# Online Appendix

## Patents, Innovation, and Growth

Davide M. Coluccia

December, 2024

# A Data and Sample Description

## A.I Details on Examiner Data

To assemble the dataset, we collect the name, surname, and USPTO division for all 204 examiners active at the patent office over the period 1919–1938. There are 65 divisions at the patent office in this period. However, divisions 47–65 are established after 1919. We thus exclude them and their patents (approximately 5% of the total) from the sample. This leaves us with 184 examiners. We successfully match 176 of them (over 96%) to Ancestry data. The matching is performed manually, and, in all cases, we match the records of the "Register" to only one person who, at some point, in the census indicates that their profession is "Patent Examiner" (or variants thereof). Their city of birth is identified either through census records close to the date of birth of the examiner or through World War 2 enlistment records, which indicate the town of birth. We then assign coordinates to the towns using a commercial geo-coding algorithm (Google Maps API). Table C.2 reports descriptive statistics on the final dataset.

## A.II Details on Patent Data

We collect the text of all patents issued in the United States between 1919 and 1938 from Google Patents. This repository also contains the CPC technology class(es) of each patent. Following Coluccia and Patacchini (2024), we apply large language models to the patent texts to extract the address of residence of the inventors (along with their name and surname, which we do not use). I geo-code the address using Google API to 1930 county borders. Patent grants indicate the county of residence of the inventor but do not report more disaggregated information. Therefore, I run the analysis at the county level. I also apply large language models to the patent texts to assign each patent to a USPTO division based on the description of each division's coverage, which I extract from the "Classification of subjects of invention," historical manuals for patent examiners.

We construct an alternative text-based similarity indicator to validate the baseline division similarity measure. Given the entire corpus of patent texts, we apply basic cleaning routines—remove stop words, lemmatization—and compute the term-frequency inverse-document frequency (TF-IDF) on the resulting dataset. The TF-IDF allows us to represent each patent as a 7,000-dimensional vector where each feature is one weighted word. For each couple of divisions $d$ and $d'$, we then compute the cosine similarity between all patents in $d$ and in $d'$ and take the average as the text-based similarity between the two divisions. Figure C.3 reports the correlation between the text-based and baseline measures.

### A.III Details on Census Data

From the population censuses 1900 to 1950, I extract information on population, adult population (aged over 18), employment, manufacturing employment ("OCC1950" 500 to 690, i.e., "Craftsmen" and "Operatives"), skilled employment ("OCC1950" 0 to 99 and 200 to 290, i.e., "Professionals" and "Managers"), occupational score ("OCCSCORE"), international immigrants and individuals born in another state (whom we label as "internal migrants"). As further controls, from the 1920 population census, I compute agricultural and mining ("IND1950" 100 to 239), construction ("IND1950" 246), manufacturing ("IND1950" 306 to 499), transportation and communications ("IND1950" 506 to 579), and trade and services employment ("IND1950" 606 to 699), and the number of illiterate individuals. I apply the method described by Eckert *et al.* (2020) to cross-walk all these variables to consistent 1930 county borders. The shares of all employment and imputed income variables are computed relative to the adult population, all other variables are relative to the overall population.

### A.IV Sample Construction

We assemble three datasets: panel "A" is a yearly panel of counties with information on patenting between 1919 and 1938; panel "B" follows county-by-division pairs over the same period; panel "C" contains data from the population census at the decade level between 1900 and 1950. In panels "A" and "B," we winsorize all patenting outcomes at the 5% level to avoid the possibility that outliers drive our estimates. We obtain very similar results without the winsorization. We use the measure developed by Kelly *et al.* (2021) to identify high-impact patents as those in the top 20% of the novelty distribution of their measure. We obtain similar results using alternative thresholds. In panels "A" (resp. panel "B") the main treatment variable is equal to one in county $c$ (resp. county $c$ and division $d$) after an examiner originating from a county closer than 100 Km from $c$ (resp. and active in division $d$) is appointed. In panel "C," we adopt the same definition except that the proximity threshold is set at 50 Km and the examiner must have been appointed in the preceding decade. Since we would only observe one post-treatment period for examiners appointed between 1930 and 1938, we drop them from the treatment definition.

## B Summary of the Robustness Checks

### B.I Descriptive Evidence

Table C.1 reports baseline descriptive statistics for the variables in the dataset. Table C.2 focuses on the individual-level examiner dataset we assembled for this paper. In figure C.2, we show that there is a (qualitative) positive correlation between the (log) number of patents (panel C.2a) and the (log) number of patents per capita (panel C.2b) in a given state and the number of examiners originating from that state. In figure C.1, we provide a visual summary of the research design. Counties in blue are the counties of origin of at least one examiner. Counties in red are those whose centroid lies within 50 Km from the blue counties, and together, these constitute the treatment group in the output regressions. Yellow counties, in turn, are those whose centroid lies within 100 Km from the blue counties, and together, these constitute the treatment group in the innovation regressions. All other counties constitute the control group.

