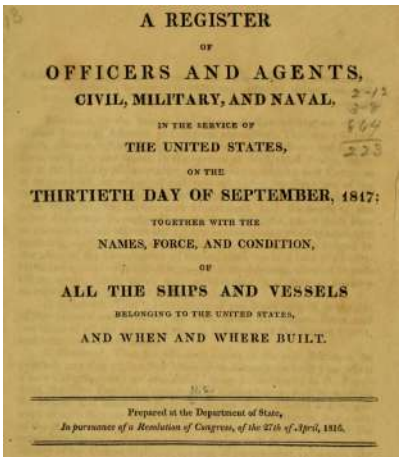
ONLINE APPENDIX

A Additional figures and tables

Figure A1: Examples of pages in the Official Register



A REGISTER

OF

OFFICERS AND AGENTS,

CIVIL, MILITARY, AND NAVAL,

IN THE SERVICE OF

THE UNITED STATES,

ON THE

THIRTIETH DAY OF SEPTEMBER, 1817:

TOGETHER WITH THE

NAMES, FORCE, AND CONDITION,

OF

ALL THE SHIPS AND VESSELS

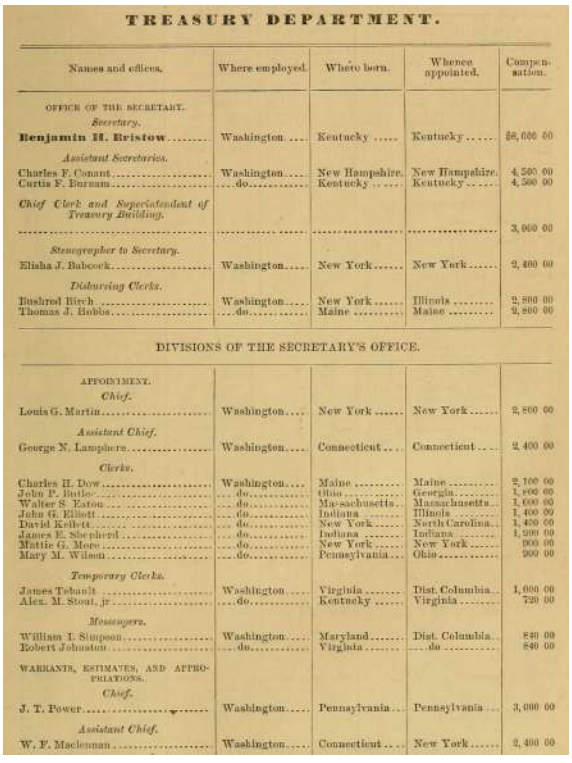
BELONGING TO THE UNITED STATES,

AND WHEN AND WHERE BUILT.

Prepared at the Department of State,

In pursuance of a Resolution of Congress, of the 27th of April, 1816.

(a) 1817 Register



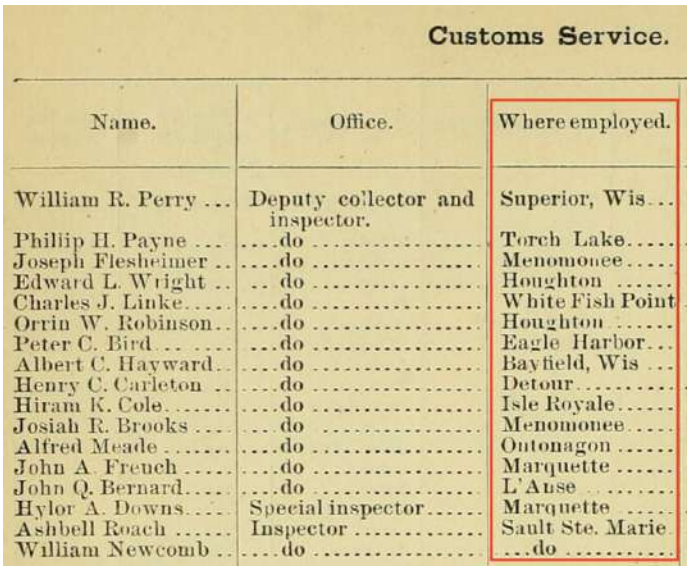
TREASURY DEPARTMENT.

Names and offices.	Where employed.	Where born.	Whence appointed.	Compensation.
OFFICE OF THE SECRETARY.				
Secretary.				
Benjamin H. Bristow	Washington	Kentucky	Kentucky	$8,000 00
Assistant Secretaries.				
Charles F. Conant	Washington	New Hampshire	New Hampshire	4,500 00
Curtis F. Burnam	do	Kentucky	Kentucky	4,500 00
Chief Clerk and Superintendent of Treasury Building.				
				3,000 00
Stenographer to Secretary.				
Elisha J. Babcock	Washington	New York	New York	2,400 00
Disbursing Clerks.				
Bushrod Birch	Washington	New York	Illinois	2,800 00
Thomas J. Hobbs	do	Maine	Maine	2,800 00
DIVISIONS OF THE SECRETARY'S OFFICE.				
APPOINTMENT.				
Chief.				
Louis G. Martin	Washington	New York	New York	2,800 00
Assistant Chief.				
George N. Lamphere	Washington	Connecticut	Connecticut	2,400 00
Clerks.				
Charles H. Dow	Washington	Maine	Maine	2,100 00
John P. Butler	do	Ohio	Georgia	1,800 00
Walter S. Eaton	do	Massachusetts	Massachusetts	1,600 00
John G. Elliott	do	Indiana	Illinois	1,400 00
David Kellett	do	New York	North Carolina	1,400 00
James E. Shepherd	do	Indiana	Indiana	1,200 00
Mattie G. More	do	New York	New York	900 00
Mary M. Wilson	do	Pennsylvania	Ohio	900 00
Temporary Clerks.				
James Tebault	Washington	Virginia	Dist. Columbia	1,000 00
Alex. M. Stout, jr	do	Kentucky	Virginia	720 00
Messengers.				
William I. Simpson	Washington	Maryland	Dist. Columbia	840 00
Robert Johnston	do	Virginia	do	840 00
WARRANTS, ESTIMATES, AND APPROPRIATIONS.				
Chief.				
J. T. Power	Washington	Pennsylvania	Pennsylvania	3,000 00
Assistant Chief.				
W. F. Maclennan	Washington	Connecticut	New York	2,400 00

(b) 1875 Register

Notes: The figure shows the cover page of the 1817's Register, and the first page of the Treasury Department in the 1875's Register.

Figure A2: "Where employed" column from the 1875 Official Register

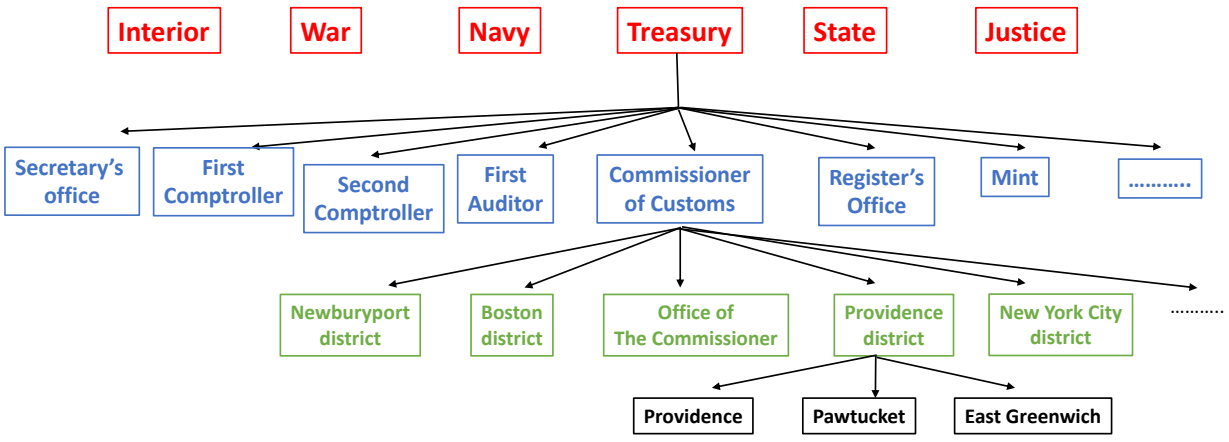


Customs Service.

Name.	Office.	Where employed.
William R. Perry ...	Deputy collector and inspector.	Superior, Wis...
Phillip H. Payne ...	....do	Torch Lake.....
Joseph Flesheimer ..	....do	Menomonee.....
Edward L. Wright ..	.. do	Houghton .....
Charles J. Linke.....	....do	White Fish Point
Orrin W. Robinson..	....do	Houghton .....
Peter C. Bird	...do	Eagle Harbor...
Albert C. Hayward..	....do	Bayfield, Wis ...
Henry C. Carleton ..	....do	Detour.....
Hiram K. Cole.....	....do	Isle Royale.....
Josiah R. Brooks	...do	Menomonee.....
Alfred Meade	....do	Ontonagon .....
John A. Freuch	....do	Marquette .....
John Q. Bernard.....	....do	L'Anse .....
Hylor A. Downs.....	Special inspector.....	Marquette .....
Ashbell Roach	Inspector	Sault Ste. Marie
William Newcomb ..	....do	...do .....

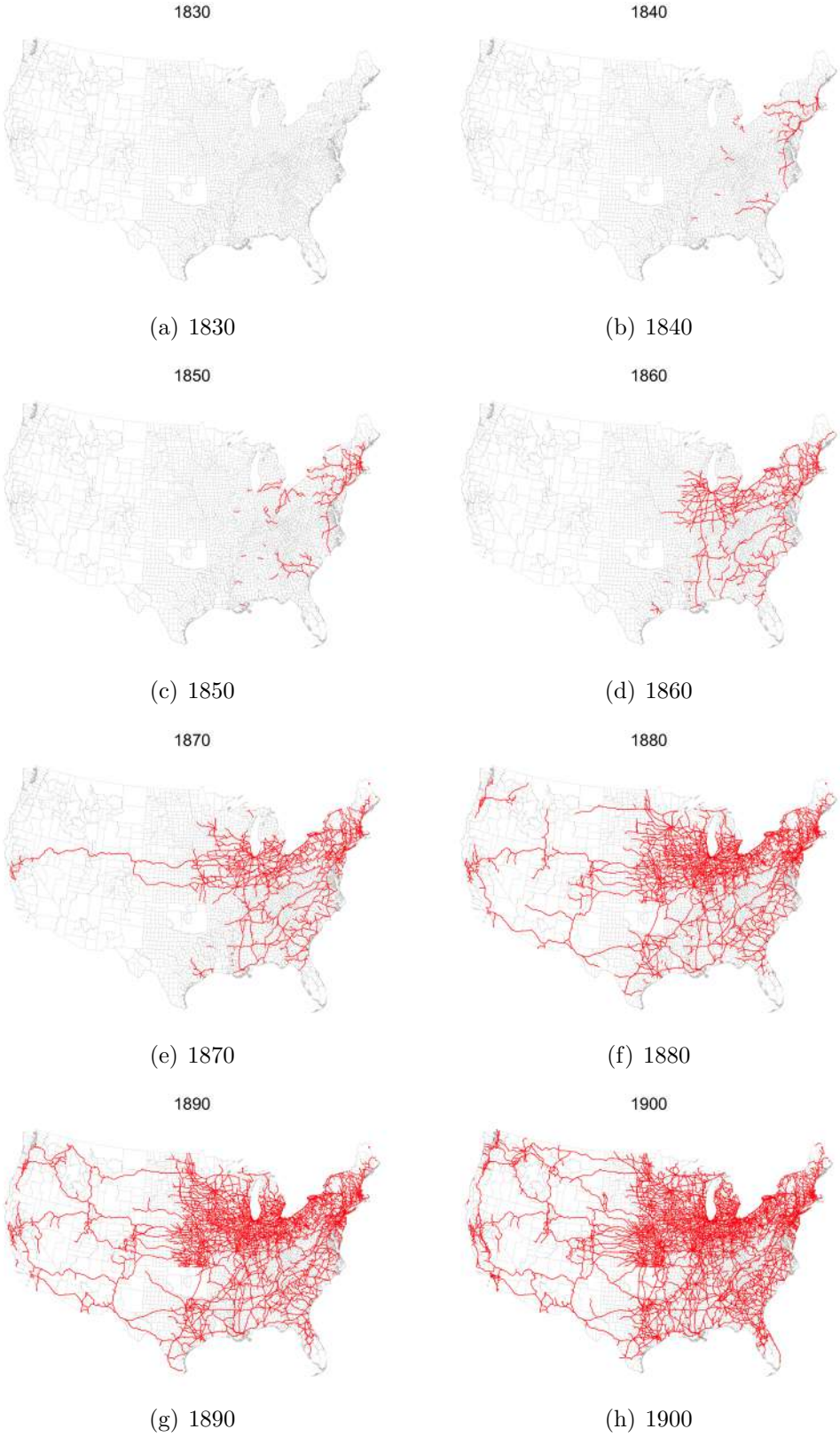
Notes: The figure shows an extract from the 1875 Register, highlighting the locations under the "where employed" column.

