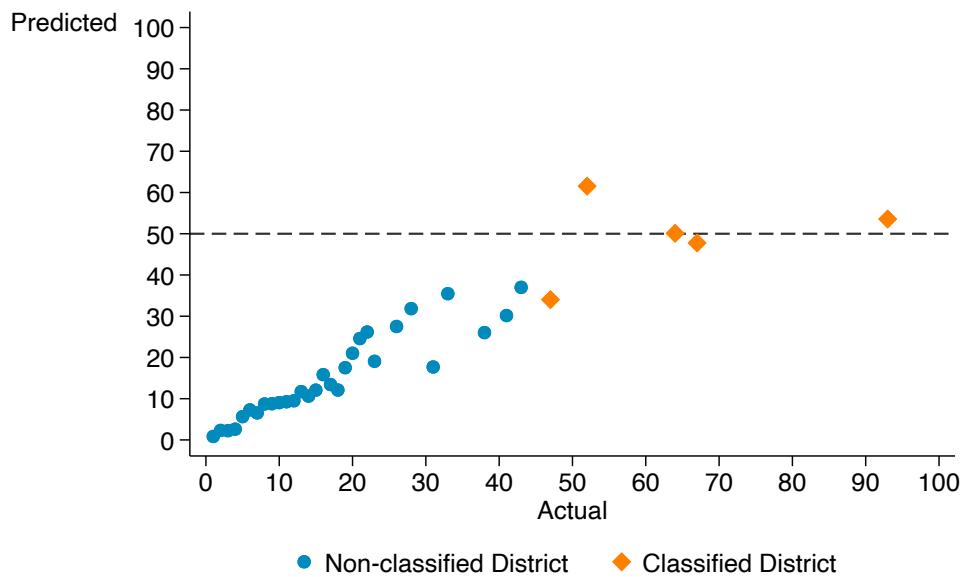
FIGURE 1: LOCATION OF US PORTS OF ENTRY



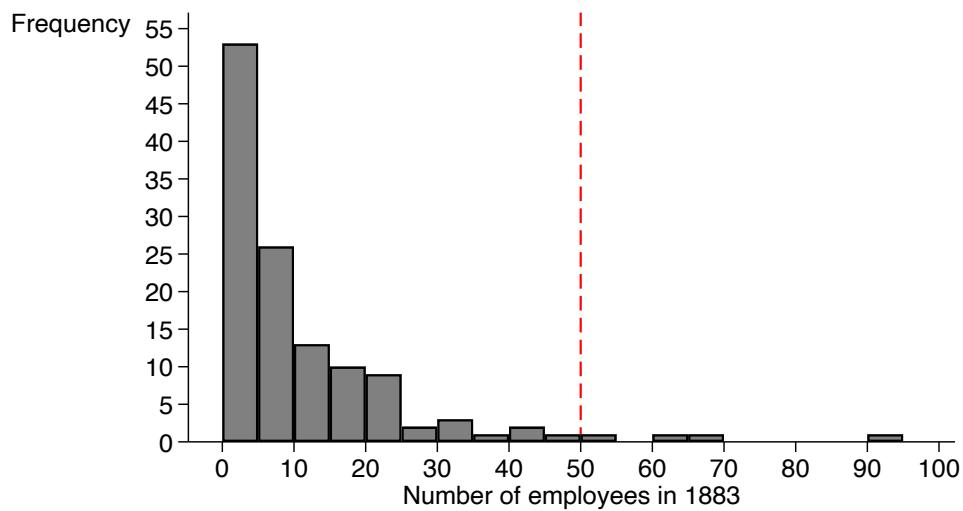
Notes: This map shows the location of customs collection districts “ports of entry”, distinguishing between those that correspond to districts that were required to hire though examinations after 1883 (“classified”) and those that were not (“non-classified”).

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

(A) PREDICTED VERSUS ACTUAL NUMBER OF EMPLOYEES IN 1883

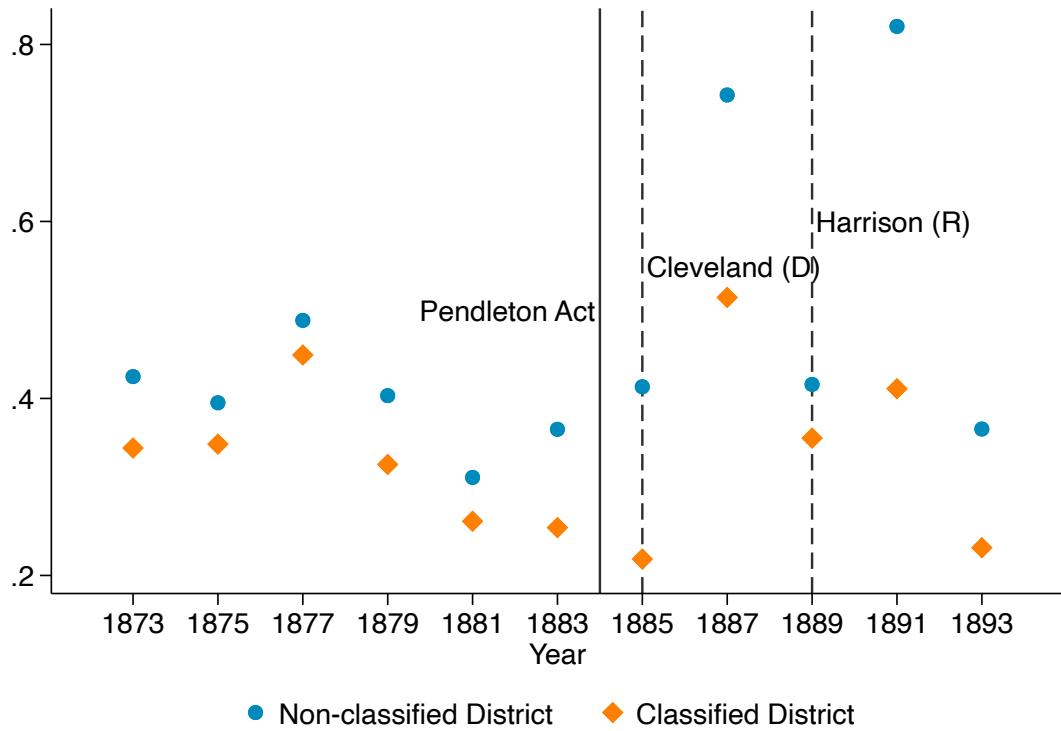


(B) DISTRIBUTION OF THE NUMBER OF EMPLOYEES IN 1883



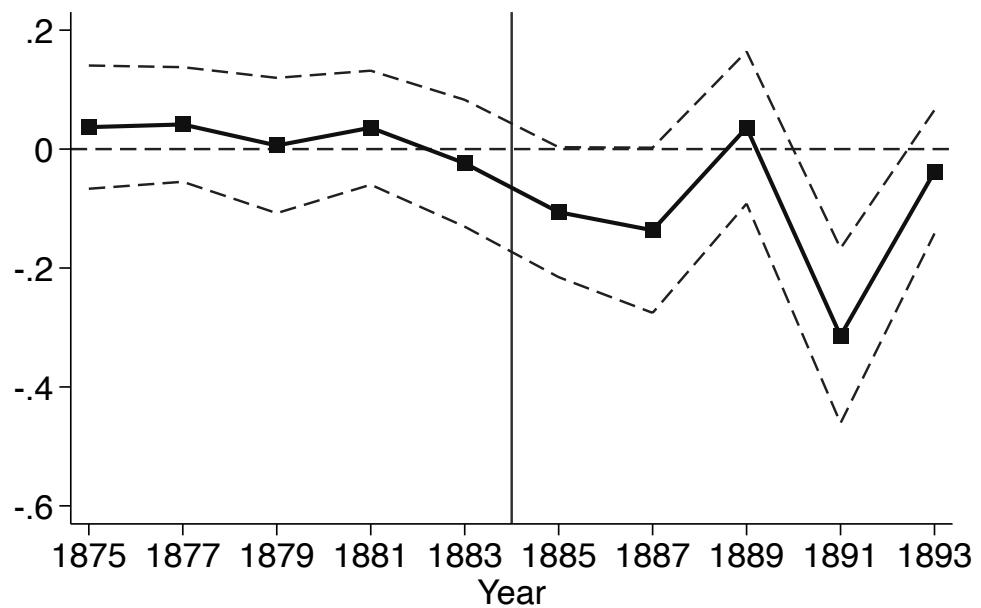
Notes: Panel (a) shows the actual (x-axis) and predicted (y-axis) number of employees of each US customs collection district in 1883. The predicted number of employees is computed by extrapolating a district's number of employees in 1879 based on its observed growth rate from 1871-1879. The figure shows that there are no districts that were predicted to be above the 50 employees threshold by 1883, but that ended up below it. Panel (b) shows the empirical distribution of the number of employees across districts in 1883. The vertical dashed line corresponds to the 50 employees cutoff above which districts were required to hire using exams. The histogram is restricted to districts with at most 100 employees in 1883 so as to more easily visualize the distribution around the cutoff. Districts are grouped in bins of 5 employees (1-5, 6-10, etc.).