### B.II Horserace Regressions

Tables C.3, C.4, and C.5 report the estimated treatment effect of local examiners on all patents, patents in the same division of the examiner, and patents in divisions other than the examiners, when controlling for a battery of county-level controls interacted with a post-examiner indicator. Figures C.9 and C.10 replicate the exercise for the output variables in levels and as shares of the population. The baseline results remain qualitatively and quantitatively unchanged throughout the specifications.

### B.III Leave-Out Estimation

Figure C.6 reports the estimated treatment effect of examiners on patenting when excluding one state at a time from the estimation sample. Figures C.7 and C.8 report the same exercise for the output variables in levels and as shares of the population. We find that no particular state drives the estimates, which remain almost always statistically significant.

### B.IV Accounting for Staggered Treatment Timing

The roll-out of examiners across counties and county-division pairs is staggered. In Figure C.11, we compare the estimated obtained using the baseline two-way fixed-effects (TWFE) estimator with those resulting from the stacked difference-in-differences estimator proposed in Cengiz *et al.* (2019). The TWFE estimates are virtually identical to the stacked estimator, indicating that non-negative weights do not appear to be a major concern in our setting. In Figures C.4 and C.5, we display fully flexible event-study estimates for all the output variables. We almost invariably find statistically insignificant pre-treatment coefficients and a gradual increase in the dependent variable after the examiner is appointed.

## B.V Assessing the Robustness of the Parallel Trends Assumption

Lastly, the validity of the research design requires that counties exposed and not exposed to examiners would not have displayed differential trajectories had the examiners not been appointed. While this assumption is hard to test, in Figure C.12, we estimate bounds on the relative magnitude of deviations from parallel trends following the procedure developed by Rambachan and Roth (2023). This exercise addresses the possibility that there may be unobserved shocks prior to the examiner's appointments that affected the exposed areas with differential intensity. For the case of overall patents (panel C.12a), the "breakdown" value for a null effect is close to 1, implying that our findings that the examiners had a significant effect on patents are valid inasmuch as we are willing to assume that the post-treatment violations of parallel trends is less than 100% as large as the largest pre-treatment violation. This value is closer to .5 when looking at patents by division (Figures C.12b–C.12c). In all cases, we interpret these results as corroborating the robustness of our findings. In figures C.13 and C.14, we replicate this exercise for the output variables in levels and as population shares. In most cases, the results confirm that our baseline estimates are robust to large deviations from the parallel trends assumption.