Figure A3: Partial Hierarchy for 1853



Notes: The figure shows a partial graphical representation of the hierarchy of the U.S. federal bureaucracy in 1853. Departments are shown in red, bureaus/offices are shown in blue, divisions are shown in green, and local offices are shown in black.

Figure A4: Expansion of the railroad network over time



(a) 1830 (b) 1840

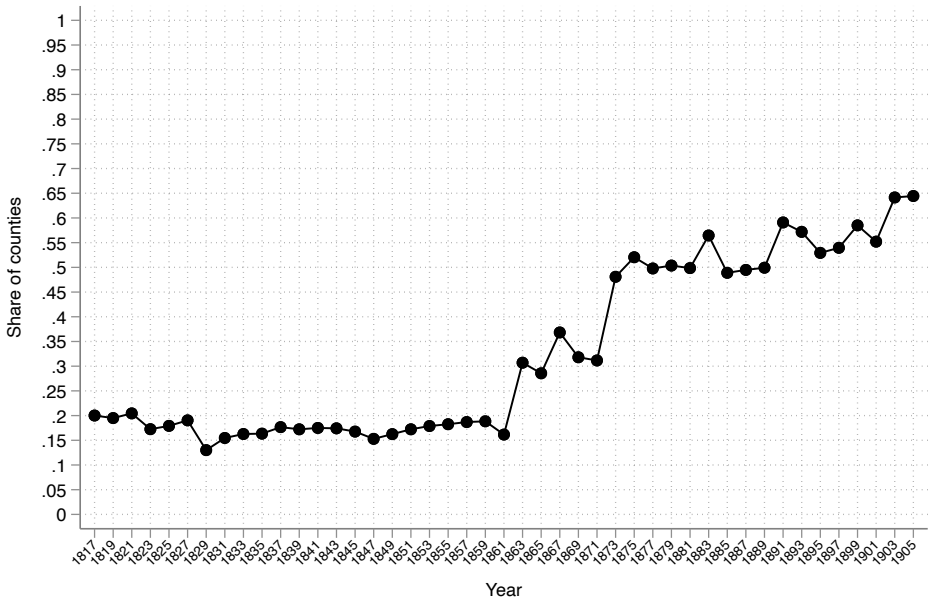
(c) 1850 (d) 1860

(e) 1870 (f) 1880

(g) 1890 (h) 1900

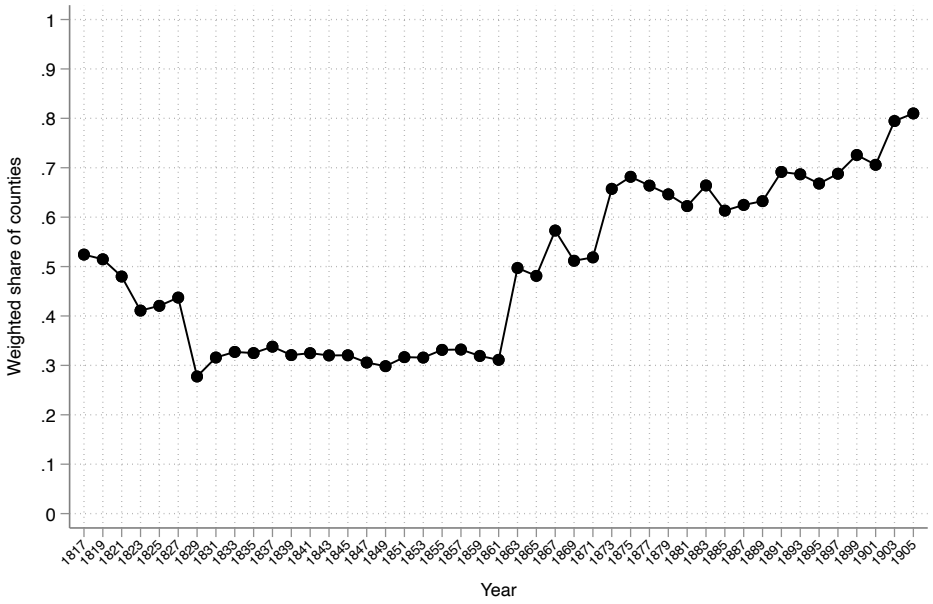
Notes: The figure shows the expansion of the railroad network over time. Source: transportation network database by Donaldson and Hornbeck (2016), based on initial GIS railroad files by Atack (2013).

Figure A5: Share of counties with state presence – sample of states as of 1817



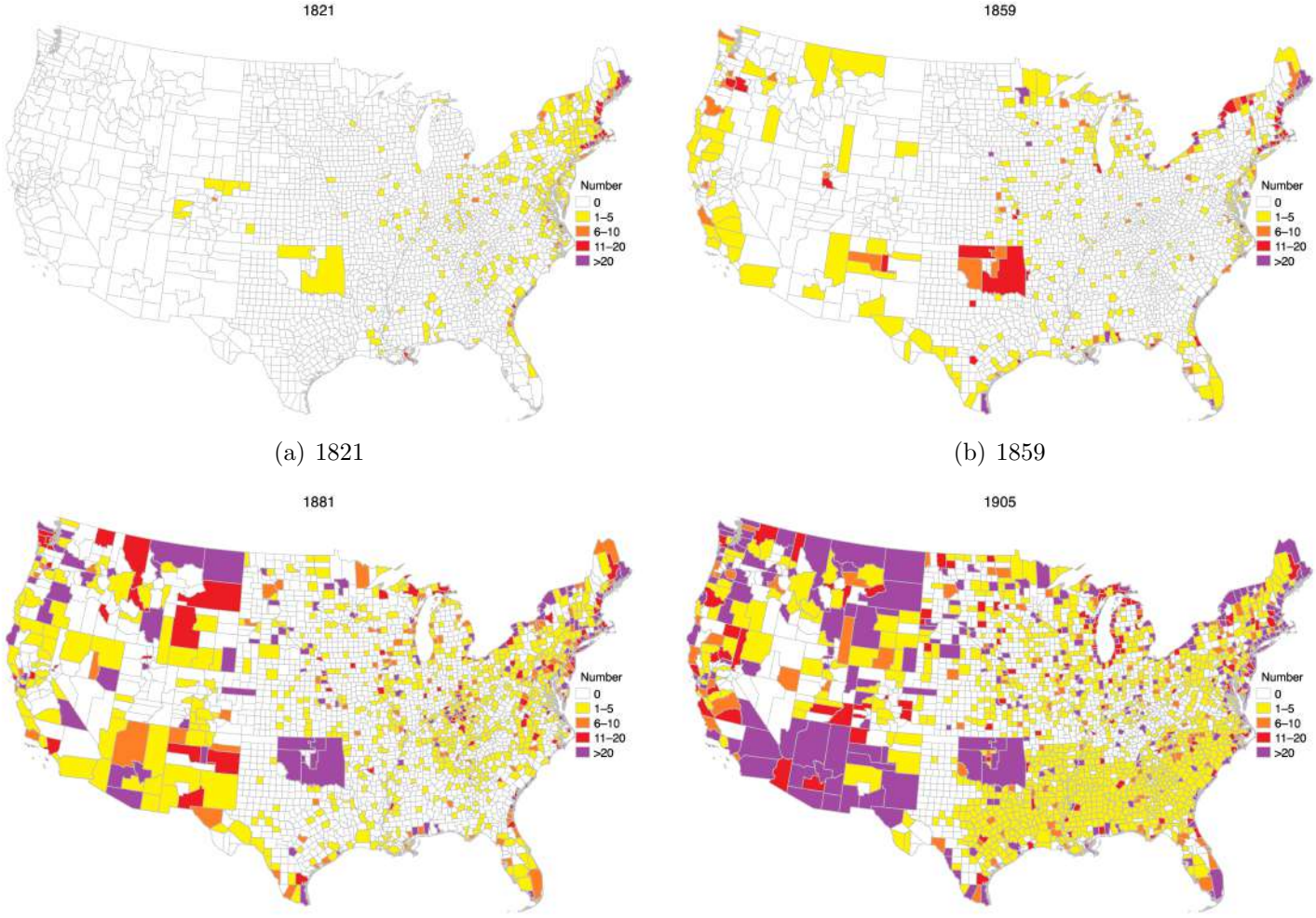
Notes: The figure shows the share of U.S. counties with federal government presence in each year from 1817 to 1905, limiting the sample to counties in states that were already part of the U.S. in 1817.

Figure A6: Share of counties with state presence – weighting by population



Notes: The figure shows the share of U.S. counties with federal government presence in each year from 1817 to 1905, weighting each county by the fraction of the U.S. population living in the county in that year.