FIGURE 3: PROBABILITY OF TURNOVER



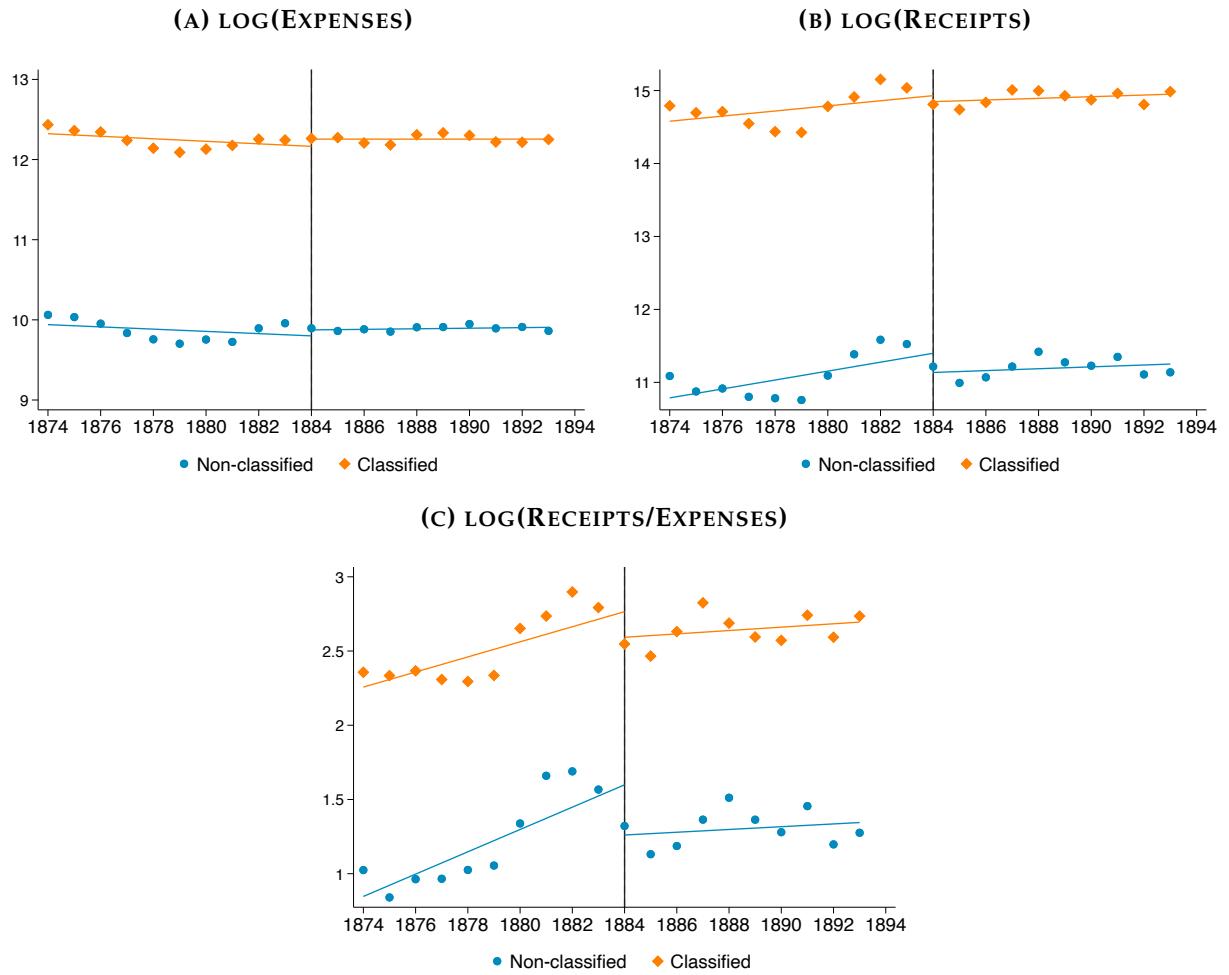
Notes: The y-axis shows the proportion of employees listed in the “Official Register” of year $t - 2$ as working in district c who were no longer listed in year’s t Register. The dashed vertical lines correspond to years in which the Presidency went from a Republican to a Democrat or vice versa. The sample is restricted to districts with at least 10 employees by 1883.

FIGURE 4: PROBABILITY OF TURNOVER, EVENT-STUDY



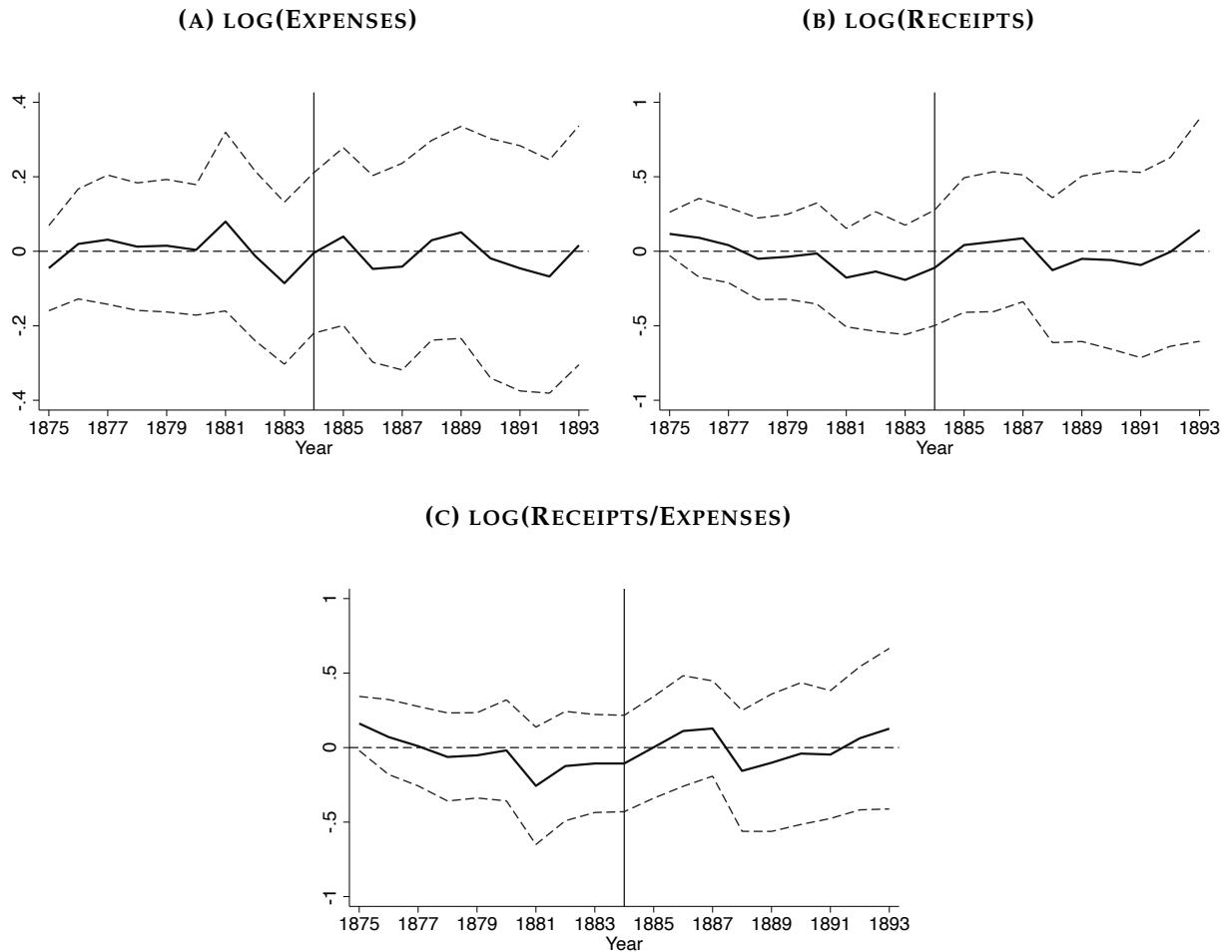
Notes: The dependent variable is an indicator that takes a value of one if employee i in district c who is listed in the "Official Register" of year $t - 2$ as working in district c is no longer listed in year t (the Registers were published every two years). The figure shows the estimated coefficients corresponding to an interaction between a "Classified" indicator and year dummies. The omitted category is 1873. The sample is restricted to districts with at least 10 employees by 1883. Standard errors clustered at the district level.