# C Tables and Figures

Table C.1. Descriptive Statistics

| | (1) Mean | (2) Std. Dev. | (3) Min. | (4) Max. | (5) Median | (6) Units | (7) Observations |
|---|---|---|---|---|---|---|---|
| **Panel A. Yearly Panel** | | | | | | | |
| N. of Patents | 8.333 | 63.982 | 0.000 | 3020.167 | 0 | 3,102 | 62,040 |
| N. of Breakthrough Patents | 0.242 | 2.758 | 0.000 | 163.500 | 0 | 3,102 | 62,040 |
| Share of Breakthrough Patents | 0.005 | 0.034 | 0.000 | 0.750 | 0 | 3,102 | 62,040 |
| Year of First Local Examiner | 1922.449 | 5.266 | 1919.000 | 1937.000 | 1,919 | 929 | 18,580 |
| Distance to Closest Examiner (Km) | 307.638 | 248.573 | 0.000 | 1419.370 | 228 | 3,102 | 46,530 |
| **Panel B. Decade-Level Panel** | | | | | | | |
| Population (1,000) | 36.910 | 124.213 | 0.004 | 4528.215 | 17 | 3,099 | 18,373 |
| International Immigrants (1,000) | 11.803 | 58.799 | 0.002 | 2925.326 | 3 | 3,099 | 18,373 |
| International Immigrants per Capita (%) | 27.582 | 19.816 | 0.342 | 100.000 | 23 | 3,099 | 18,373 |
| Internal Immigrants (1,000) | 7.721 | 35.897 | 0.001 | 2481.280 | 3 | 3,099 | 18,373 |
| Internal Immigrants per Capita (%) | 21.378 | 16.400 | 0.034 | 97.507 | 17 | 3,099 | 18,373 |
| Employment (1,000) | 12.174 | 47.230 | 0.000 | 1795.795 | 5 | 3,099 | 18,373 |
| Employment Rate (%) | 49.629 | 5.946 | 0.000 | 100.000 | 49 | 3,099 | 18,373 |
| Employment in Manufacturing (1,000) | 4.045 | 19.296 | 0.000 | 728.223 | 1 | 3,099 | 18,373 |
| Employment in Manufacturing per Capita (%) | 10.248 | 7.117 | 0.000 | 74.255 | 8 | 3,099 | 18,373 |
| High-Skill Employment (1,000) | 1.861 | 8.970 | 0.000 | 360.801 | 1 | 3,099 | 18,373 |
| High-Skill Employment Rate (%) | 5.971 | 2.134 | 0.000 | 20.833 | 6 | 3,099 | 18,373 |
| Imputed Income (10,000) | 31.458 | 134.986 | 0.000 | 5134.753 | 11 | 3,099 | 18,373 |
| Imputed Income (10,000) | 31.458 | 134.986 | 0.000 | 5134.753 | 11 | 3,099 | 18,373 |
| Imputed Income per Capita (%) | 1108.534 | 185.449 | 0.000 | 2158.654 | 1,075 | 3,099 | 18,373 |
| **Panel C. Cross-Sectional Controls (in 1920)** | | | | | | | |
| Agriculture & Mining (1,000) | 24.587 | 60.716 | 0.037 | 1691.997 | 15 | 3,094 | 3,094 |
| Manufacturing (1,000) | 3.159 | 16.581 | 0.000 | 406.902 | 0 | 3,094 | 3,094 |
| Transportation & Communication (1,000) | 1.116 | 5.319 | 0.000 | 181.539 | 0 | 3,094 | 3,094 |
| Finance & Business (1,000) | 0.000 | 0.000 | 0.000 | 0.000 | 0 | 3,094 | 3,094 |
| Trade & Services (1,000) | 2.538 | 13.872 | 0.000 | 419.962 | 1 | 3,094 | 3,094 |
| Illiterate (1,000) | 9.095 | 25.316 | 0.009 | 702.783 | 5 | 3,094 | 3,094 |

*Notes.* This table presents key descriptive statistics for the main variables used in the analysis. In panel A, the unit of observation is a county at a yearly level between 1918 and 1939. The panel reports patent and examiner—i.e., treatment—outcomes. In panel B, the unit of observation is a county at a decade frequency between 1900 and 1950. The panel reports the outcome variables constructed from the population censuses in levels and as shares of the population. The variables in levels are expressed in thousand units; the population shares are expressed in percentage terms. In panel C, the table reports the control variables constructed from the 1920 population census, which, in the horserace regressions, we interact with a post-treatment indicator to assess the robustness of the estimates. These variables are expressed in thousand units.

TABLE C.2. Descriptive Statistics on Examiner Data

| | (1) Mean | (2) Std. Dev. | (3) Min. | (4) Max. | (5) Median | (6) Observations |
|---|---|---|---|---|---|---|
| **Panel A. All Examiners** | | | | | | |
| Matched to Ancestry Data | 0.962 | 0.192 | 0.000 | 1.000 | 1.000 | 184 |
| Birth Year | 1881.403 | 14.872 | 1844.000 | 1909.000 | 1881.000 | 176 |
| White | 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 177 |
| Male | 0.994 | 0.075 | 0.000 | 1.000 | 1.000 | 177 |
| First-generation Immigrant | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 177 |
| Second-generation Immigrant | 0.220 | 0.415 | 0.000 | 1.000 | 0.000 | 173 |
| From North-east | 0.333 | 0.473 | 0.000 | 1.000 | 0.000 | 177 |
| From South | 0.090 | 0.288 | 0.000 | 1.000 | 0.000 | 177 |
| From Midwest | 0.294 | 0.457 | 0.000 | 1.000 | 0.000 | 177 |
| From West | 0.028 | 0.166 | 0.000 | 1.000 | 0.000 | 177 |
| From DC Area (DC, VA, MD) | 0.147 | 0.355 | 0.000 | 1.000 | 0.000 | 177 |
| **Panel B. Geo-coded Examiners** | | | | | | |
| Tenure (Years) | 9.308 | 7.392 | 0.000 | 35.000 | 8.000 | 198 |
| Latitude | 40.291 | 2.644 | 29.324 | 45.108 | 40.242 | 176 |
| Longitude | -80.807 | 8.750 | -120.620 | -67.162 | -77.069 | 176 |

*Notes.* This table presents key descriptive statistics of the individual-level dataset of principal examiners active at the United States Patent and Trademark Office between 1919 and 1938. Examiner names and divisions have been collected from the "Official Register of the United States" at a biennial frequency until 1921 and yearly thereafter. We then manually link the records of the Register to genealogy data held by `Ancestry.com`. In panel A, the first row of the table reports the matching rate. All variables are constructed from either the population censuses or other available information (such as World War 2 enlistment records). We assign the county of birth of the examiners to the six United States Census Bureau census regions, plus a seventh area that collects the Washington, DC, and the surrounding states (Virginia and Maryland). In panel B, we report information on the tenure and the latitude and longitude of the town of birth of each successfully geo-coded examiner.