Figure A7: The Geographic Expansion of the U.S. Federal Government

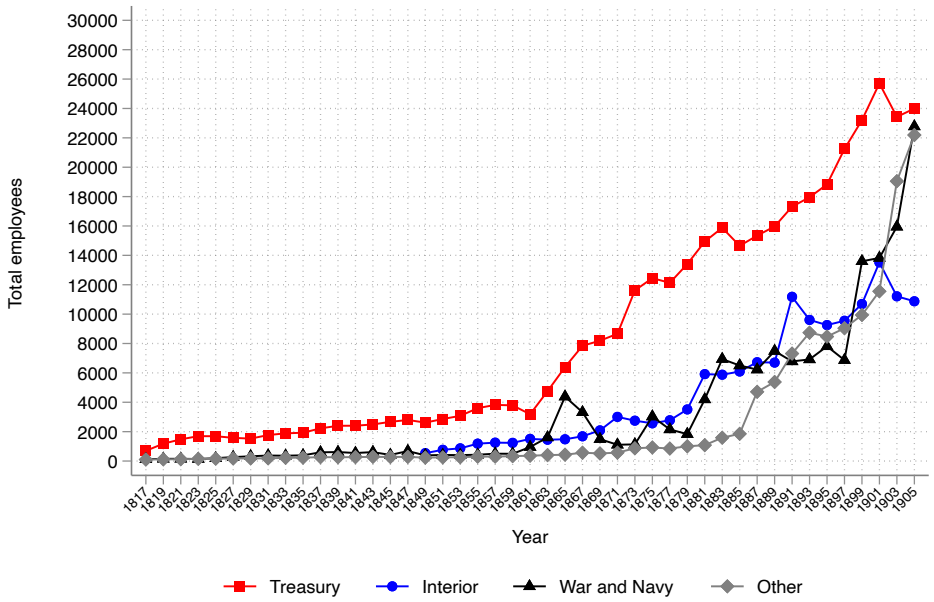


(a) 1821 (b) 1859

(c) 1881 (d) 1905

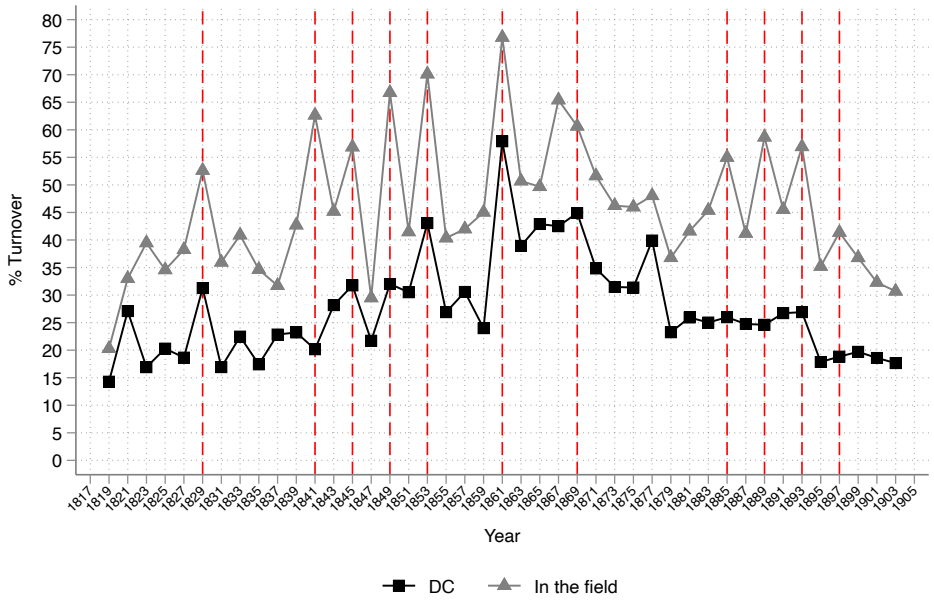
Notes: The figure shows the number of workers of the federal government employed in each U.S. county (using fixed 1890 county border), in 1821 (Panel A), 1859 (Panel B), 1881 (Panel C), 1905 (Panel D).

Figure A8: Number of workers of the federal government by department



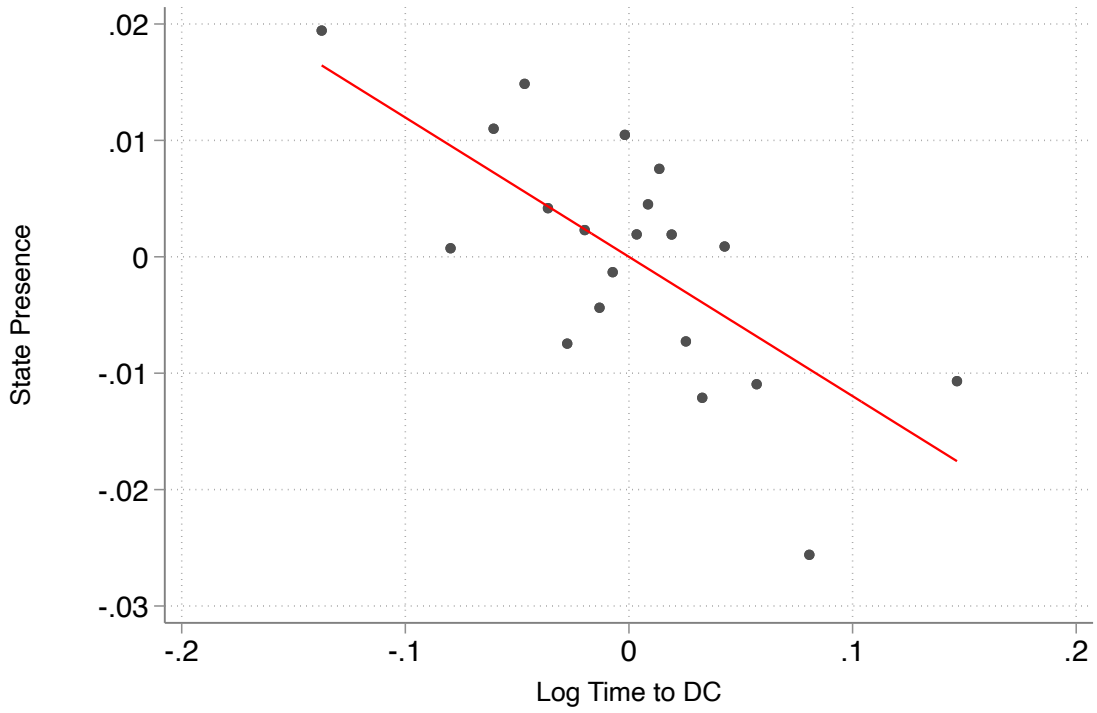
Notes: The figure shows the number of workers of the federal government by department.

Figure A9: Turnover - D.C. vs Field



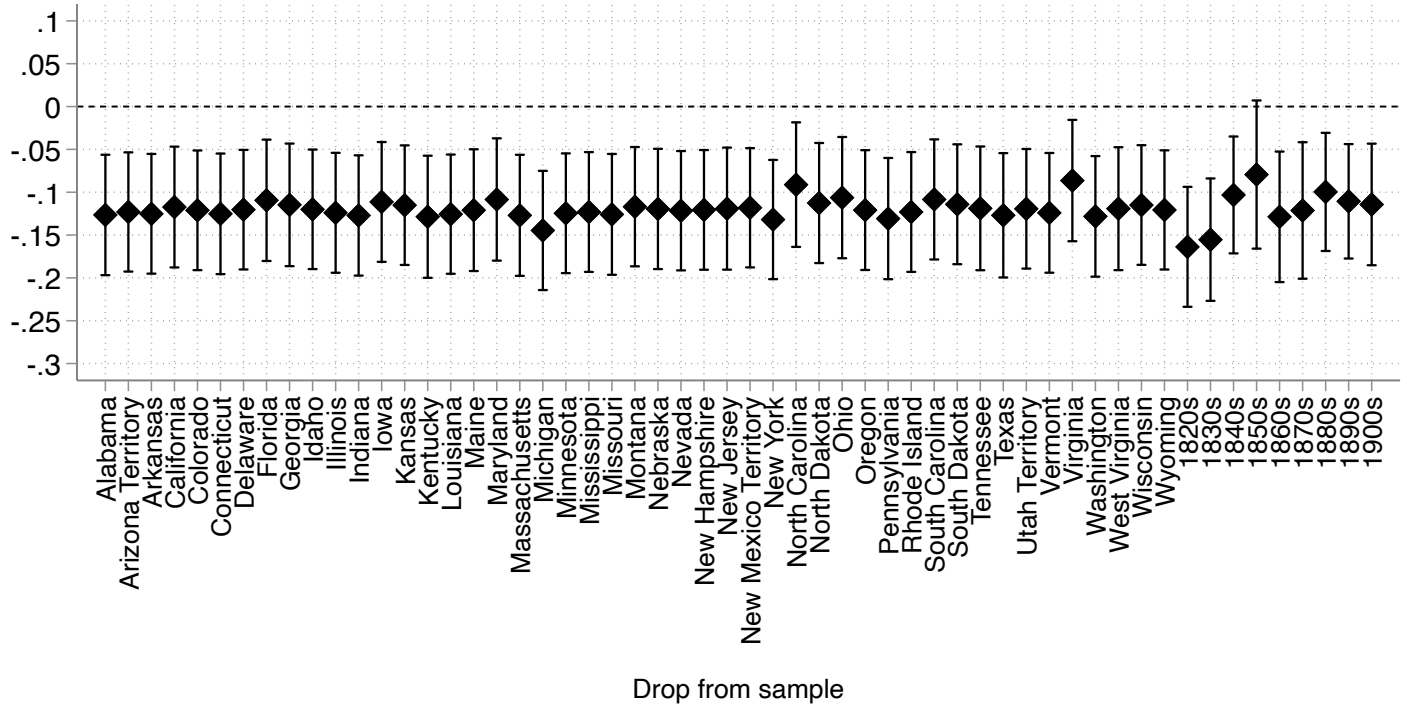
Notes: The figure shows turnover rates, i.e. the share of employees leaving the federal government in a given year, over 1817-1905, separately for employees in D.C. (black) and outside of D.C. (gray).

Figure A10: Travel time to D.C. and probability of state presence



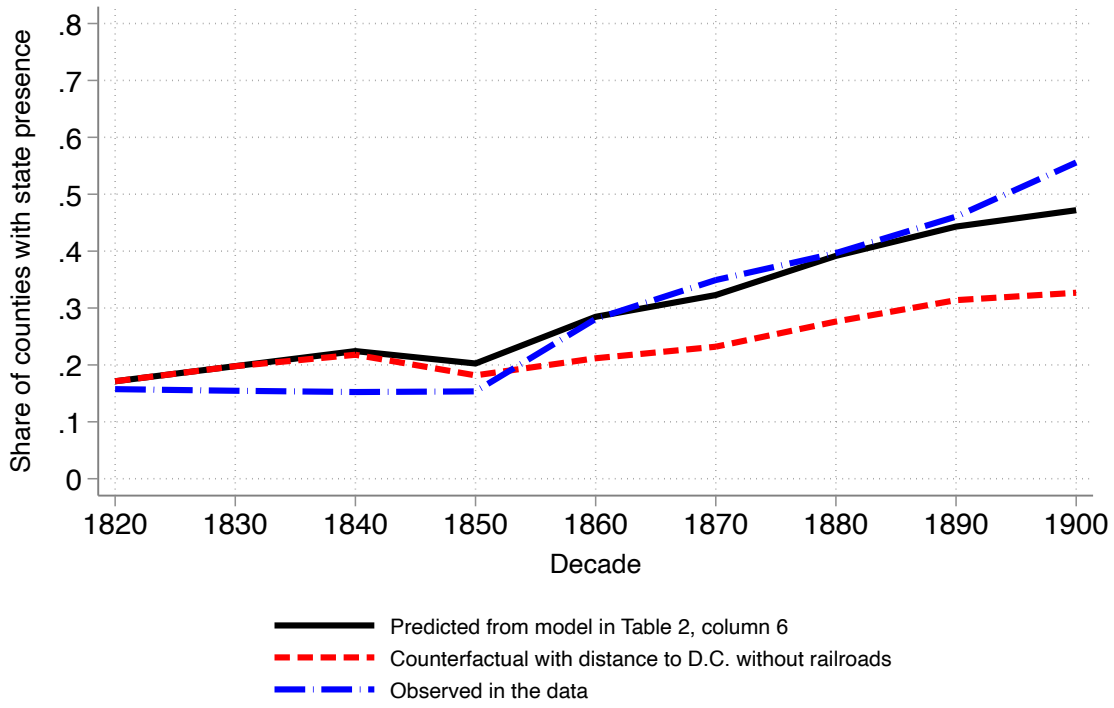
Notes: The figure shows the partial correlation between *State presence* (an indicator that takes value one if the federal government is present in the county) and *Log Time to DC* (the log of travel time in minutes between D.C. and the county's centroid), in a bin scatter plot. The relationship shown is after partialing out the full list of controls from the specification in column 6 of Table 2.