FIGURE 5: AVERAGE EXPENSES AND RECEIPTS, 1874-1893



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

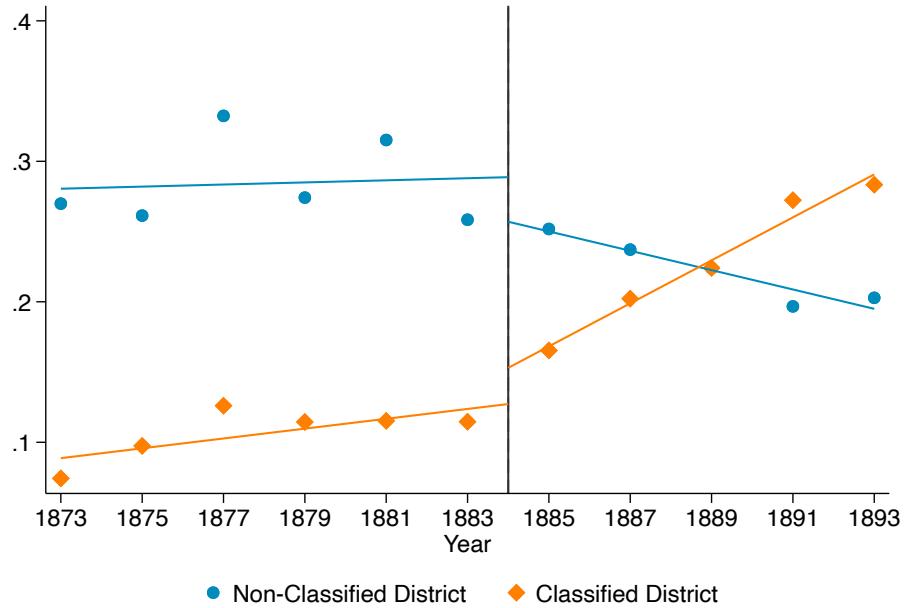
FIGURE 6: EFFECTS OF THE REFORM ON RECEIPTS AND EXPENSES, EVENT-STUDY



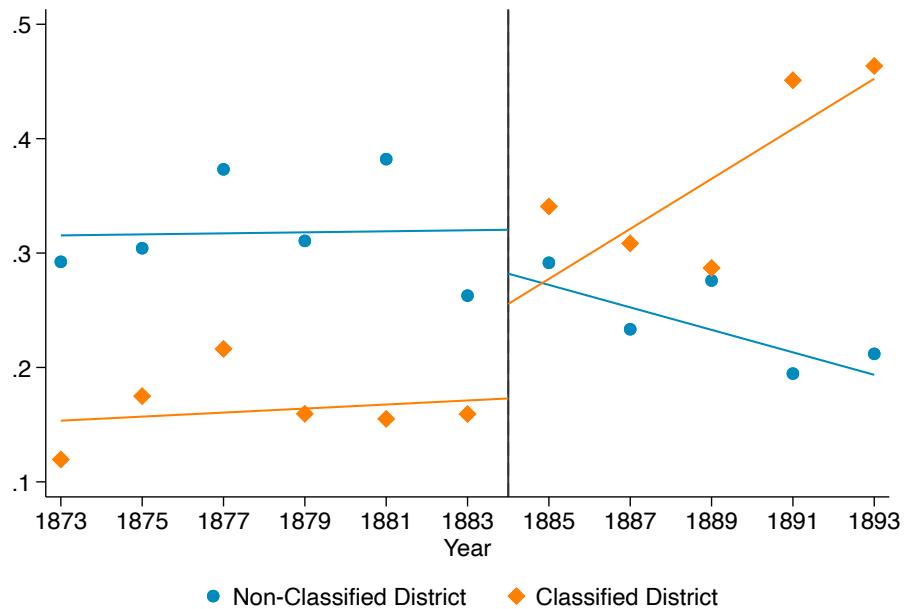
Notes: This figure uses the data on receipts and expenses from the Annual Reports of the Secretary of the Treasury ([US Congress, 1874-1893](#)). The figure shows the estimated coefficients corresponding to an interaction between a "Classified" indicator and year dummies. The omitted category is 1874.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

FIGURE 7: SHARE OF EMPLOYEES BELOW THE EXAM CUTOFF

(A) STOCK OF EMPLOYEES

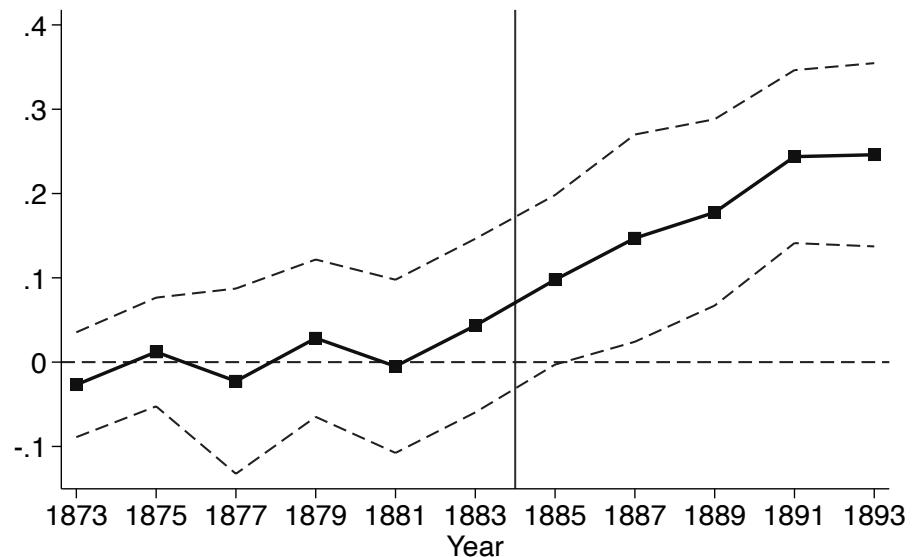


(B) FLOW OF NEWLY-HIRED EMPLOYEES



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 districts). Panel (a) focuses on the stock of employees in year t , whereas panel (b) focuses on the flow of newly hired employees. The sample is restricted to districts with at least 10 employees by 1883.

FIGURE 8: SHARE OF EMPLOYEES BELOW THE EXAM CUTOFF, EVENT-STUDY



Notes: The dependent variable is an indicator that takes a value of one if an employee worked 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. The omitted category is 1871. The sample is restricted to districts with at least 10 employees by 1883. Standard errors clustered at the district level.

TABLE 1: F-TESTS FOR PRE-REFORM EVENT-STUDY COEFFICIENTS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean	p-value	Mean	p-value	Mean	p-value	Mean	p-value
<i>A. Expenses and Receipts</i>								
log(Expenses)	0.028	0.112	0.039	0.111	0.002	0.550	-0.072	0.230
log(Receipts)	0.079	0.110	0.078	0.112	-0.040	0.370	-0.069	0.248
log(Receipts/Expenses)	0.051	0.030	0.040	0.023	-0.042	0.399	0.003	0.056
N	1740		1580		940		640	
<i>B. Personnel Outcomes</i>								
Turnover	0.009	0.170	0.006	0.323	0.019	0.469	0.015	0.575
Professional Occ.	-0.020	0.641	-0.020	0.762	-0.037	0.296	-0.033	0.138
Below cutoff	0.034	0.180	0.011	0.267	-0.003	0.433	0.012	0.111
N (Turnover)	44450		43194		41267		37968	
N (Professional)	6433		6223		5938		5469	
N (Below cutoff)	48293		46904		44922		41459	
Comparison group	All		5+		10+		20+	

Notes: Each row corresponds to a different outcome variable. Each group of columns corresponds to a different estimation sample based on the minimum number of employees (0, 5, 10 or 20) of the districts in the control group. Columns 1, 3, 5 and 7 report the mean value of the pre-reform event-study coefficients based on estimating equation 2. Columns 2, 4, 6 and 8 report p-values for the hypothesis that all the pre-1883 event-study coefficients are equal to zero. Panel (a) focuses on outcomes measured at the district 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”. Standard errors clustered at the district level.

TABLE 2: EMPLOYEES' PROFESSIONAL BACKGROUND

(A) PROFESSIONAL OCCUPATION IN CENSUS

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

(B) UNSKILLED OR NO OCCUPATION IN CENSUS

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

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

TABLE 3: PROBABILITY OF TURNOVER

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

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

TABLE 4: EFFECTS OF THE REFORM ON RECEIPTS AND EXPENSES, DIFFERENCE-IN-DIFFERENCES

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

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

TABLE 5: SHARE OF EMPLOYEES IN POSITIONS EXEMPTED FROM EXAMS

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

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

TABLE 6: COLLECTORS AND DISTRICT PERFORMANCE

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

(B) COLLECTOR'S DEATH WALD TEST			
	(1) J	(2) p-value(chi)	(3) $N_z J$
log(Expenses)	1.716	0.006	56.634
log(Receipts)	1.263	0.143	41.669
log(Receipts/Expenses)	1.353	0.085	44.656

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

Online Appendix- Not for Publication

A Data Appendix

A.1 Linking the Official Registers to the Census: Linking Algorithm and Robustness

Linking algorithm. 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 no birthplace information (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 they are observed in the Official Registers (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 censuses 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 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

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

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 for c_1 (including just using exact matches, i.e. $c_1 = 0$).

Figure A5 shows the proportion of individuals that we match to at least one working-age (i.e. when the individual was 18 to 65) observation in the census (and to at least two, three and four, 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 an individual to the 1850, 1860 or 1870 censuses). On average, we are able to find at least one match for about 20% of Customs Service employees. We expect a lower proportion of individuals in later years to be matched to at least one adult observation, as the latest 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 analysis using linked data, we assess how the professional background of Customs Service employees changed with the passing of the Pendleton Act. Our sample in this analysis includes *only* employees of the Customs Service who were successfully linked to at least one observation in the census. Specifically, we compare the characteristics of bureaucrats in classified districts to those in non-classified districts, 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 *differently* for individuals in classified districts after the reform. This is unlikely because our linking procedure is exactly the same throughout all sample years and across districts.

To further alleviate this concern, we estimate our main difference-in-differences specification using as outcome variables: (1) the total number of censuses to which we link an employee, or (2) an indicator that takes a value of one if the employee is linked to at least one census. Figure A6 shows that, while employees in non-classified districts are overall 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 (which *do not* require the linked data) are very similar when we estimate them on the smaller linked sample.

TABLE A1: NO EFFECTS OF THE REFORM ON PROBABILITY OF MATCHING

	(1)	(2)
	At least 1 match	N. of matches
Classified X After	-0.0106 (0.0204)	-0.00342 (0.0286)
District FE	Yes	Yes
Year FE	Yes	Yes
Observations	45344	45344

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 Customs Service employee is successfully matched to at least one observation in the census. The dependent variables in column 2 is instead the total number of censuses to which a Customs Service employee is matched to. Standard errors are clustered at the district level.