TABLE C.3. Horserace Regressions on the Impact of Local Examiners on Patenting: All Divisions

| | Baseline | One Control at a Time | | | | | | | | | | All Controls |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Local Examiner × Post | 0.152*** | 0.103*** | 0.110*** | 0.098*** | 0.105*** | 0.103*** | 0.111*** | 0.123*** | 0.105*** | 0.125*** | 0.091*** | 0.077** |
| | (0.027) | (0.032) | (0.031) | (0.033) | (0.031) | (0.031) | (0.031) | (0.030) | (0.032) | (0.030) | (0.032) | (0.039) |
| Population × Post | | 0.003*** | | | | | | | | | | 0.100** |
| | | (0.001) | | | | | | | | | | (0.047) |
| Employment × Post | | | 0.006*** | | | | | | | | | -0.062 |
| | | | (0.002) | | | | | | | | | (0.048) |
| Agriculture Empl. × Post | | | | 0.005*** | | | | | | | | -0.087* |
| | | | | (0.001) | | | | | | | | (0.049) |
| Construction Empl. × Post | | | | | 0.094*** | | | | | | | -0.089 |
| | | | | | (0.021) | | | | | | | (0.178) |
| Manufacturing Empl. × Post | | | | | | 0.018*** | | | | | | -0.045 |
| | | | | | | (0.004) | | | | | | (0.038) |
| Transportation Empl. × Post | | | | | | | 0.056*** | | | | | -0.122* |
| | | | | | | | (0.014) | | | | | (0.065) |
| Trade Empl. × Post | | | | | | | | 0.016*** | | | | -0.058 |
| | | | | | | | | (0.006) | | | | (0.062) |
| Illiterate × Post | | | | | | | | | 0.011*** | | | 0.007 |
| | | | | | | | | | (0.003) | | | (0.051) |
| Immigrants × Post | | | | | | | | | | 0.006*** | | -0.014 |
| | | | | | | | | | | (0.002) | | (0.014) |
| Internal Migrants × Post | | | | | | | | | | | 0.021*** | 0.006 |
| | | | | | | | | | | | (0.004) | (0.011) |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Counties | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 |
| Number of Observations | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 |
| $R^2$ | 0.701 | 0.701 | 0.701 | 0.701 | 0.701 | 0.701 | 0.701 | 0.701 | 0.701 | 0.701 | 0.701 | 0.701 |
| Mean Dep. Var. | 2.879 | 2.879 | 2.879 | 2.879 | 2.879 | 2.879 | 2.879 | 2.879 | 2.879 | 2.879 | 2.879 | 2.879 |
| Std. Dev. Dep. Var. | 5.716 | 5.716 | 5.716 | 5.716 | 5.716 | 5.716 | 5.716 | 5.716 | 5.716 | 5.716 | 5.716 | 5.716 |

*Notes.* This table reports the effect of newly appointed examiners on patenting. The observation units are counties at a yearly frequency between 1919 and 1938. The dependent variable is the number of patents. The treatment is an indicator variable equal to one in counties that are exposed to an examiner after the examiner is appointed and zero otherwise. A county is exposed to examiners who are born in a county within 100 kilometers. In column (1), we report the baseline specification without any controls beyond the fixed effects; in columns (2–11), we include one control at a time. Each control is an interaction term between one variable computed on data from the 1920 population census and a post-treatment indicator. These variables are population (col. 2), employment (col. 3), agriculture employment (col. 4), construction employment (col. 5), manufacturing employment (col. 6), transportation and communication employment (col. 7), trade and services employment (col. 8), the number of illiterate individuals (col. 9), the number of international immigrants (col. 10), and the number of people born in another state (col. 11). In column (12), we include all the controls. All regressions include county and year fixed effects. Standard errors are clustered at the county level and are shown in parentheses.
***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.10$