Figure A11: Travel time to D.C. and state presence. Robustness to dropping single states or single decades



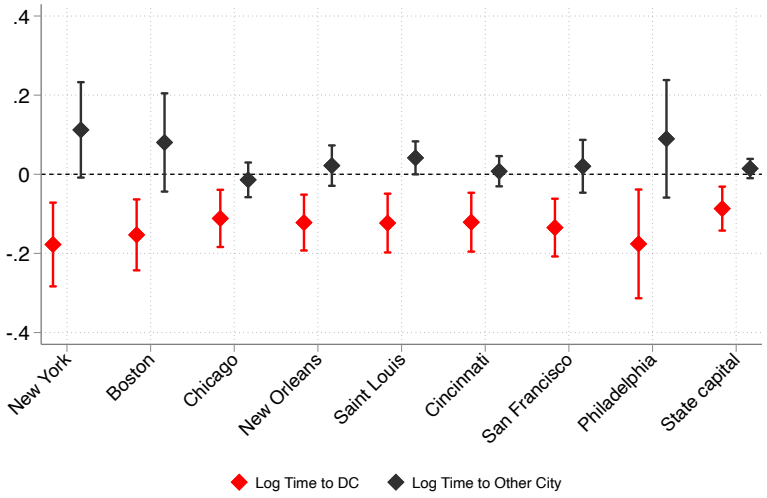
Notes: The figure shows estimated coefficient on *Log Time to* DC_{ct} and 95 percent confidence intervals based on standard errors clustered at the county level, from the estimation of the specification in column 6 of Table 2, dropping from the sample the state or the decade indicated on the x-axis.

Figure A12: Counterfactual share of counties with state presence, with travel time to D.C. fixed at 1820s values

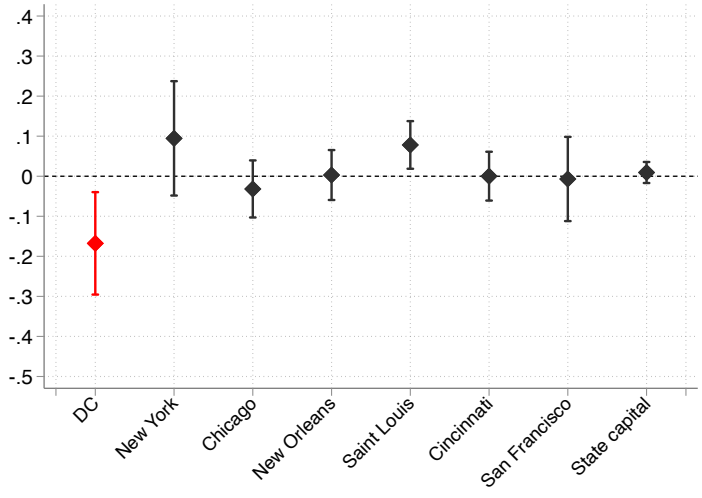


Notes: The figure shows the share of counties with state presence predicted by the estimation of equation 2 (in black), the counterfactual share of counties with state presence, had the travel times between each county and D.C. remained fixed at their values in 1820 (in red), and the actual share of counties with state presence that we observe in the data (in blue).

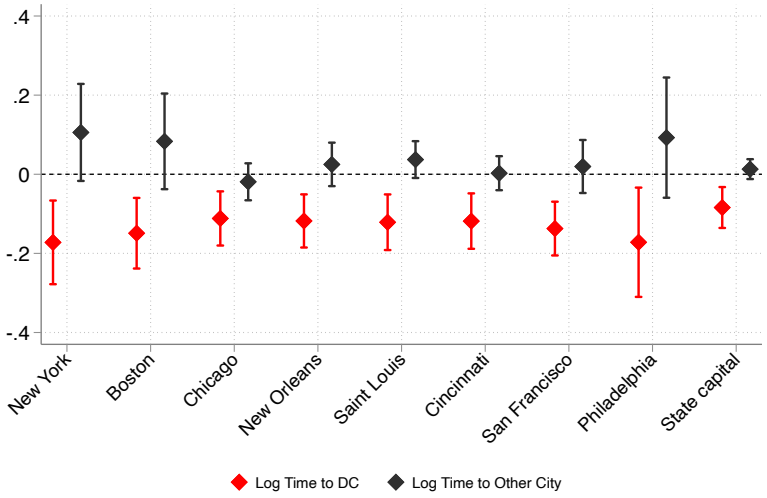
Figure A13: Lower travel time to other cities is not associated with the presence of more managers



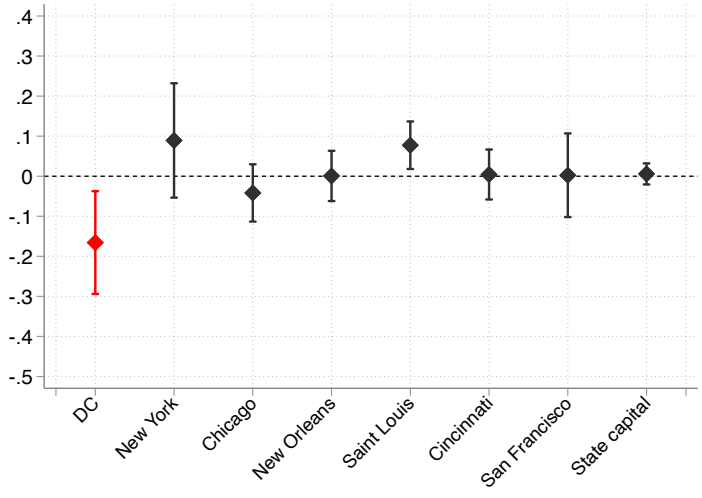
(a)



(b)



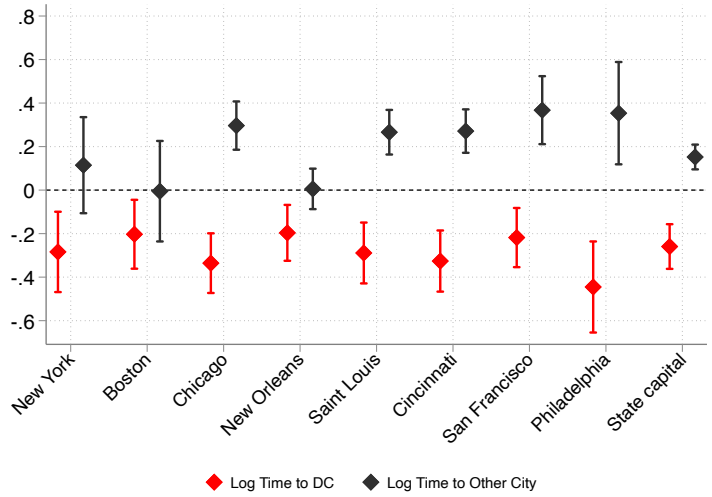
(c)



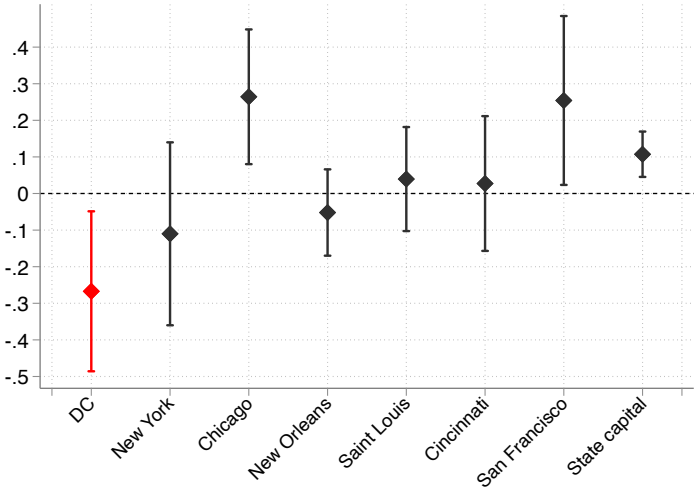
(d)

Notes: Results from estimation of the specification in column 1, panel A, Table 3, adding log of travel time between city P and the county's centroid. Panel (a) reports effects for D.C. (in red) and for each other city (in black) from a series of regressions for each city P. Panel (b) reports results from a single regression with all cities included. Panels (c) and (d) report results from similar regressions, but with $log(MA)$ calculated using travel time instead of transportation costs. 95 percent confidence intervals are based on standard errors clustered at the county-level.

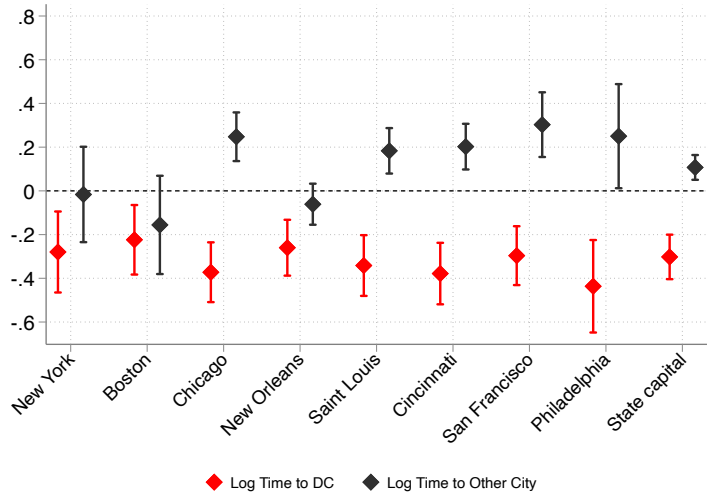
Figure A14: Lower travel time to other cities is not associated with the presence of more clerks



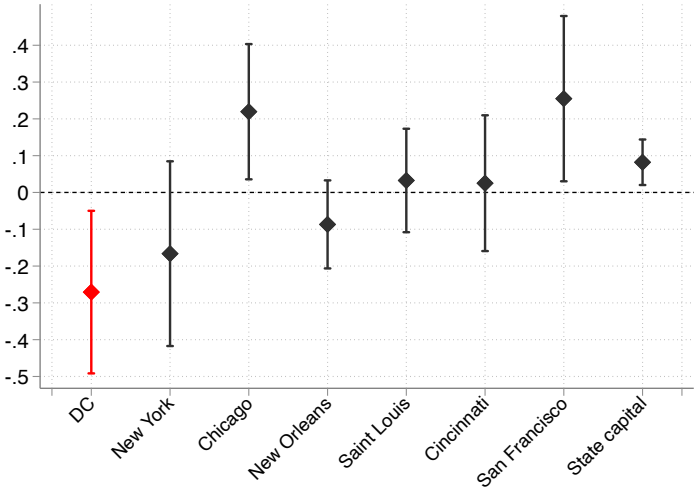
(a)



(b)



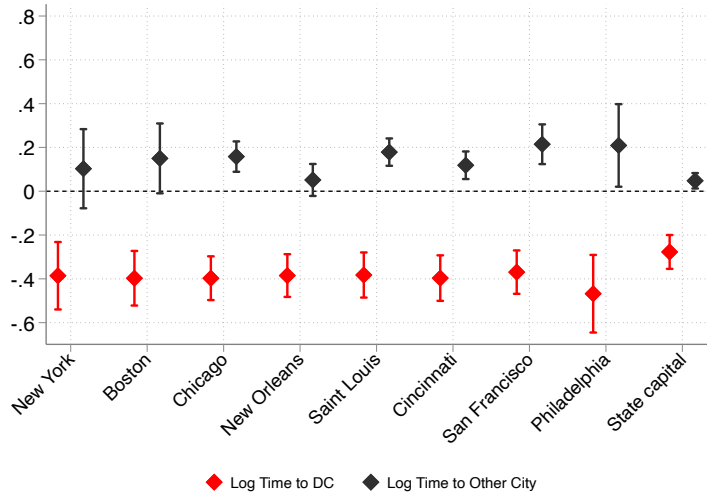
(c)