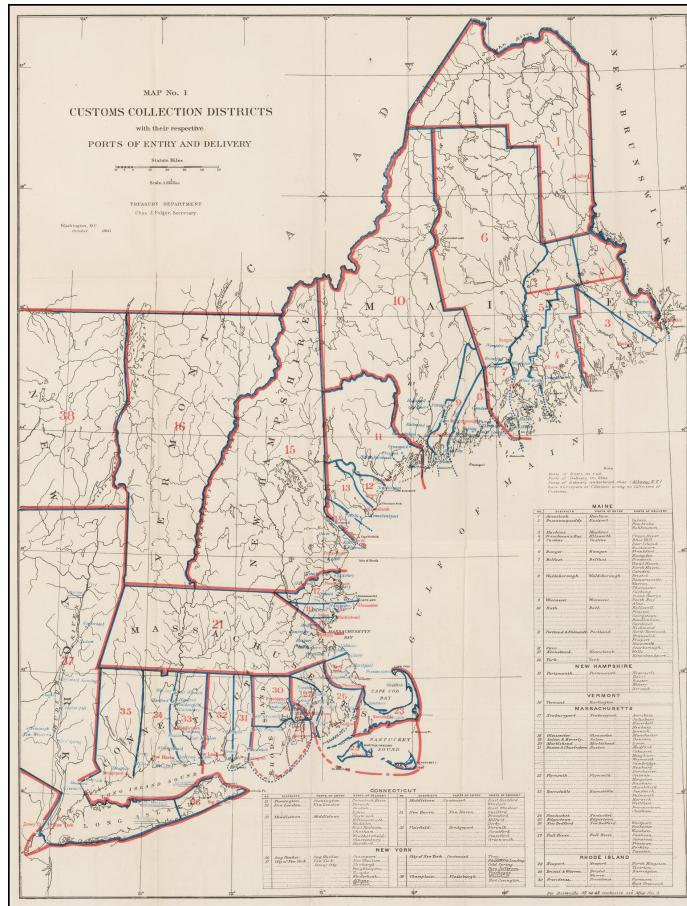
TABLE A2: ALL PERSONNEL OUTCOMES, LINKED SAMPLE

	Turnover		Below Cutoff	
	(1)	(2)	(3)	(4)
Classified X After	-0.126*** (0.0271)	-0.122*** (0.0405)	0.178*** (0.0294)	0.169*** (0.0354)
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	41267	7598	44922	8215
Sample	All	Linked	All	Linked

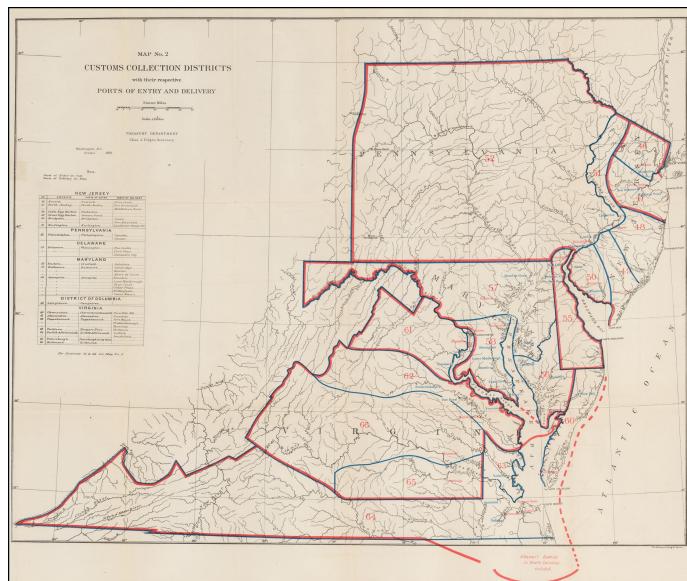
Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table shows the robustness of our personnel results that do not rely on linked register-to-census data to using the linked sample that we use when focusing on employees' occupational background. Standard errors are clustered at the district level.

FIGURE A1: CUSTOMS COLLECTION DISTRICTS IN 1883

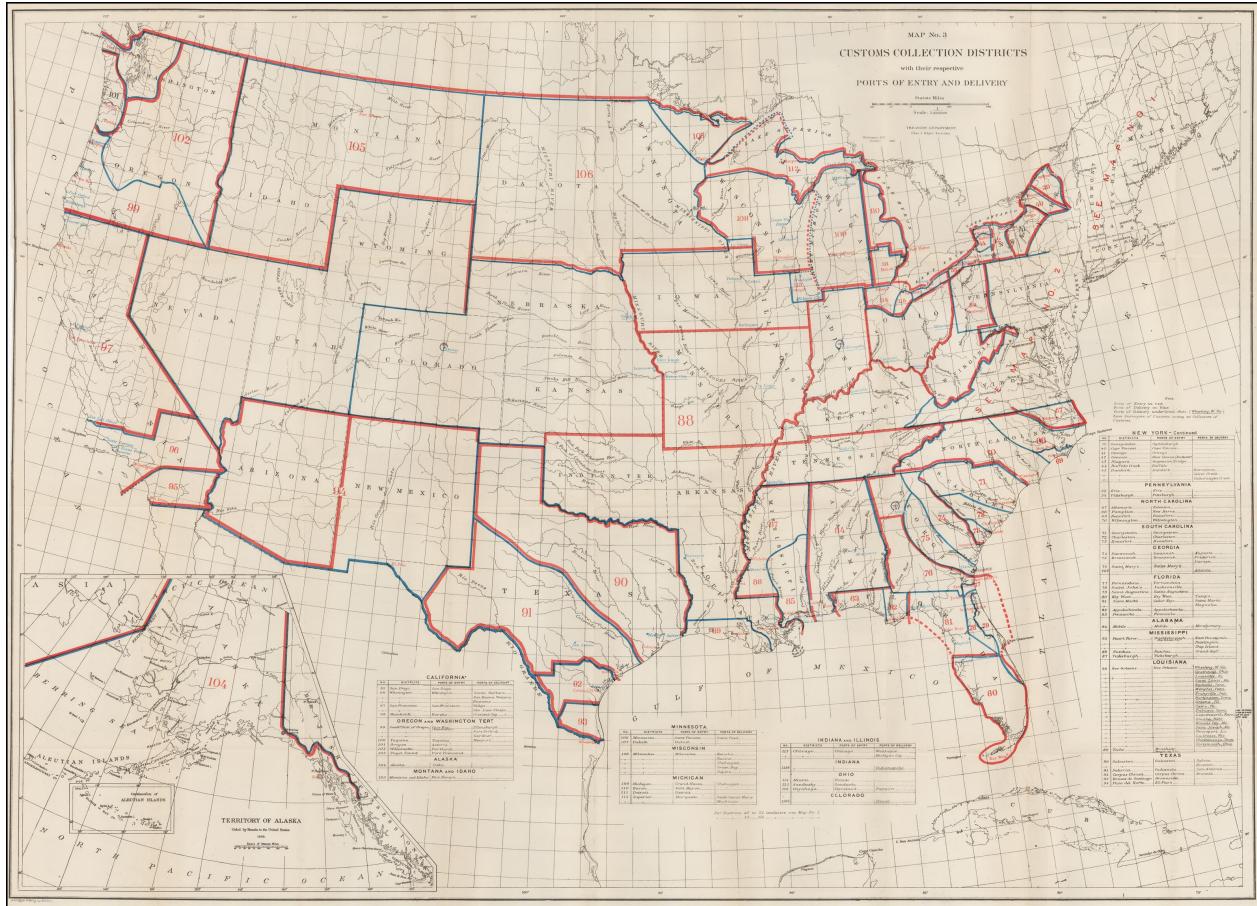
(A) NEW ENGLAND



(B) MIDDLE ATLANTIC



(C) REST OF THE COUNTRY



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

FIGURE A2: EXAMPLE QUESTIONS

(A) ARITHMETIC

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

=====

(B) BOOKKEEPING

SIXTH SUBJECT.—Elements of book-keeping and of accounts.

Exercise.—Samuel Adams, a contractor, had the following dealings with the Treasury Department: He furnished, January 3, 1883, 2,575 lb. of twine, at 12 cents a lb.; April 4, '83, 25 doz. gold pens, at \$25 a doz.; May 7, '83, 645 reams letter-paper, at \$2 a ream; July 9, '83, 45 doz. qt. Arnold's ink, at \$3 a doz.; October 30, '83, 1,000,000 envelopes, at \$2 a thousand; and December 5, '83, 8 doz. inkstands, at \$1.97 a doz. He was paid cash as follows: February 4, 1883, \$175; April 30, \$350; July 15, \$700; November 5, \$2,300; and December 31, 1883, he was allowed on settlement \$45 for cartage, and charged \$75 for breakage and \$60 for shortage on envelopes. State his account in the blank below, with proper heading, and show the balance, if any, due him.

--	--	--	--	--	--	--	--

=====

Notes: This figure shows example questions for applicants to the classified Customs Service. Panel (a) shows a question corresponding to the arithmetic exam, whereas Panel (b) shows a question corresponding to the bookkeeping exam.

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

TREASURY DEPARTMENT.					
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 New York's collector's office 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 Pittsburgh, in the State of Virginia, to succeed Peter F. Cogbill, deceased.