Table C.4. Horserace Regressions on the Impact of Local Examiners on Patenting: Treated Divisions

| | Baseline | One Control at a Time | | | | | | | | | | All Controls |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Local Examiner × Post | 0.350*** | 0.347*** | 0.346*** | 0.350*** | 0.342*** | 0.343*** | 0.353*** | 0.344*** | 0.350*** | 0.352*** | 0.345*** | 0.443*** |
| | (0.055) | (0.074) | (0.072) | (0.075) | (0.074) | (0.076) | (0.070) | (0.067) | (0.073) | (0.066) | (0.078) | (0.093) |
| Population × Post | | 0.000 | | | | | | | | | | 0.113** |
| | | (0.001) | | | | | | | | | | (0.049) |
| Employment × Post | | | 0.000 | | | | | | | | | -0.046 |
| | | | (0.001) | | | | | | | | | (0.060) |
| Agriculture Empl. × Post | | | | -0.000 | | | | | | | | -0.193*** |
| | | | | (0.001) | | | | | | | | (0.067) |
| Construction Empl. × Post | | | | | 0.005 | | | | | | | 0.039 |
| | | | | | (0.022) | | | | | | | (0.256) |
| Manufacturing Empl. × Post | | | | | | 0.001 | | | | | | -0.014 |
| | | | | | | (0.004) | | | | | | (0.026) |
| Transportation Empl. × Post | | | | | | | -0.002 | | | | | -0.134*** |
| | | | | | | | (0.014) | | | | | (0.049) |
| Trade Empl. × Post | | | | | | | | 0.001 | | | | -0.003 |
| | | | | | | | | (0.003) | | | | (0.076) |
| Illiterate × Post | | | | | | | | | -0.000 | | | 0.166* |
| | | | | | | | | | (0.003) | | | (0.088) |
| Immigrants × Post | | | | | | | | | | -0.000 | | -0.054** |
| | | | | | | | | | | (0.001) | | (0.023) |
| Internal Migrants × Post | | | | | | | | | | | 0.001 | -0.034*** |
| | | | | | | | | | | | (0.006) | (0.012) |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Counties | 2,554 | 2,554 | 2,554 | 2,554 | 2,554 | 2,554 | 2,554 | 2,554 | 2,554 | 2,554 | 2,554 | 2,554 |
| Number of Observations | 51,080 | 51,080 | 51,080 | 51,080 | 51,080 | 51,080 | 51,080 | 51,080 | 51,080 | 51,080 | 51,080 | 51,080 |
| $R^2$ | 0.639 | 0.639 | 0.639 | 0.639 | 0.639 | 0.639 | 0.639 | 0.639 | 0.639 | 0.639 | 0.639 | 0.639 |
| Mean Dep. Var. | 1.716 | 1.716 | 1.716 | 1.716 | 1.716 | 1.716 | 1.716 | 1.716 | 1.716 | 1.716 | 1.716 | 1.716 |
| Std. Dev. Dep. Var. | 3.758 | 3.758 | 3.758 | 3.758 | 3.758 | 3.758 | 3.758 | 3.758 | 3.758 | 3.758 | 3.758 | 3.758 |