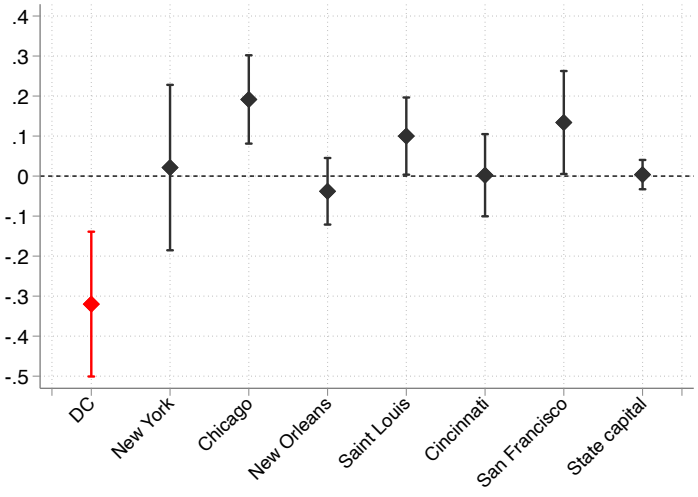
(d)

Notes: Results from estimation of the specification in column 2, panel A, Table 3, adding log of travel time between city P and the county's centroid. Panel (a) reports effects for D.C. (in red) and for each other city (in black) from a series of regressions for each city P. Panel (b) reports results from a single regression with all cities included. Panels (c) and (d) report results from similar regressions, but with $log(MA)$ calculated using travel time instead of transportation costs. 95 percent confidence intervals are based on standard errors clustered at the county-level.

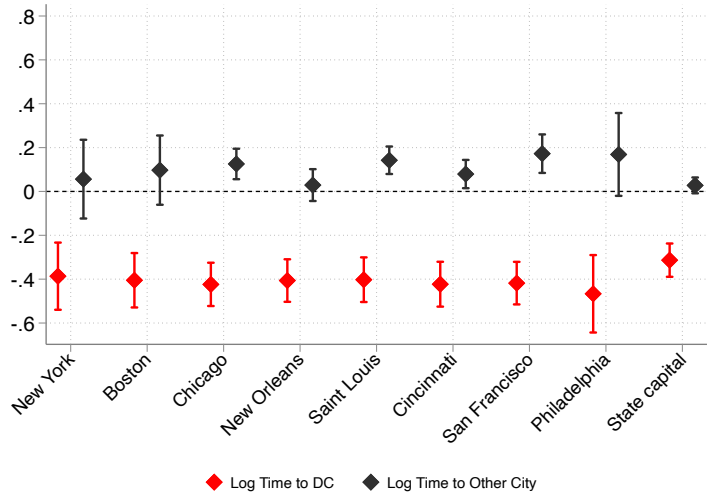
Figure A15: Lower travel time to other cities is not associated with the presence of bureaus related to taxation



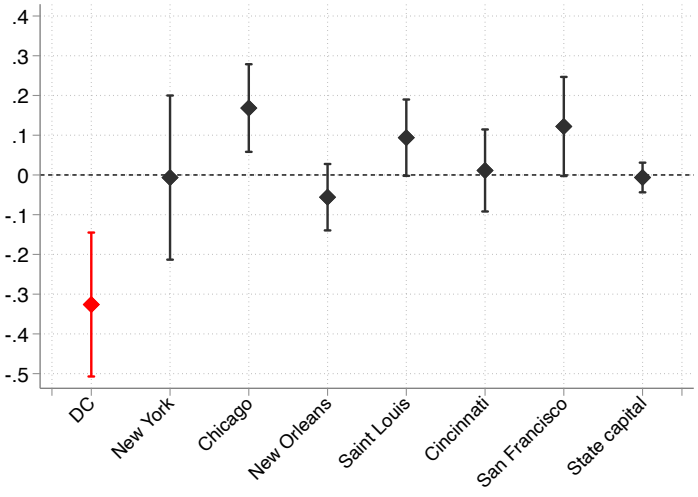
(a)



(b)



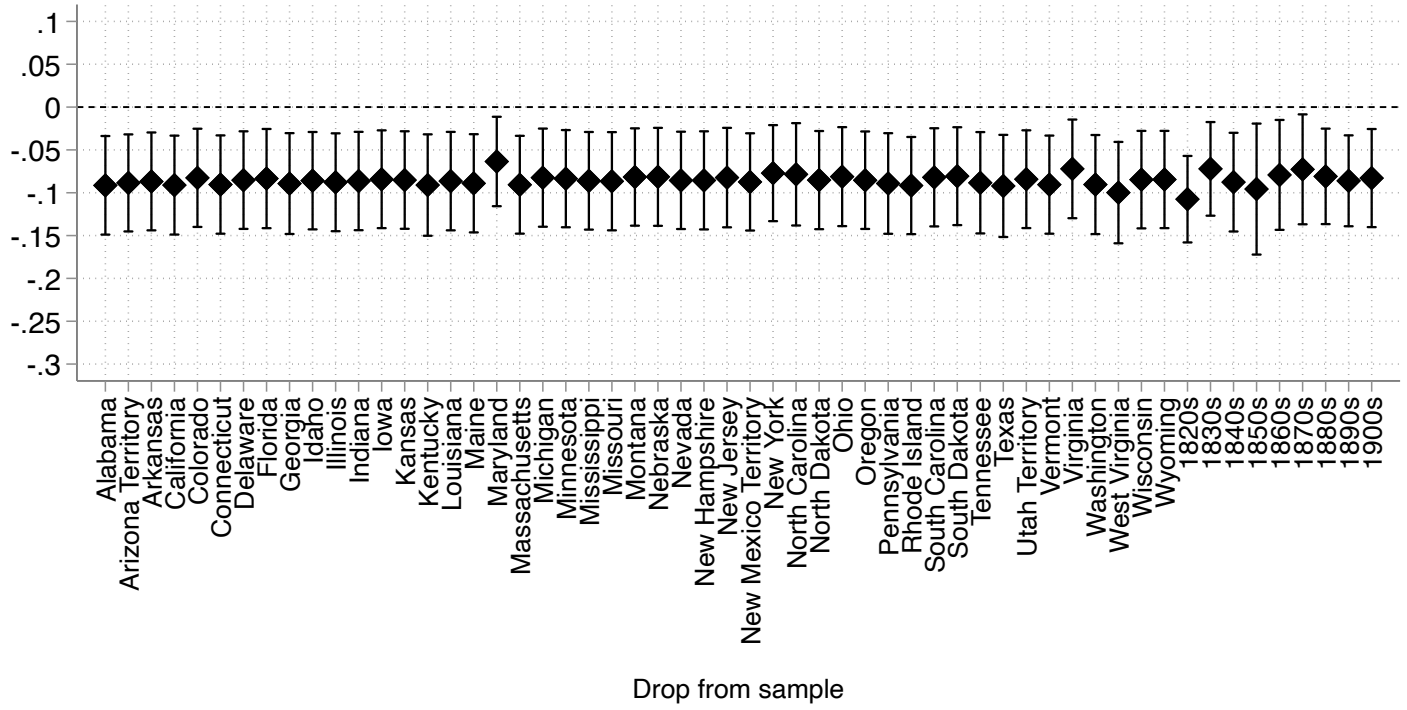
(c)



(d)

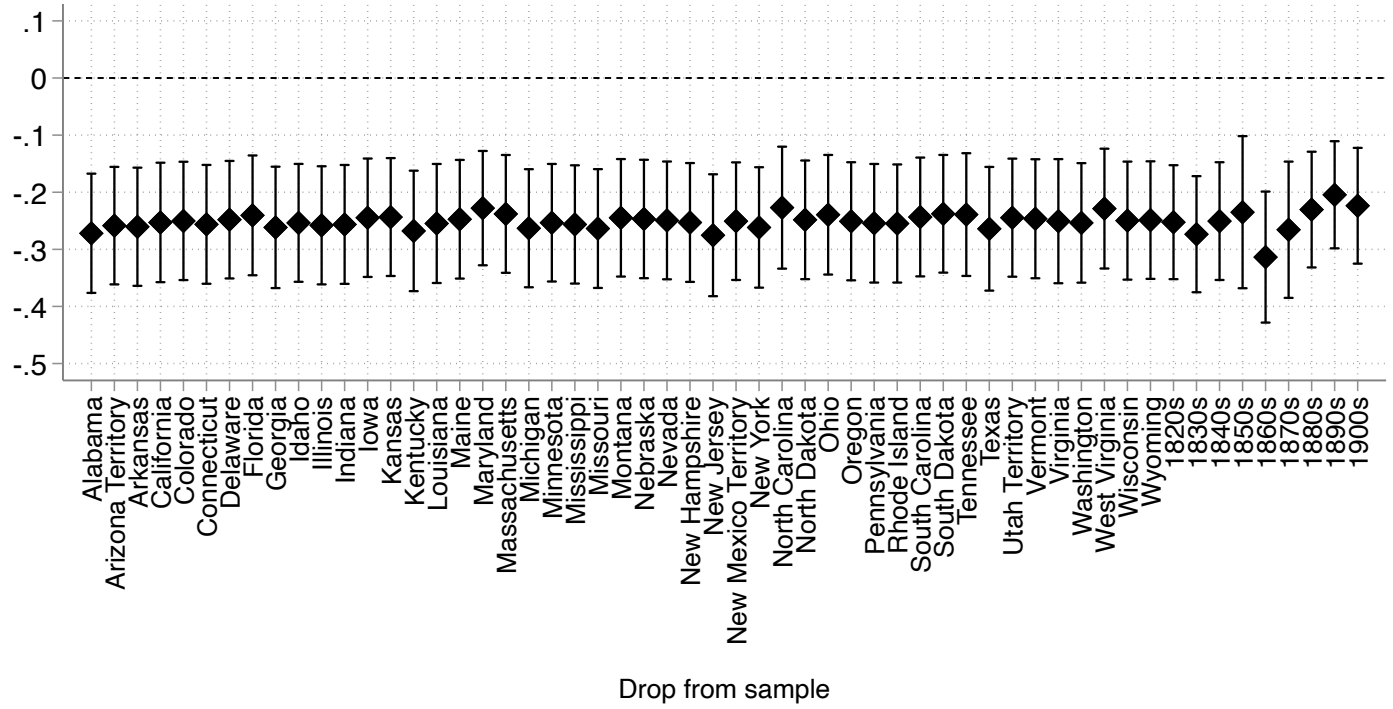
Notes: Results from estimation of the specification in column 2, panel B, Table 3, adding log of travel time between city P and the county's centroid. Panel (a) reports effects for D.C. (in red) and for each other city (in black) from a series of regressions for each city P. Panel (b) reports results from a single regression with all cities included. Panels (c) and (d) report results from similar regressions, but with $log(MA)$ calculated using travel time instead of transportation costs. 95 percent confidence intervals are based on standard errors clustered at the county-level.