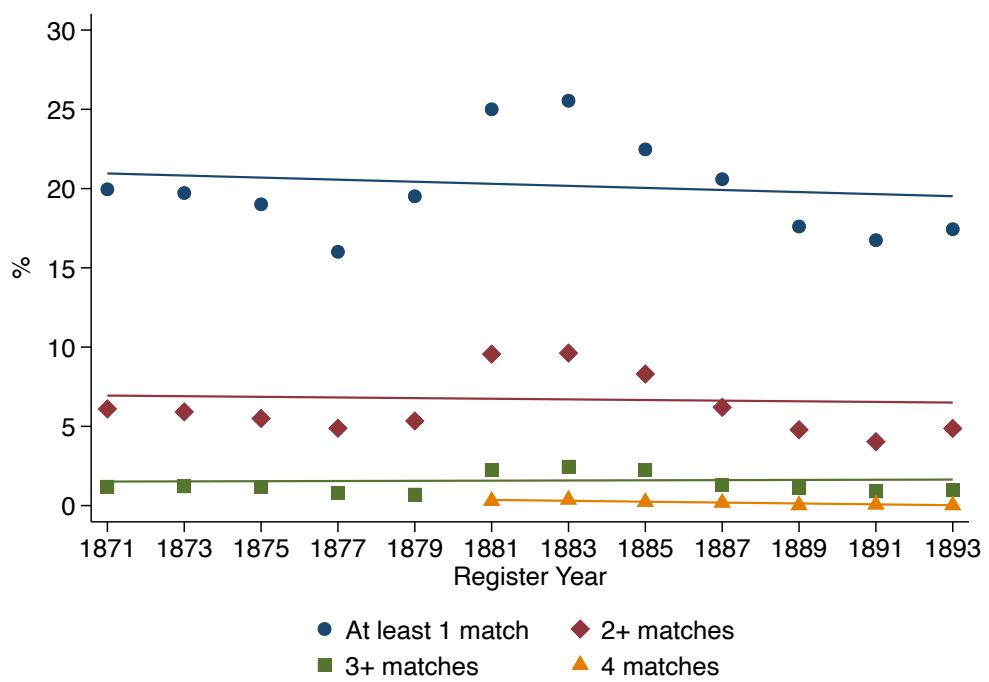
Mr. Jarrett was temporarily commissioned during the recess of the Senate, June 13, 1889.

BENJ. HARRISON.

EXECUTIVE MANSION, December 19, 1889.

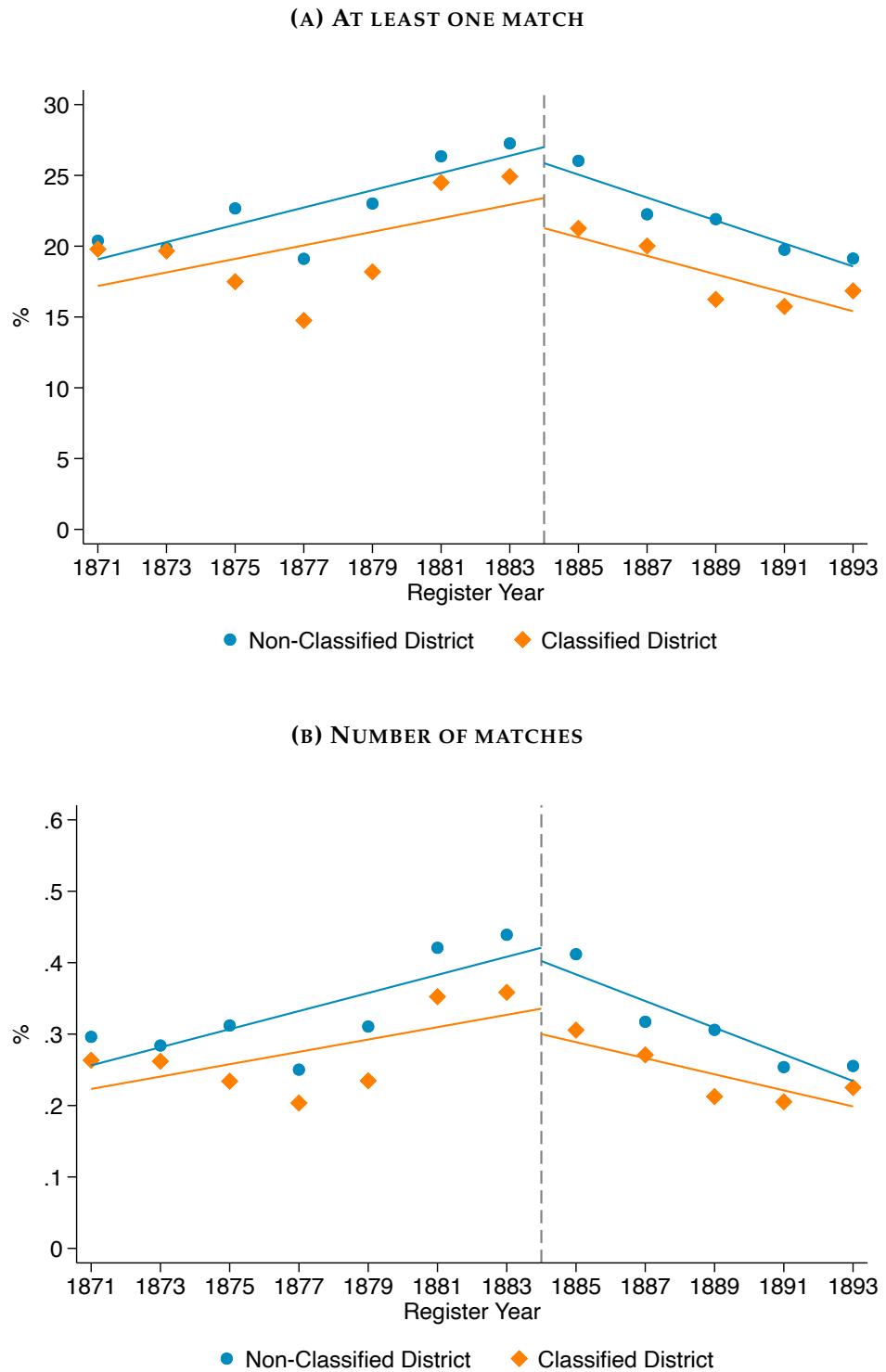
Notes: This figure shows an example page of the *Journal of the Executive Proceedings of the United States Senate (Senate, 1875)*. This page lists the nomination of a new collector who would replace a collector who died while in office.

FIGURE A5: MATCH RATE BY REGISTER YEAR



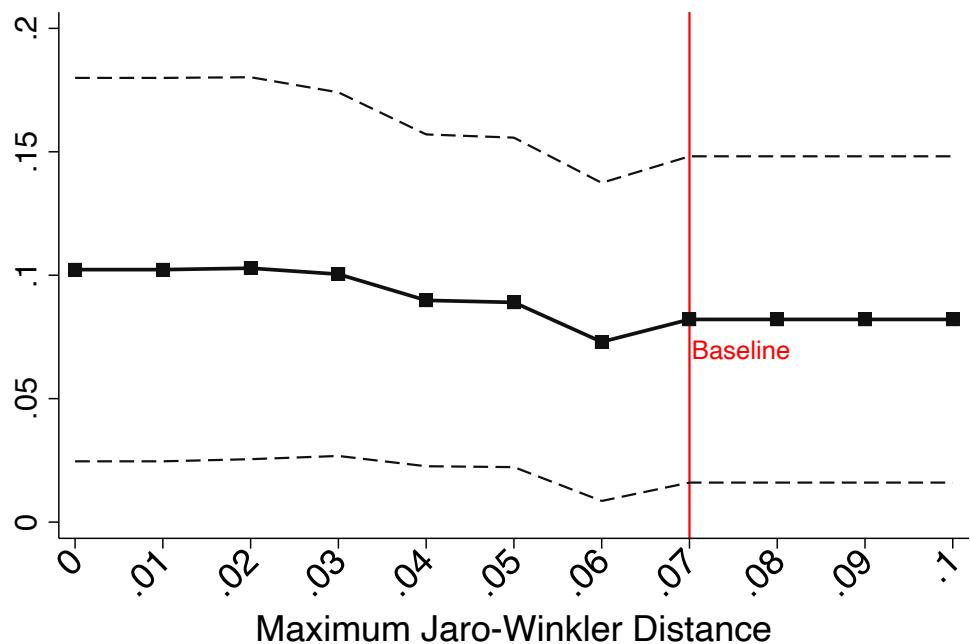
Notes: This figure shows the percent of Customs Service employees that are matched to at least one, two, three or four working-age (aged 18 to 65) observations in the census, by register year.

FIGURE A6: MATCH RATE BY REGISTER YEAR AND CLASSIFICATION STATUS



Notes: Panel (a) shows the percent of Customs Service 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 district. Panel (b) shows instead the average number of censuses to which an observation is linked.

FIGURE A7: ROBUSTNESS TO ALTERNATIVE JARO-WINKLER CUTOFFS



Notes: This figure shows the estimated effect of the reform on the likelihood that a Customs Service employee would have held a professional occupation (y-axis), as a function of the minimum Jaro-Winkler string distance above which an observation would no longer be considered a match (x-axis). Lower values of the Jaro-Winkler distance represent more conservative matches: A Jaro-Winkler distance of zero correspond to two identical strings, whereas a distance of one correspond to two strings with no common characters. The red vertical bar corresponds to the cutoff used in the baseline approach.

B Additional Results

B.1 Robustness of personnel results

Table B3 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 district would have been subject to the reform. In this table, we focus on districts with less than 50 employees and estimate our main difference-in-differences specification using these placebo cutoffs. We restrict the sample to districts with fewer than 50 employees because otherwise the “placebo” treatment group would mechanically include the actual set of reformed districts (i.e. those with 50 or more employees).

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 districts 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 districts 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 districts, a concern is that the effects of the reform might have been driven by changes taking place in one specific classified district. In Figure B2, however, we show that the results are similar when we estimate our baseline difference-in-differences specification while excluding one classified district at a time. The x-axis in this figure indicates the district that we exclude from the regression, and the y-axis shows the estimated effect of the reform on each personnel outcome. The figure shows that the results are very stable regardless of which district we exclude. Hence, our findings are unlikely to be driven by concurrent changes unrelated to the reform that took place in a specific district.