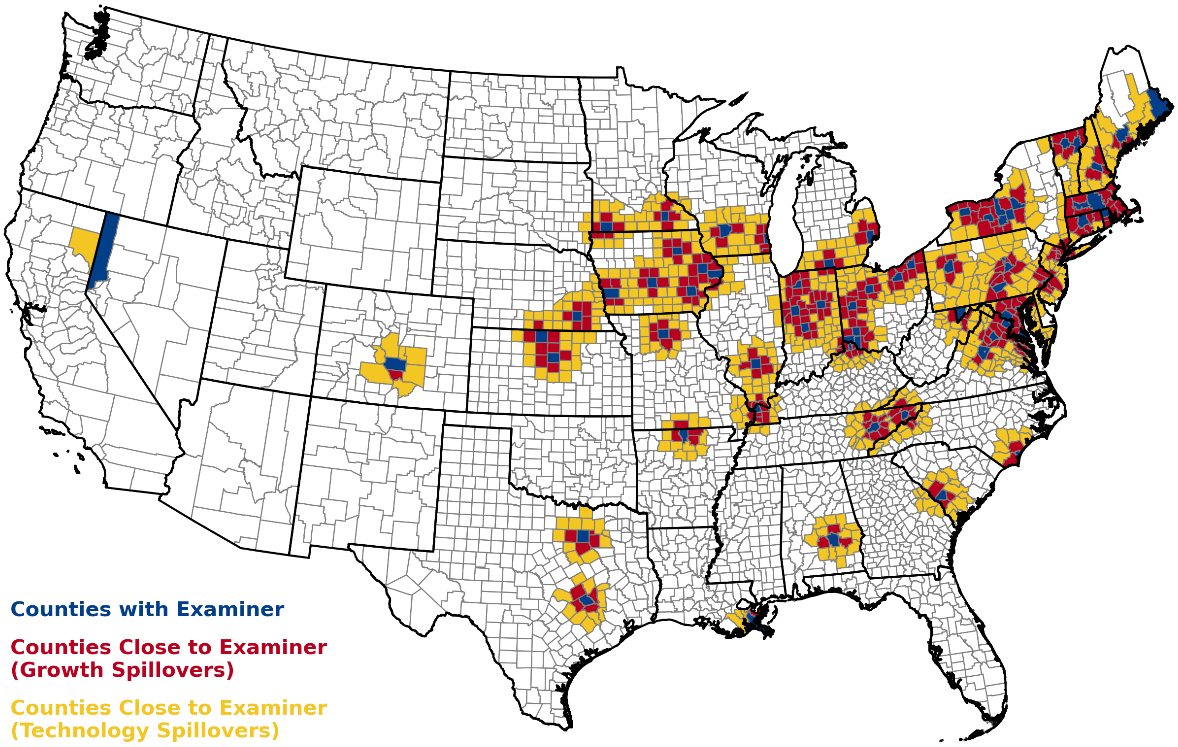
*Notes.* This table reports the effect of newly appointed examiners on patenting. The observation units are counties at a yearly frequency between 1919 and 1938. The dependent variable is the number of patents in the division of the examiner who triggers the treatment. The treatment is an indicator variable equal to one in counties that are exposed to an examiner after the examiner is appointed and zero otherwise. A county is exposed to examiners who are born in a county within 100 kilometers. In column (1), we report the baseline specification without any controls beyond the fixed effects; in columns (2–11), we include one control at a time. Each control is an interaction term between one variable computed on data from the 1920 population census and a post-treatment indicator. These variables are population (col. 2), employment (col. 3), agriculture employment (col. 4), construction employment (col. 5), manufacturing employment (col. 6), transportation and communication employment (col. 7), trade and services employment (col. 8), the number of illiterate individuals (col. 9), the number of international immigrants (col. 10), and the number of people born in another state (col. 11). In column (12), we include all the controls. All regressions include county and year fixed effects. Standard errors are clustered at the county level and are shown in parentheses.
***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.10$

Table C.5. Horserace Regressions on the Impact of Local Examiners on Patenting: Non-Treated Divisions

| | Baseline | One Control at a Time | | | | | | | | | | All Controls |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Local Examiner × Post | 0.147*** | 0.095*** | 0.103*** | 0.089*** | 0.099*** | 0.097*** | 0.103*** | 0.117*** | 0.097*** | 0.120*** | 0.084*** | 0.048 |
| | (0.027) | (0.031) | (0.031) | (0.032) | (0.031) | (0.031) | (0.031) | (0.030) | (0.031) | (0.030) | (0.031) | (0.039) |
| Population × Post | | 0.003*** | | | | | | | | | | 0.101* |
| | | (0.001) | | | | | | | | | | (0.054) |
| Employment × Post | | | 0.007*** | | | | | | | | | -0.076 |
| | | | (0.002) | | | | | | | | | (0.062) |
| Agriculture Empl. × Post | | | | 0.006*** | | | | | | | | -0.067 |
| | | | | (0.001) | | | | | | | | (0.056) |
| Construction Empl. × Post | | | | | 0.105*** | | | | | | | -0.148 |
| | | | | | (0.024) | | | | | | | (0.200) |
| Manufacturing Empl. × Post | | | | | | 0.020*** | | | | | | -0.047 |
| | | | | | | (0.004) | | | | | | (0.043) |
| Transportation Empl. × Post | | | | | | | 0.065*** | | | | | -0.120 |
| | | | | | | | (0.016) | | | | | (0.074) |
| Trade Empl. × Post | | | | | | | | 0.018*** | | | | -0.047 |
| | | | | | | | | (0.006) | | | | (0.075) |
| Illiterate × Post | | | | | | | | | 0.012*** | | | -0.023 |
| | | | | | | | | | (0.003) | | | (0.058) |
| Immigrants × Post | | | | | | | | | | 0.007*** | | -0.011 |
| | | | | | | | | | | (0.003) | | (0.016) |
| Internal Migrants × Post | | | | | | | | | | | 0.024*** | 0.012 |
| | | | | | | | | | | | (0.005) | (0.012) |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Counties | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 | 2,992 |
| Number of Observations | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 | 59,840 |
| $R^2$ | 0.661 | 0.661 | 0.661 | 0.661 | 0.661 | 0.661 | 0.661 | 0.661 | 0.661 | 0.661 | 0.661 | 0.661 |
| Mean Dep. Var. | 2.556 | 2.556 | 2.556 | 2.556 | 2.556 | 2.556 | 2.556 | 2.556 | 2.556 | 2.556 | 2.556 | 2.556 |
| Std. Dev. Dep. Var. | 4.735 | 4.735 | 4.735 | 4.735 | 4.735 | 4.735 | 4.735 | 4.735 | 4.735 | 4.735 | 4.735 | 4.735 |