Figure A16: Travel time to D.C. and number of managers. Robustness to dropping single states or single decades



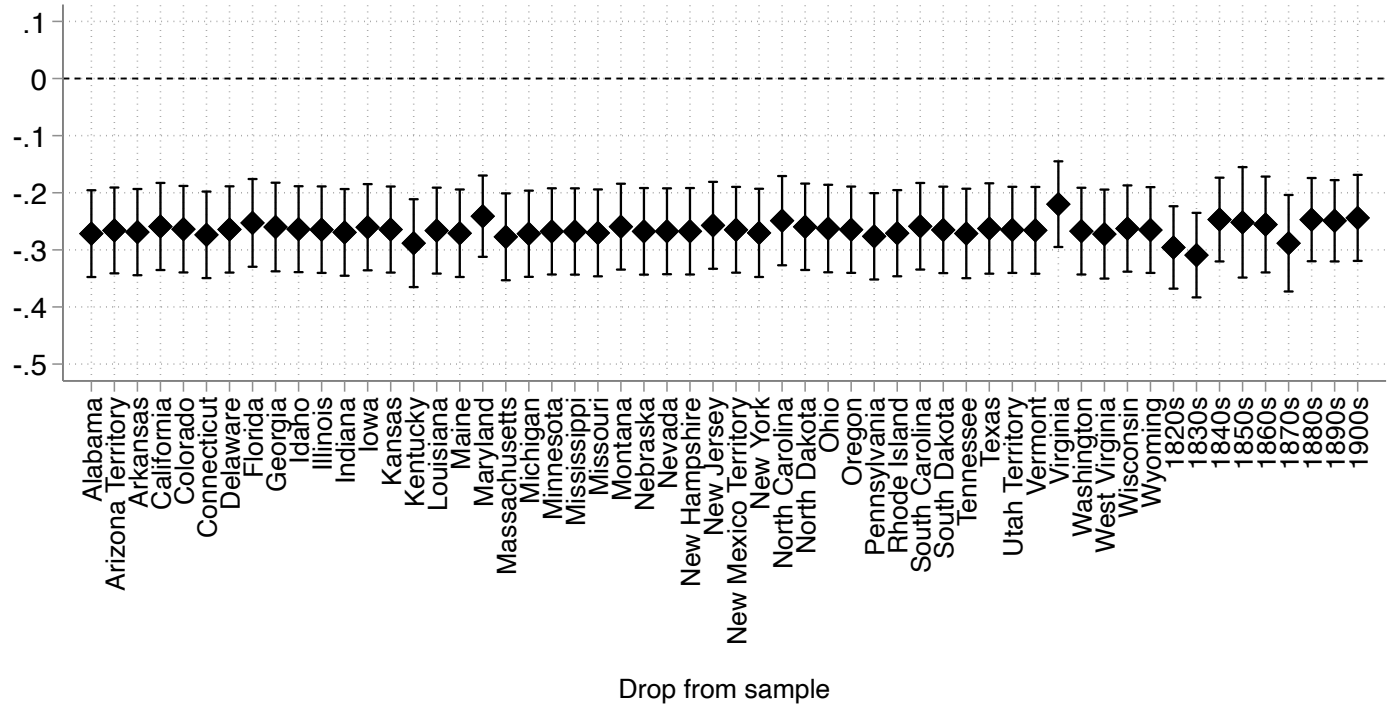
Notes: The figure shows estimated coefficient on *Log Time to* DC_{ct} and 95 percent confidence intervals based on standard errors clustered at the county level, from the estimation of the specification in column 1, panel A, of Table 3, dropping from the sample the state or the decade indicated on the x-axis.

Figure A17: Travel time to D.C. and number of clerks. Robustness to dropping single states or single decades



Notes: The figure shows estimated coefficient on *Log Time to* DC_{ct} and 95 percent confidence intervals based on standard errors clustered at the county level, from the estimation of the specification in column 2, panel A, of Table 3, dropping from the sample the state or the decade indicated on the x-axis.

Figure A18: Travel time to D.C. and state presence in taxation bureaus. Robustness to dropping single states or single decades



Notes: The figure shows estimated coefficient on *Log Time to* DC_{ct} and 95 percent confidence intervals based on standard errors clustered at the county level, from the estimation of the specification in column 2, panel B, of Table 3, dropping from the sample the state or the decade indicated on the x-axis.

Table A1: The federal government does not substitute for lower levels of government

	(1) State govt. presence	(2) Local govt. presence	(3) State govt. workers per capita	(4) Local govt. workers per capita
State presence	0.027**	0.004		
	(0.012)	(0.007)		
Federal workers per capita			-0.003	0.013*
			(0.003)	(0.007)
Observations	13,157	13,157	13,072	13,072
Mean dep. var.	0.445	0.884	0.069	0.670

Notes: The unit of observation is a county-year, with the sample limited to the years 1851, 1861, 1871, 1881, 1891, and 1901. *State presence* takes value one if the federal government is present in the county. *State govt. presence* (respectively, *Local govt. presence*) takes value one if the state government (respectively, a local government) is present in the county with at least one employee. *Federal workers per capita* is the number of federal employees per 1,000 people. *State govt. workers per capita* (respectively, *Local govt. workers per capita*) is the number of state (respectively, local) government employees per 1,000 people. In all columns, we control for county fixed effects and year fixed effects. Standard errors in parentheses, clustered by county. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: Largest bureaus outside of D.C.

Bureau	*Number of workers*
Customs	134,239
Internal revenue	49,975
Indian office	43,939
Ordnance - War	33,001
Quartermaster - War	28,253
Lighthouse board	27,644
Mint	16,254
General land office	13,322
Life saving service	11,863
Weather bureau	7,404
Other bureaus	69,190

Notes: Number of total workers employed across all years by bureau of the federal government. "Other bureaus" include all bureaus with less than 7,000 workers.

Table A3: Turnover is lower in D.C. than in the field

	(1)	(2)	(3)
	Turnover	Turnover	Turnover
Employed in DC	-0.188***	-0.199***	-0.181**
	(0.001)	(0.002)	(0.003)
Observations	657,341	657,246	652,722
Mean Turnover in Field	0.434	0.434	0.434
Fixed effects	Year	Year-Bureau	Year-Bureau-Occ. layer

Notes: The unit of observation is the employee-year. Turnover takes value one if the employee leaves the organization. *Employed in DC* takes value one if the employee is employed in D.C. Column 1 includes year fixed effects, column 2 includes year-bureau fixed effects, column 3 includes year-bureau-occupational layer fixed effects. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Correlation between travel time to D.C., state presence, and organizational features.

	(1) State Presence	(2) Log (1 plus) Managers	(3) Log (1 plus) Clerks	(4) Manager present	(5) Taxation Bureaus	(6) Turnover White Collars	(7) Shared Last name
Log Time to DC	-0.112***	-0.058***	-0.082***	-0.063***	-0.133***	0.028***	0.004**
	(0.008)	(0.010)	(0.016)	(0.009)	(0.007)	(0.004)	(0.002)
Observations	96,875	96,875	96,875	29,532	96,875	20,874	29,521
Mean dep. var.	0.307	0.094	0.263	0.317	0.192	0.615	0.040
Std. dev. Log Time to DC	0.814	0.814	0.814	0.893	0.814	0.909	0.892

Notes: The unit of observation is a county-year. The dependent variable in column 1 is an indicator that takes value one if the federal government is present in the county; in column 2 and 3 is the logarithm of (one plus) the total number of workers employed in managerial and clerical occupations, respectively; in column 4 is an indicator equal to one if there is at least one manager in the county; in column 5 is an indicator equal to one if a bureau of the federal government related to taxation is present in the county; in column 6 is the share of white collars who leave their local office between year t and year t-2; in column 7 is the share of employees who share a last name with at least one other worker in the same local office. *Log Time to DC* is the log of travel time (in minutes) between D.C. and the county's centroid. All specifications control for year fixed effects. In column 4 and 7, we additionally control for fixed effects for the number of workers in the county. In columns 4, 6, 7 we restrict the sample only to county-year observations with state presence. In column 6 we also exclude the year 1905, as we cannot measure turnover for employees in 1905, and counties with no white collar worker employed. Standard errors in parentheses, clustered at the county-level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5: Travel time to D.C. and state presence. Coefficients on other variables

	(1) Coefficient	(2) Standard error	(3) Std. dev. variable
Log Time to DC	-0.120***	(0.035)	0.816
Log Market Access	0.046***	(0.014)	0.943
Railroad	0.023**	(0.011)	0.496
Length local railroad track	0.086**	(0.037)	0.381
Log Population	0.016***	(0.003)	2.057
Share Manufacturing Employment	0.060	(0.057)	0.055
Urbanization	0.203***	(0.032)	0.163
Observations: 94,071			
Mean dep. var.: 0.312			

Notes: The unit of observation is a county-year. The dependent variable is an indicator that takes value one if the federal government is present in the county. We report estimated coefficient, standard errors (clustered at the county level), and standard deviation, for some of the controls included in the specification from column 6 of Table 2. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6: Travel time to D.C. and state presence. Robustness checks

	(1) Coefficient	(2) Standard error	(3) Observations	(4) Mean dep. var.	(5) Std. dev. Log Time to DC
1. Table 2, column 6	-0.120***	(0.035)	94,071	0.312	0.816
2. Drop if major battle	-0.125***	(0.037)	88,905	0.307	0.792
3. Drop if Sherman march	-0.112***	(0.036)	91,561	0.314	0.822
4. Drop if any battle	-0.128***	(0.038)	85,078	0.301	0.787
5. Drop South in 1860s-1870s	-0.103**	(0.043)	79,534	0.308	0.797
6. Drop states after 1860	-0.117***	(0.038)	82,917	0.315	0.815
7. Drop states after 1820	-0.085*	(0.044)	54,521	0.338	0.780
8. Log MA with travel time costs	-0.171***	(0.035)	94,071	0.312	0.816
9. Travel time parameters fixed 1860	-0.134***	(0.031)	86,306	0.326	0.820
10. Travel time parameters fixed 1900	-0.172***	(0.027)	86,306	0.326	0.820
11. Only middle year of each decade	-0.131***	(0.044)	19,549	0.336	0.821
12. Previous decade	-0.015	(0.045)	16,354	0.307	0.765

Notes: The unit of observation is a county-year. The dependent variable is an indicator that takes value one if the federal government is present in the county. We report estimated coefficients and standard errors (clustered at the county level) for the impact of log of travel time between D.C. and the county's centroid. All specifications control for county fixed effects, year fixed effects, a second-order polynomial in the distance between the county and D.C. interacted with year fixed effects, and the full set of controls from column 6 of Table 2. Row 1 reports the baseline estimate from Table 2, column 6. In row 2-4, we drop from the sample all counties with Civil War battles with over 500 casualties, all counties affected by General Sherman's march, and all counties with any Civil War battle, respectively. In row 5, we drop from the sample all counties in a Confederate state in the 1860s and 1870s. In row 6, we drop all counties in states that were not already part of the U.S. in 1860. In row 7, we drop all counties in states that were not already part of the U.S. in 1820. In row 8, we use an alternative control for market access that relies on county-to-county travel times in minutes, rather than on costs of transportation per ton-mile. In rows 9 and 10, we use a measure of $Log\ Time\ to\ DC_{ct}$ computed using time-invariant travel time parameters reflecting travel times in 1860 and 1900, respectively (dropping from the sample the 1820s decade, as there is no variation in travel time to D.C. relative to the 1820s). In row 11, we keep only the middle year for each decade. In row 12, the dependent variable is the lagged value of state presence. Specifically, for each decade, we use state presence from the final year of the preceding decade (e.g., for 1830, we use state presence in 1819; for 1840, we use 1829). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: Travel time to D.C. and additional measures of state presence. Robustness checks