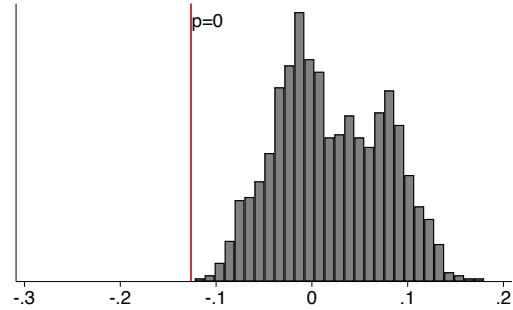
In our baseline specification, our control group is comprised of districts with 10 or more employees by 1883. We use this control group since it has similar pre-trends than the classified districts with respect to all of our main personnel and financial outcomes. However, Table B2 shows that we continue 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.

In Table B6, we show that our main results on personnel outcomes are similar when we estimate regressions at the district level (rather than at the employee level). In this table, the data are collapsed at the district-year level.

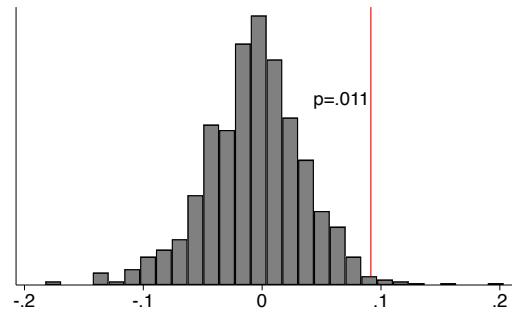
Finally, Table B5 shows that our personnel results are also robust to controlling for: (1) census region-year fixed effects, (2) interactions between a district number of employees in 1883 and year fixed effects, or (3) both at the same time. These results make it unlikely that our findings would be driven by pre-existing differential trends between districts of different size, or by differential trends across broad US regions.

FIGURE B1: PERSONNEL OUTCOMES, RANDOMIZATION INFERENCE

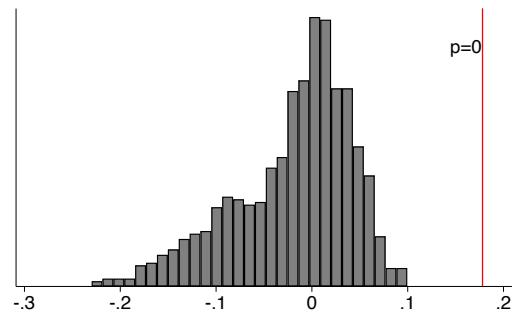
(A) TURNOVER



(B) PROFESSIONAL OCCUPATION



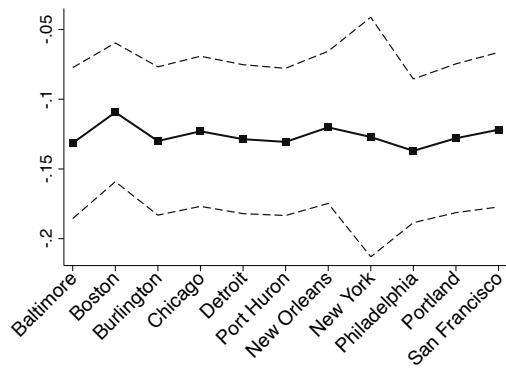
(C) BELOW CLASSIFICATION CUTOFF



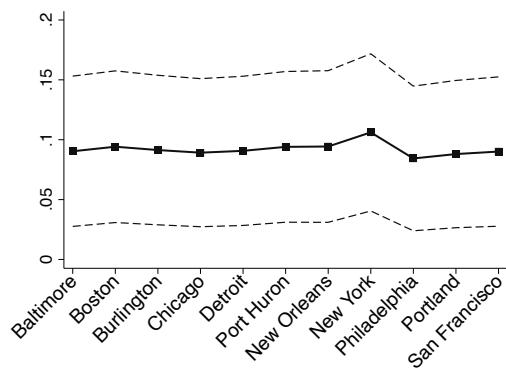
Notes: These figures show the empirical distribution of estimated effects when we implement a randomization inference approach. In this exercise, we randomly select eleven districts as being classified and estimate the “effects” of the reform using our baseline differences-in-differences model. We repeat 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 classified districts. 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 panel (a), column 4 of Table 2. The specification and outcome in Panel (c) correspond to those in column 5 of Table 5.

FIGURE B2: PERSONNEL OUTCOMES, EXCLUDING ONE DISTRICT AT A TIME

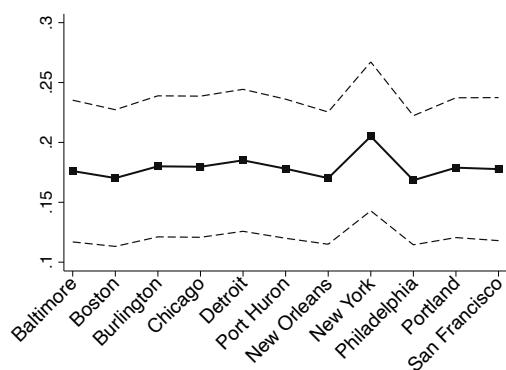
(A) TURNOVER



(B) PROFESSIONAL OCCUPATION

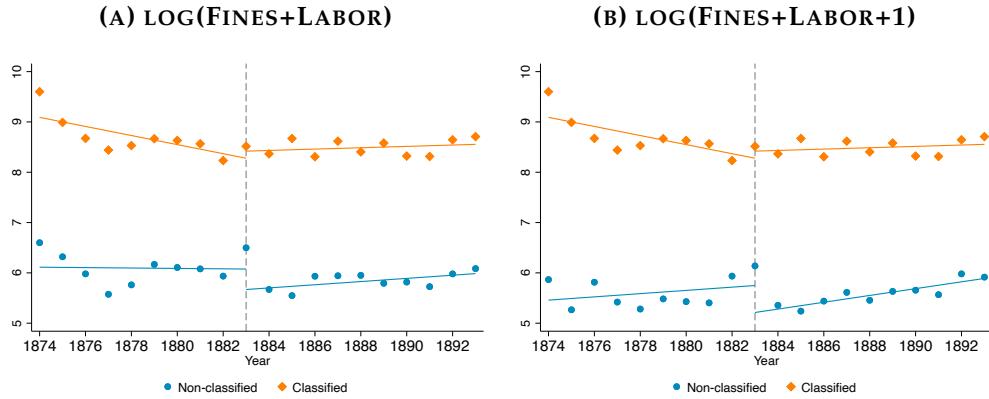


(C) BELOW CLASSIFICATION CUTOFF



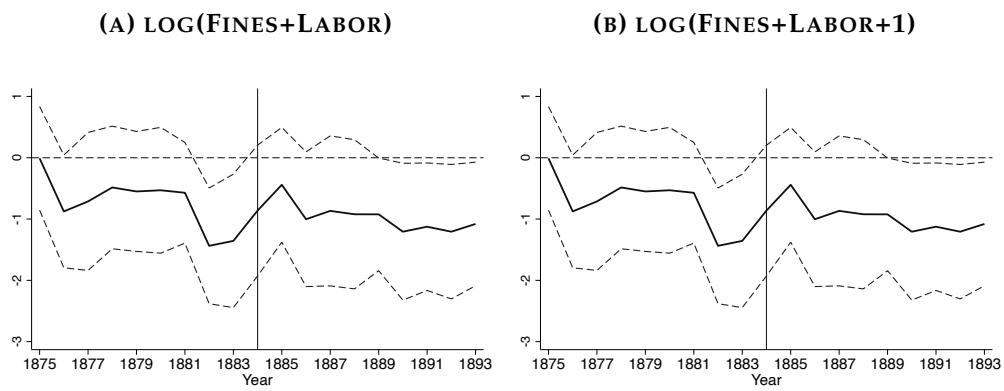
Notes: These figures show the sensitivity of the personnel results to excluding one classified district at a time. The y-axis shows our baseline difference-in-difference estimates around a 95% confidence interval when estimated excluding each of the classified districts indicated in the x-axis. The specification 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 panel (a), column 4 of Table 2. The specification and outcome in panel (c) correspond to those in column 1 of Table 5.

FIGURE B3: AVERAGE REVENUE FROM FINES AND LABOR, 1874-1893



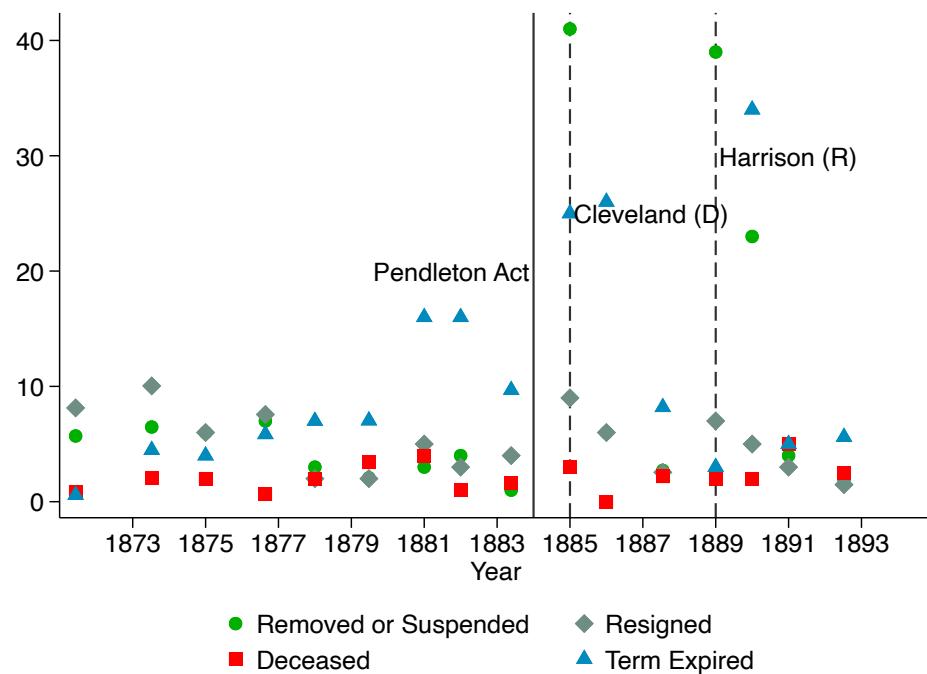
Notes: This figure uses data on the amount of revenue collected from “fines, penalties and forfeitures” and “labor, drayage and storage” from the Annual Reports of the Secretary of the Treasury ([US Congress, 1874-1893](#)). The figure shows yearly average log fines, separately for classified and non-classified districts from 1874 to 1893. The outcome in panel (a) is the log of fines, whereas in panel (b) it is the log of fines plus one.