*Notes.* This table reports the effect of newly appointed examiners on patenting. The observation units are counties at a yearly frequency between 1919 and 1938. The dependent variable is the number of patents in divisions other than that of the examiner who triggers the treatment. The treatment is an indicator variable equal to one in counties that are exposed to an examiner after the examiner is appointed and zero otherwise. A county is exposed to examiners who are born in a county within 100 kilometers. In column (1), we report the baseline specification without any controls beyond the fixed effects; in columns (2–11), we include one control at a time. Each control is an interaction term between one variable computed on data from the 1920 population census and a post-treatment indicator. These variables are population (col. 2), employment (col. 3), agriculture employment (col. 4), construction employment (col. 5), manufacturing employment (col. 6), transportation and communication employment (col. 7), trade and services employment (col. 8), the number of illiterate individuals (col. 9), the number of international immigrants (col. 10), and the number of people born in another state (col. 11). In column (12), we include all the controls. All regressions include county and year fixed effects. Standard errors are clustered at the county level and are shown in parentheses.
***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.10$

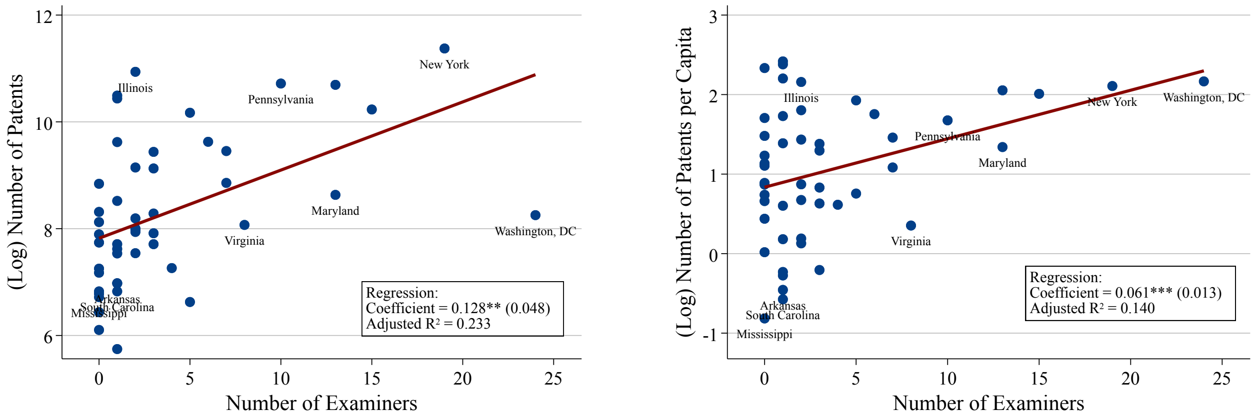
Figure C.1. Map of Counties with an Examiner, close to an Examiner, and without any Examiner




*Notes.* This map reports the distribution of treatment and control counties across the various empirical exercises. The counties of origin of at least one principal examiner active at the USPTO between 1919 and 1938 are displayed in blue; counties whose centroid lies within 50 Km from the blue counties are marked in red and are considered treated in the growth analysis; counties whose centroid lies within 100 Km from the blue counties are marked in yellow and are considered treated in the innovation analysis. All other counties are part of the control group. The county borders refer to 1930. The solid black lines superimpose the contemporaneous state borders.

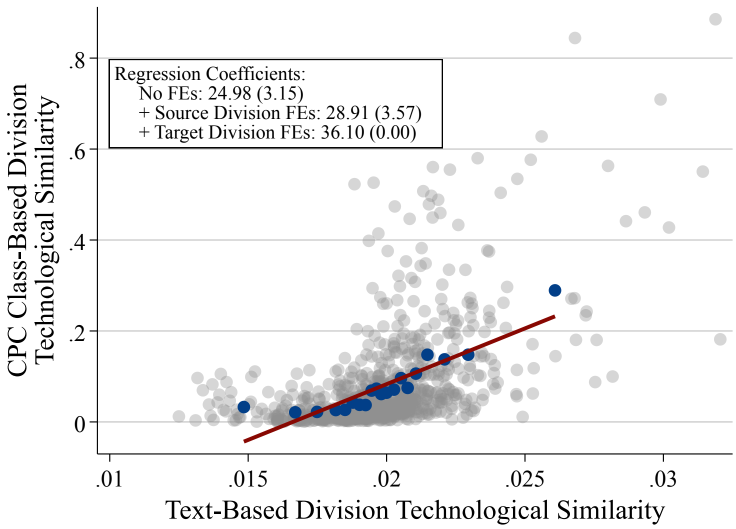
FIGURE C.2. Correlation Between USPTO Examiners and Patenting