	(1) Coefficient	(2) Standard error	(3) Observations	(4) Mean dep. var.	(5) Std. dev. Log Time to DC
Panel A: Log (1 plus) number of workers employed as managers					
1. Table 2, column 6	-0.085***	(0.029)	94,071	0.096	0.816
2. Drop if major battle	-0.089***	(0.030)	88,905	0.092	0.792
3. Drop if Sherman march	-0.085***	(0.029)	91,561	0.097	0.822
4. Drop if any battle	-0.080***	(0.030)	85,078	0.088	0.787
5. Drop South in 1860s-1870s	-0.081**	(0.035)	79,534	0.092	0.797
6. Drop states after 1860	-0.081**	(0.032)	82,917	0.096	0.815
7. Drop states after 1820	-0.060	(0.037)	54,521	0.115	0.780
8. Log MA with travel time costs	-0.087***	(0.027)	94,071	0.096	0.816
9. Travel time parameters fixed 1860	-0.082***	(0.023)	86,306	0.096	0.820
10. Travel time parameters fixed 1900	-0.073***	(0.019)	86,306	0.096	0.820
11. Only middle year of each decade	-0.088***	(0.034)	19,549	0.099	0.821
12. Previous decade	0.027	(0.029)	16,498	0.088	0.766
Panel B: Log (1 plus) number of workers employed as clerks					
1. Table 2, column 6	-0.251***	(0.052)	94,071	0.266	0.816
2. Drop if major battle	-0.285***	(0.054)	88,905	0.263	0.792
3. Drop if Sherman march	-0.250***	(0.053)	91,561	0.269	0.822
4. Drop if any battle	-0.275***	(0.054)	85,078	0.258	0.787
5. Drop South in 1860s-1870s	-0.341***	(0.063)	79,534	0.267	0.797
6. Drop states after 1860	-0.229***	(0.056)	82,917	0.258	0.815
7. Drop states after 1820	-0.163**	(0.066)	54,521	0.281	0.780
8. Log MA with travel time costs	-0.311***	(0.051)	94,071	0.266	0.816
9. Travel time parameters fixed 1860	-0.176***	(0.046)	86,306	0.284	0.820
10. Travel time parameters fixed 1900	-0.159***	(0.038)	86,306	0.284	0.820
11. Only middle year of each decade	-0.366***	(0.061)	19,549	0.285	0.821
12. Previous decade	-0.192***	(0.059)	16,498	0.205	0.766
Panel C: State presence in taxation bureaus					
1. Table 2, column 6	-0.265***	(0.038)	94,071	0.196	0.816
2. Drop if major battle	-0.277***	(0.040)	88,905	0.192	0.792
3. Drop if Sherman march	-0.260***	(0.039)	91,561	0.197	0.822
4. Drop if any battle	-0.278***	(0.041)	85,078	0.189	0.787
5. Drop South in 1860s-1870s	-0.264***	(0.046)	79,534	0.191	0.797
6. Drop states after 1860	-0.261***	(0.041)	82,917	0.210	0.815
7. Drop states after 1820	-0.187***	(0.050)	54,521	0.244	0.780
8. Log MA with travel time costs	-0.311***	(0.038)	94,071	0.196	0.816
9. Travel time parameters fixed 1860	-0.283***	(0.031)	86,306	0.207	0.820
10. Travel time parameters fixed 1900	-0.278***	(0.026)	86,306	0.207	0.820
11. Only middle year of each decade	-0.315***	(0.045)	19,549	0.203	0.821
12. Previous decade	-0.010	(0.042)	16,498	0.176	0.766

Notes: The unit of observation is a county-year. The dependent variable is the logarithm of (one plus) the total number of workers employed in managerial occupations (Panel A), the logarithm of (one plus) the total number of workers employed in clerical occupations (Panel B), an indicator equal to one if a bureau of the federal government related to taxation is present in the county (Panel C). See notes to Appendix Table A6. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A8: Monitoring capacity and Local Economic Growth

	(1) Number of establishments	(2) Log revenue per establishment	(3) Log workers per establishment	(4) Urbanization
Log Time to DC	-0.703	-0.986***	-0.734***	-0.044***
	(14.714)	(0.183)	(0.198)	(0.015)
Observations	12,482	11,467	11,456	19,853
Std. dev. DV	171.979	1.022	1.098	0.167
Std. dev. Log Time to DC	0.780	0.780	0.781	0.819

Notes: The unit of observation is a county-year. The sample in columns 1-3 include counties in 1860, 1870, 1880, 1890, and 1900. The sample in column 4 includes counties for all decades between 1820 and 1900. All specifications include the full set of controls from the specification in column 6 of Table 2, excluding controls for local economic growth. Standard errors in parentheses, clustered at the county-level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Decomposing the Sources of Growth

In this section, we provide a decomposition of the drivers of growth of the U.S. federal government between 1817 and 1905.

There are three possible sources of growth in a state organization. First, a state can grow because it starts to perform a higher number of functions (the "functions" component of state growth). Second, a state can grow because it increases the number of locations across the territory in which it is present (the "geographic expansion" component). Third, a state can grow by increasing the intensity of its presence, i.e. by increasing the number of employees performing a given function in a given location (the "intensity" component).

Appendix Figure B19, Panel A, shows the growth in the number of bureaus of the U.S. federal government over the nineteenth century, which we consider as a proxy for a specific function performed by the state. Their number steadily increased in the first half of the century, from 25 in 1817 to 46 in 1859. The rate of growth was higher in the second half of the century, when the organization added an average of 3.7 new functions every two years, reaching a total of 132 separate functions in 1905.

In contrast, as shown in Panel C of Figure 1 (and also in Appendix Figure B19, Panel B), the state did not start to expand its geographical presence until the 1860s. We plot the share of U.S. counties where we observe a presence of the federal government (i.e. with at least one individual employed within the county borders). For each year, the number of counties with potential state presence (i.e., the denominator of this share) is the number of counties in states and territories that were included in the most recent census. In this way, we account for the enormous territorial expansion of the U.S. over the nineteenth century. This share hovered around 15 percent between 1817 and 1859, and does not display any increasing trend over this period. In the second half of the nineteenth century, the state begins to increase its presence across the territory: it is present in 24% of counties by 1871, in 38% of counties by 1881, and in 61% of counties by 1905.

In Appendix Figure B19, Panel C, we show how the average number of employees for each county-office pair, i.e. our measure of the intensity of state presence, changed over time. We observe a steady growth in this measure during the sample period, from 1.9 average employees in 1817 to 6.7 in 1859 and to 14.5 in 1905.

To provide a formal decomposition of state growth between these three sources, we compute counterfactual growth rates between 1817 and 1859, and between 1859 to 1905, had each of the three components remained constant at its level at the beginning of the period.

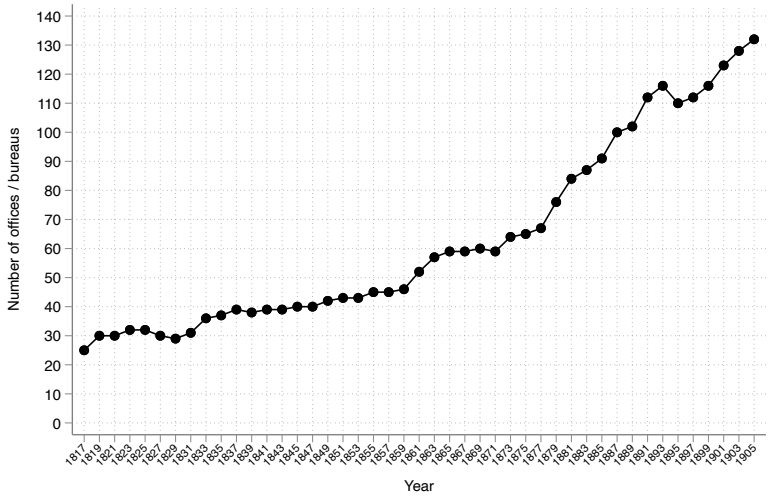
Specifically, we define the total number of workers employed by the state in year t as:

$$\text{Workers}_t = B_t \times \frac{1}{B_t} \sum_b L_{bt} \times \frac{1}{\sum_b L_{bt}} \sum_{blt} W_{blt} \tag{3}$$

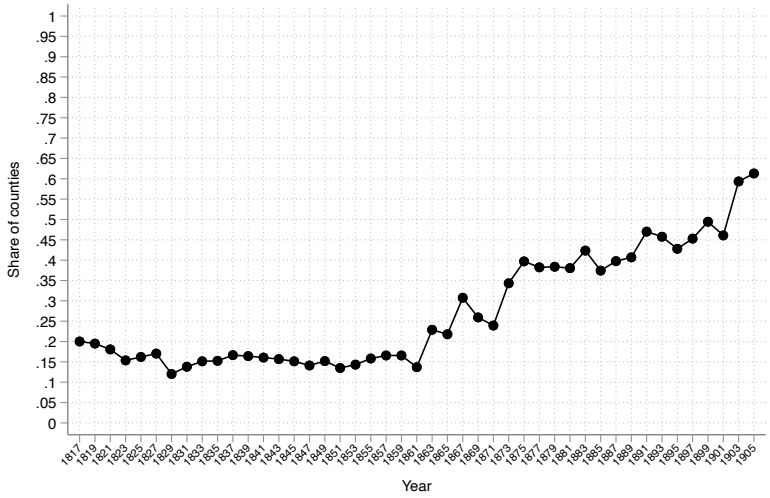
where B_t is the number of bureaus in year t, L_{bt} is the number of counties where bureau b is present in t, and W_{blt} is the number of workers employed in bureau-county bl in t. The three terms captures the function, geographic expansion, and intensity components, respectively. We compute each of the three terms for 1817, 1859, and 1905, their change from 1817 to 1859 and from 1859 to 1905, and counterfactual growths in Workers_t had each of the three components remained constant at its level at the beginning of the period.

As shown in Appendix Figure B19, Panel D, the growth of the U.S. federal government between 1817 and 1859 was entirely driven by higher number of functions and by an increase of employees per office. Consistent with the trends in Panel B of the figure, the geographic expansion component did not lead to any state growth in the 1817-1859 period. In contrast, after 1859, the geographic expansion component accounted for about 29 percent of the growth of the state, with the intensity component accounting for 32 percent and the functions component for the remaining 39 percent.