FIGURE B4: EFFECTS OF REFORM ON REVENUE FROM FINES AND LABOR, EVENT-STUDY REGRESSIONS



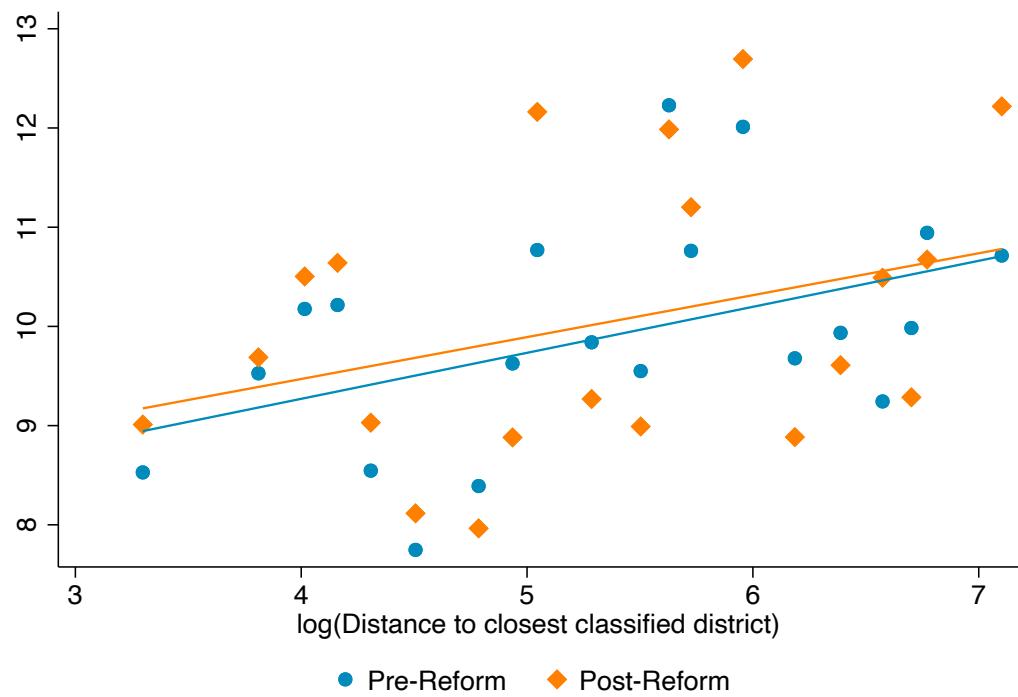
Notes: This figure uses data on the amount of revenue collected from “fines, penalties and forfeitures” and “labor, drayage and storage” from the Annual Reports of the Secretary of the Treasury ([US Congress, 1874-1893](#)). The figure shows event-study coefficients corresponding to estimating equation 2 in the main body of the paper. The outcome in panel (a) is the log of fines, whereas in panel (b) it is the log of fines plus one.

FIGURE B5: MOTIVES FOR COLLECTOR TURNOVER, 1871-1893



Notes: The y-axis shows the yearly number of nominations for the position of collector of customs, separately based on the motive for which a new collector had to be nominated. The dashed vertical lines correspond to years in which the Presidency went from a Republican to a Democrat or vice versa.

FIGURE B6: PROXIMITY TO CLASSIFIED DISTRICTS AND REVENUE



Notes: This figure shows the correlation between the distance to the nearest classified district (based on a district port's of entry) and total receipts, before and after 1883. The sample is restricted to non-classified districts.

TABLE B1: SUMMARY STATISTICS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All	Non-Classified		Non-Classified (10+ emp.)			Classified		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
<i>A. District-level statistics</i>									
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	.	.
# Districts	76	.	.	40	.	.	11	.	.
<i>B. Employee-level statistics</i>									
Turnover	0.47	0.00	0.50	0.47	0.00	0.50	0.34	0.00	0.47
Professional Occ.	0.05	0.00	0.21	0.04	0.00	0.21	0.04	0.00	0.19
Below cutoff	0.31	0.00	0.46	0.26	0.00	0.44	0.16	0.00	0.37
# Observations	12344	.	.	8661	.	.	36683	.	.

Notes: Panel (a) presents district-level summary statistics based on data 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 district-year. These data cover the 1874-1893 period and are annual. Panel (b) is based on Customs Service 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 districts (i.e. those that were not required to hire through competitive exams after 1883). Columns 4 to 6 show statistics for non-classified districts that had 10 or more employees by 1883. Columns 7 to 9 show statistics for the 11 classified districts.

TABLE B2: PERSONNEL OUTCOMES, ROBUSTNESS TO ALTERNATIVE CONTROL GROUPS

	Turnover			Professional Occ.			Below Cutoff		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Classified X After	-0.106*** (0.0252)	-0.126*** (0.0271)	-0.132*** (0.0288)	0.0857*** (0.0306)	0.0915*** (0.0317)	0.0937** (0.0368)	0.174*** (0.0269)	0.178*** (0.0294)	0.174*** (0.0358)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	44683	41267	37968	1960	1855	1702	48286	44922	41459
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 districts (regardless of their number of employees in 1883). Columns 2, 5 and 8 correspond to our baseline sample (using districts with 10+ employees in 1883 as the control group). In columns 3, 6 and 9, the control group is restricted to districts with 20+ employees by 1883. The specification and outcome in columns 1 to 3 correspond to those in column 1 of Table 3. The specification and outcome in columns 4 to 6 correspond to those in panel (a), column 4 of Table 2. The specification and outcome in columns 7 to 9 correspond to those in column 5 of Table 5. Standard errors are clustered at the district level.

TABLE B3: PLACEBO, PERSONNEL PRACTICES

(A) 20+ EMPLOYEES

	(1) Turnover	(2) Professional Occ.	(3) Below Cutoff
Placebo Classified X After	0.0127 (0.0354)	0.0129 (0.0731)	0.0162 (0.0453)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	8470	477	8979

(B) 30+ EMPLOYEES

	(1) Turnover	(2) Professional Occ.	(3) Below Cutoff
Placebo Classified X After	-0.0195 (0.0296)	-0.0666 (0.0604)	0.00774 (0.0504)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	8470	477	8979

(C) 40+ EMPLOYEES

	(1) Turnover	(2) Professional Occ.	(3) Below Cutoff
Placebo Classified X After	-0.0135 (0.0237)	0.0233 (0.0835)	0.0106 (0.0551)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	8470	477	8979

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table show the results of estimating our baseline difference-in-differences model using placebo cutoffs for the number of employees above which a district would have been classified. Panel (a) uses a placebo cutoff of 20 employees, panel (b) a cutoff of 30 and panel (c) a cutoff of 40. The specification and outcome in column 2 correspond to those in column 1 of Table 3. The specification and outcome in column 2 correspond to those in panel (a), column 4 of Table 2. The specification and outcome in column 3 correspond to those in column 5 of Table 5. The sample is restricted to employees who worked in districts with less than 50 employees (so as to exclude the actual set of “treated” districts from the sample). Standard errors are clustered at the district level.

TABLE B4: EFFECTS ON OCCUPATIONAL BACKGROUND, ADJUSTING FOR AGE, BIRTHPLACE AND RACE FIXED EFFECTS

	Professional Occ.				Unskilled			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Classified X After	0.0915*** (0.0317)	0.0845*** (0.0299)	0.0912*** (0.0306)	0.0894*** (0.0316)	-0.109** (0.0478)	-0.0984** (0.0440)	-0.0984** (0.0434)	-0.0957** (0.0463)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	No	No	No	Yes	No	No
Birthplace FE	No	No	Yes	No	No	No	Yes	No
Race FE	No	No	No	Yes	No	No	No	Yes
Observations	1855	1855	1855	1855	1855	1855	1855	1855

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In this table, we estimate our baseline effects of the reform on employees' professional background while adjusting for age, birthplace and race fixed effects. The specification and outcome in columns 1 to 4 correspond to those in panel (a), column 4 of Table 2. The specification and outcome in columns 5 to 8 correspond to those in panel (b), column 4 of Table 2. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B5: EFFECTS ON PERSONNEL OUTCOMES, ADDITIONAL CONTROL VARIABLES

	Turnover			Professional Occ.			Below Cutoff		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Classified X After	-0.120** (0.0480)	-0.120*** (0.0294)	-0.117** (0.0509)	0.111*** (0.0348)	0.0869*** (0.0302)	0.112*** (0.0352)	0.208*** (0.0376)	0.182*** (0.0278)	0.212*** (0.0335)
District 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	41267	41267	41267	1855	1855	1855	44922	44922	44922

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 a district number of employees in 1883 and year dummies, and (2) interactions between census region and year dummies. The specification and outcome in columns 1 to 3 correspond to those in column 1 of Table 3. The specification and outcome in columns 4 to 6 correspond to those in panel (a), column 4 of Table 2. The specification and outcome in columns 7 to 9 correspond to those in column 5 of Table 5. The sample is restricted to districts with at least 10 employees by 1883. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B6: EFFECTS ON PERSONNEL OUTCOMES ESTIMATED AT THE DISTRICT LEVEL

	Turnover		Professional Occ.		Below Cutoff	
	(1)	(2)	(3)	(4)	(5)	(6)
Classified X After	-0.126*** (0.0271)	-0.0884** (0.0424)	0.0915*** (0.0317)	0.0797* (0.0409)	0.178*** (0.0294)	0.141*** (0.0426)
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41267	539	1855	324	44922	588
Unit of analysis	Individual	District	Individual	District	Individual	District

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In this table, we aggregate the employee-level data at the district-year level. The specification and outcome in columns 1 and 2 correspond to those in column 1 of Table 3. The specification and outcome in columns 3 and 4 correspond to those in panel (a), column 4 of Table 2. The specification and outcome in columns 5 and 6 correspond to those in column 5 of Table 5. The sample is restricted to districts with at least 10 employees by 1883. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B7: EFFECTS ON PERSONNEL OUTCOMES, INCLUDING PORTS OF DELIVERY IN CLASSIFIED DISTRICTS IN THE TREATMENT GROUP

	Turnover		Professional Occ.		Below Cutoff	
	(1)	(2)	(3)	(4)	(5)	(6)
Classified X After	-0.126*** (0.0271)	-0.127*** (0.0270)	0.0915*** (0.0317)	0.0758** (0.0325)	0.194*** (0.0293)	0.177*** (0.0288)
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41267	41702	1855	1860	44922	45356
Ports of delivery	No	Yes	No	Yes	No	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In this table, “Ports of Delivery” within classified districts are included in the treatment group. The specification and outcome in columns 1 and 2 correspond to those in column 1 of Table 3. The specification and outcome in columns 3 and 4 correspond to those in panel (a), column 4 of Table 2. The specification and outcome in columns 5 and 6 correspond to those in column 5 of Table 5. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B8: EFFECTS ON PERSONNEL OUTCOMES, EXCLUDING PORTS OF DELIVERY

	Turnover		Professional Occ.		Below Cutoff	
	(1)	(2)	(3)	(4)	(5)	(6)
Classified X After	-0.126*** (0.0271)	-0.106*** (0.0314)	0.0915*** (0.0317)	0.0806** (0.0326)	0.178*** (0.0294)	0.181*** (0.0304)
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41267	40518	1855	1809	44922	44111
Ports of delivery	Yes	No	Yes	No	Yes	No

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In this table, we exclude from the sample those observations that correspond to employees who worked in "ports of delivery". The specification and outcome in columns 1 and 2 correspond to those in column 1 of Table 3. The specification and outcome in columns 3 and 4 correspond to those in panel (a), column 4 of Table 2. The specification and outcome in columns 5 and 6 correspond to those in column 5 of Table 5. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B9: 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 (0.111)	-0.115 (0.0866)	0.0231 (0.244)	-0.187 (0.229)	0.0381 (0.171)	-0.0715 (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 X After* takes a value of one for districts that were part of the classified Customs Service after 1883. Odd columns include year and district fixed effects. Even columns also include *Region X Time* fixed effects. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B10: ASSOCIATION BETWEEN THE SHARE OF WORKERS WITH A PROFESSIONAL BACKGROUND AND DISTRICTS' FINANCIAL OUTCOMES

	log(Expenses)	log(Receipts)	log(Receipts/Expenses)
	(1)	(2)	(3)
Share with Professional Occ.	-0.232 (0.291)	0.553*** (0.179)	0.785* (0.400)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	140	140	140

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. An observation corresponds to district-year. The dependent variable in column 1 is the log of expenses in district c in year t . The dependent variable in column 2 is the log receipts. The dependent variable in column 3 is the log of receipts over expenses. The sample is restricted to the pre-reform period. Standard errors are clustered at the district level.

TABLE B11: EFFECTS OF THE REFORM ON THE NUMBER OF EMPLOYEES

	Total	Non-Exempted	Exempted
	(1)	(2)	(3)
Classified X After	0.0908 (0.0883)	-0.203 (0.126)	0.619*** (0.148)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	588	588	588

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. An observation corresponds to a district-year. The dependent variable in column 1 is the log number of employees in district 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 districts that were made part of the classified system after 1883. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B12: EMPLOYEES PAID BELOW THE EXAM CUTOFF HAD WEAKER PROFESSIONAL BACKGROUNDS

	(1)	(2)	(3)
	Professional Occ.	Unskilled	Literate
Below Exam Cutoff	-0.00548 (0.0195)	0.0998*** (0.0278)	-0.0336* (0.0190)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	1362	1362	1362

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 \$900 a year. The sample is restricted to employees in the *pre-reform* period and to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B13: ACCOUNTING FOR COMPOSITIONAL CHANGES

	Turnover		Professional Occ.	
	(1)	(2)	(3)	(4)
Classified X After	-0.126*** (0.0271)	-0.151*** (0.0267)	0.0278 (0.0176)	0.0364* (0.0191)
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Position FE	No	Yes	No	Yes
Observations	41267	41267	7652	7652

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table shows how the estimated effects of the reform on employee turnover and the likelihood than an employee would have held a professional occupation change as we add position fixed effects. A position is defined as a combination between an occupation (for instance, “clerk”) and a compensation. Adding position fixed effects shuts down the effects of the reform that stem from compositional changes in a district’s workforce. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B14: MOTIVES FOR COLLECTOR TURNOVER, 1871-1893

	(1) Number	(2) %
Deceased	47	8.4
Removed or Suspended	183	32.6
Resigned	116	20.7
Term Expired	215	38.3
Total	561	100

Notes: This table shows the motives why a new collector had to be nominated based on data from the “Journals of Executive Proceedings of the Senate” ([Senate, 1875](#)).

TABLE B15: COLLECTORS TRANSITION AND (LAGGED) DISTRICT PERFORMANCE

	Transition					
	(1)	(2)	(3)	(4)	(5)	(6)
log(Receipts) at t-1	-0.00323 (0.0195)			-0.0120 (0.0208)		
log(Expenses) at t-1		0.0148 (0.0319)			0.00594 (0.0328)	
log(Receipts/Expenses) at t-1			-0.00873 (0.0198)			-0.0140 (0.0205)
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region X Time FE	No	No	No	Yes	Yes	Yes
Observations	939	935	935	939	935	935

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table shows the results of estimating specification 6. The outcome variable *Transition* measures whether there is a new collector at time t . The independent variables are measures of district performance assessed at time $t - 1$. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B16: EFFECTS OF THE REFORM ON REVENUE FROM FINES AND LABOR, DIFFERENCE-IN-DIFFERENCES

	log(Fines+Labor)		log(Fines+Labor+1)	
	(1)	(2)	(3)	(4)
Classified X After	-0.00539 (0.256)	-0.129 (0.199)	-0.173 (0.262)	-0.311 (0.211)
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Region X Time FE	No	Yes	No	Yes
Observations	895	895	940	940

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table uses data on the amount of revenue collected from “fines, penalties and forfeitures” and “labor, drayage and storage” from the Annual Reports of the Secretary of the Treasury ([US Congress, 1874-1893](#)). An observation corresponds to a district-year. *Classified* \times *After* takes a value of one for districts that were part of the classified Customs Service after 1883. All columns include year and district fixed effects. Even columns also include *Region* \times *Time* fixed effects. The sample is restricted to districts with at least 10 employees by 1883. Standard errors are clustered at the district level.

TABLE B17: SPILLOVERS TO NON-CLASSIFIED DISTRICTS, RECEIPTS

	(1)	(2)	(3)	(4)
Distance Closest Classified X After	-0.0413 (0.126)	-0.0511 (0.137)	0.152 (0.143)	-0.00241 (0.215)
District FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1500	1340	720	420
Comparison group	0+	5+	10+	20+

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. An observation corresponds to a district-year. This table shows the correlation between distance to the closest classified district and a district's total receipts. The sample is restricted to non-classified districts.

TABLE B18: SPILLOVERS TO NON-CLASSIFIED DISTRICTS, PERSONNEL OUTCOMES

	(1) Turnover	(2) Professional Occ.	(3) Below Cutoff
Distance to Closest Classified X After	-0.0180 (0.0193)	0.0108 (0.0343)	0.0330 (0.0224)
District FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	7753	449	8238

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 district and personnel outcomes. The sample is restricted to non-classified districts.

TABLE B19: SPILLOVERS TO NON-CLASSIFIED DISTRICTS, EXCLUDING NON-CLASSIFIED DISTRICTS IN CLOSE PROXIMITY TO A CLASSIFIED DISTRICT

	(1)	(2)	(3)	(4)
Classified X After	0.0251 (0.185)	0.00703 (0.194)	-0.00443 (0.206)	0.0283 (0.208)
District 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 districts whose port of entry was at least 50, 100 or 200 miles away from the closest classified district's port of entry.

TABLE B20: EFFECTS OF THE REFORM ON RECEIPTS AND EXPENSES OVER THE LONGER-TERM

	1874-1893		1874-1904	
	(1)	(2)	(3)	(4)
Classified X After	0.0358 (0.137)	-0.0543 (0.138)	0.00656 (0.161)	-0.0665 (0.168)
Region X Time FE	No	Yes	No	Yes
Customhouse FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Comparison group	10+	10+	10+	10+
Observations	940	940	1457	1457

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 districts that were part of the classified Customs Service starting in 1883. Odd columns include year and district fixed effects. Even columns also include *Region* \times *Time* fixed effects. 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). The sample is restricted to districts with at least 10 employees by 1883. Standard errors clustered at the district level.