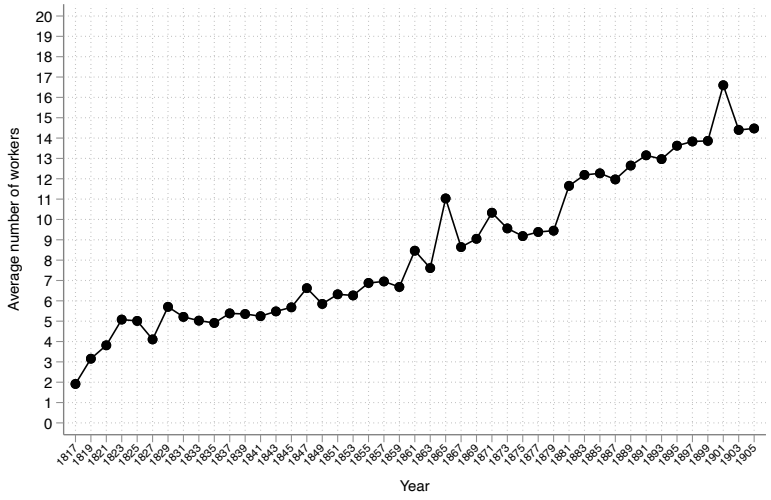
Figure B19: Decomposing the Sources of Growth



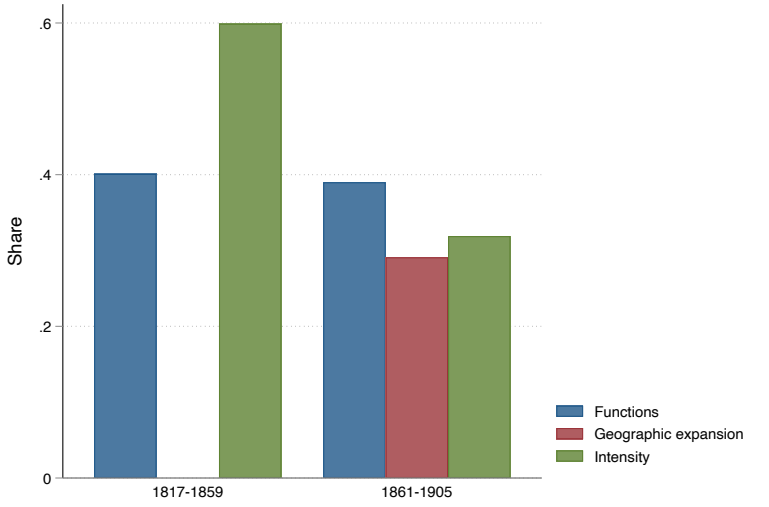
(a) Number of bureaus



(b) Share of counties with state presence



(c) Average workers across county-bureaus



(d) Decomposition

Notes: The figure shows the number of bureaus over time (Panel A), the share of counties with state presence over time (Panel B), the average number of workers across county-bureaus over time (Panel C), the share of each component's contribution to state growth between 1817-1859 and 1861-1905, following equation 3 (Panel D).

C Data construction appendix

In this section, we provide details on the additional variables used in the analysis.

Logarithm of county population. Source: U.S. censuses, 1820-1900 (Haines, 2002). Area-based weights were applied to reflect county borders as of 1890. We use the logarithm of one plus the county population, to account for the handful of counties with zero population in a given census year.

Share of county population employed in manufacturing. Source: U.S. censuses, 1820-1900 (Haines, 2002). Area-based weights were applied to reflect county borders as of 1890. Since information on the number of persons employed in manufacturing is not present in the 1830 census, we interpolate the values for this decade.

Urbanization. Definition: the share of the population living in places with at least 2,500 inhabitants. Source: U.S. censuses, 1820-1900 (Haines, 2002). Area-based weights were applied to reflect county borders as of 1890.

Immigration. Definition: the share of the population that is foreign born. Source: U.S. censuses, 1820-1900 (Haines, 2002). Area-based weights were applied to reflect county borders as of 1890. Since information on the number of persons employed in manufacturing is not present in the 1840 census, we interpolate the values for this decade.

Civil war battles. Source: Hornbeck and Rotemberg (2024), which consider all battlefields identified as significant by the Civil War Sites Advisory Commission's Report on the Nation's Civil War Battlefields of 1993. We build the following variables: the number of casualties; the number of battles; an indicator equal to one if a major battle (defined as one with more than 500 recorded casualties) took place in the county.

General Sherman march. Definition: indicator equal to one if the county is inside a ten mile geographic band of the main route of General Sherman's March. Source: Feigenbaum et al. (2022).

Average agricultural suitability. Definition (following Bazzi et al. (2020)): average of attainable yields for alfalfa, barley, buckwheat, cane sugar, carrot, cabbage, cotton, ax, maize, oats, onion, pasture grasses, pasture legumes, potato, pulses, rice, rye, sorghum, sweet potato, tobacco, tomato, and wheat. Each product's value is normalized by dividing it by the maximum value for that product in the sample. Measures of attainable yields originally constructed using data from the FAO's Global Agro-Ecological Zones project v3.0 (IIASA/FAO, 2012) for intermediate levels of inputs/technology and rain-fed conditions. Source of data: Fiszbein (2022).

Rainfall. Definition: county-level average annual precipitation. Source of data: Fiszbein

(2022), using data from FAO's Global Agro-Ecological Zones project v3.0 (IIASA/FAO, 2012).

Mineral deposits. Definition: logarithm of one plus the number of mineral deposits in the county. Source: shapefile on mineral deposits downloaded from the USGS (United States Geological Survey) website.[73]

Waterpower potential. Definition: logarithm of one plus waterpower potential of the county. For the construction of a county waterpower potential, we follow Hornbeck et al. (2024). Specifically, we compute, for any river segment in a county, its theoretical potential for generating waterpower; this is given by the product of: (1) the flow rate of water, defined as the average flow rate over the three lowest months of the year; (2) fall height, defined as the difference between the maximum and minimum elevation along the river segment; and (3) a gravitational constant equal to 0.1134. We drop all river segments that are not classified as either stream river or artificial path. We sum across all river segments in a county, and for segments intersecting county boundaries we allocate waterpower potential in proportion to the share of its length inside each county. Finally, we divide by the area of the county. Source: data on the water network is from the National Hydropgraphy Dataset Plus (NHDPlusV2, https://www.epa.gov/waterdata/nhdplus-national-data), compiled by the EPA and the U.S. Geological Survey (USGS).

[73] https://www.usgs.gov/data/prospect-and-mine-related-features-us-geological-survey-75-and-15-minute

D Travel time parameters

In order to calculate the shortest travel time between D.C. and the centroid of each county, we need to specify travel time parameters for railroads, navigable waterways, and stagecoaches. Our assumptions are motivated by historical sources documenting nineteenth-century transportation conditions.

Railroads. For railroad travel, we follow the estimates provided by White (1979).[74] In 1840, we assign a speed of 15 mph, based on White's observation that "the average on all United States roads was about 15 miles per hour" (p. 73). For 1850, we increase the speed to 20 mph, reflecting the improvement in average train speeds - although New York passenger trains averaged 24 mph by the mid-1850s, White notes that this "should be viewed as higher than the probable national average" (p 73). For 1860, we assign a speed of 22 mph, based on a conservative average of the three main train types recorded in 1864: express trains (32 mph), regular passenger trains (26 mph), and branch line trains (17 mph) (p. 74). For 1887, White (1979) reports that "the new York Central's fastest train averaged just over 40 miles per hour; [...] The Pennsylvania Railroad express between Jersey City and Washington averaged 45 miles per hour, while its fast train between Jersey City and Pittsburgh averaged 39.2 miles per hour" (page 74). Based on this data, we can conclude that the average speed for express trains increased from about 32 mph in 1860 to about 40 mph in 1890. Thus, we increase the speed by 8 mph over this period, assigning a speed of 30 mph in 1890. Interpolating linearly, we assign speeds of 25 mph in 1870, 27 mph in 1880, and 33 mph in 1900. For 1830, we reduce the assigned speed to 13 mph (relative to 15 mph in 1840) to reflect the earlier and less developed state of railroad technology.

Navigable waterways. For navigable waterways, we rely on information from Hunter (2012) until the 1850s.[75] For 1820 and 1830, we assign a speed of 4 mph, consistent with Hunter's estimate that by 1825 "many steamboats were making the fast time of one hundred miles a day" (p. 22). For 1840, we increase the speed to 5 mph, based on the 1836 record journey on the Mississippi river from St. Louis to Galena at just over 5 mph (p. 25). In 1850, we assign 9 mph, reflecting multiple sources: faster steamboats by the mid-1850s reportedly averaged 10 mph from New Orleans to Louisville (p. 23), while a Missouri River route between St. Louis and St. Joseph averaged just over 8 mph in 1856 (p. 24). From 1860 onward, steamboat race records suggest continued modest improvements. For instance, the Robert E. Lee steamboat traveled from New Orleans to St. Louis (1,278 miles) in 1870 in

[74] White John H. 1979. "A History of the American Locomotive: Its Development 1830-1880." New York: Dover Publications.

[75] Hunter, L. C. (2012). "Steamboats on the Western Rivers: An Economic and Technological History." United States: Dover Publications.

just under 4 days (approximately 14 mph), and from New Orleans to Natchez (268 miles) in 1880 in 17 hours and 11 minutes (about 15 mph).[76] Based on this evidence, we assign speeds of 10 mph in 1860, 12 mph in 1870, 13 mph in 1880, 15 mph in 1890, and 17 mph in 1900. Besides reflecting modest increases in speed over time, this parameterization maintains an approximate 2:1 ratio between rail and water speeds after 1860.

Stagecoaches: For stagecoach travel, we assign a constant speed of 7.5 mph across all decades. This conservative estimate is based on multiple sources reporting a range of 5 to 10 mph. For instance, Bancroft (1888) notes that "over a good road, ten miles an hour were readily made" in California (p. 328).[77]. In Iowa, the main stage lines averaged 8 mph, with some reaching 9 mph.[78] The Butterfield Overland Mail route averaged approximately 5 mph across its eight divisions.[79]

[76]Way, Frederick, Jr. "She Takes the Horns: Steamboat Racing on Western Waters", pp. 71–81, Young & Klein, Inc., Cincinnati, Ohio, 1953.

[77]H. H. Bancroft, "California Inter Pocula", San Francisco, 1888

[78]See Explorations in Iowa History Project, University Of Northern Iowa, https://iowahist.uni.edu/Frontier_Life/Stagecoach/Stagecoach.htm.

[79]Wright, Muriel H. "Historic Places on the Old Stage Line from Fort Smith to Red River – Appendix A", Chronicles of Oklahoma 11:2 (June 1933) 821–822.

E Estimated effect of the development of the railroad network on appraisers' travel time between custom offices

To illustrate how the development of the railroad network materially affected monitoring capacity, in section 2.2 we examined its impact on the district appraiser system established in 1851 to inspect custom offices. This section provides additional details on how we performed this calculation.

First, we obtain information on the 1851 geographic division of the country into 8 districts (U.S. Customs Service, 1988, pp. 44). For each district, we identify all counties with a custom office. Then, we use our data on travel time in 1850 between each pair of counties in the United States, to calculate the shortest path that could allow appraisers to visit each county with a custom office within their district.

Column 1 of Table E9 reports the minutes of uninterrupted travel time needed to visit each office in 1850. Columns 2-6 report counterfactual travel times, had the railroad network of each decade between 1860 and 1900 been in place in 1850. Finally, column 7 reports the percentage decrease in travel time between the estimate for 1850 (column 1) and the estimate for 1890 (column 5).

Table E9: Travel times across custom offices of each appraiser district between 1850 and 1900.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Minutes of uninterrupted travel in:						% change
District	1850	1860	1870	1880	1890	1900	1850-1890
Maine, New Hampshire, Massachusetts	7950	7071	6605	5134	4892	4040	38%
Rhode Island, Connecticut, New York	8887	6807	5721	4680	4126	3704	54%
New Jersey, Pennsylvania, Delaware	5891	4327	3051	2617	2391	2061	59%
Maryland, Virginia	5972	5460	5282	4587	4188	3855	30%
North Carolina, South Carolina, Georgia, Florida	18211	16145	14737	13083	10728	8773	41%
Florida, Alabama, Missouri, Louisiana, Texas	26477	22529	20938	16186	14041	11921	47%
California, Oregon	26351	24011	18668	15110	12080	9448	54%

Notes: the table presents, by decade, the estimated duration (in minutes) of uninterrupted travel required for each district appraiser to visit all the custom offices within their district and return to D.C. The last column shows the percentage change in travel time between 1890 and 1850. The boundaries of each district are taken in 1851.