(A) (Log) Patents (B) (Log) Patents per Capita



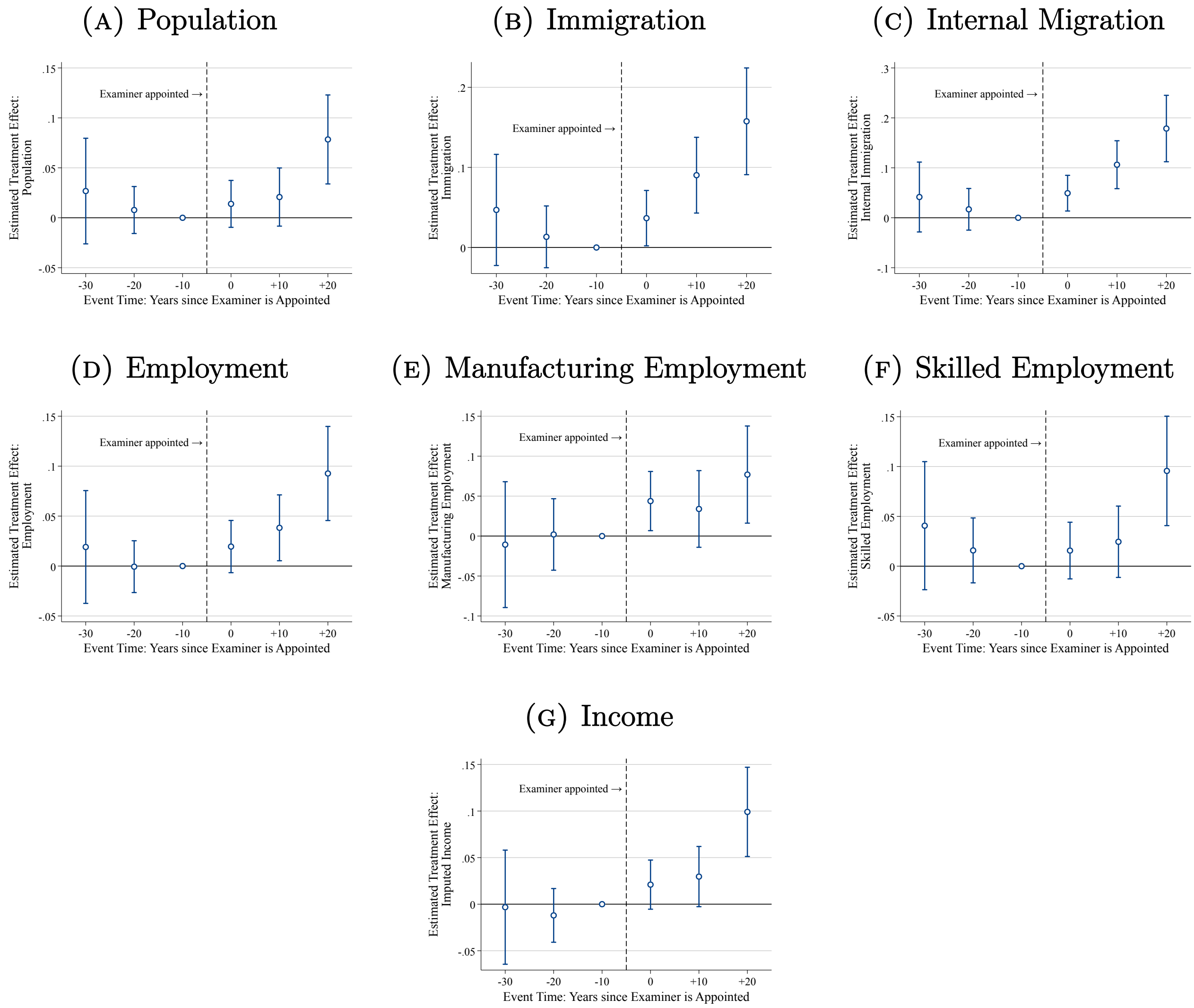

*Notes.* The graphs display the correlation between patenting and the number of USPTO principal examiners. Each dot refers to one state. On the $x$-axis, we display the number of examiners active between 1919 and 1938 who originated from each state. On the $y$-axis, panel C.2a reports the (log) number of patents issued to inventors residing in the given state over the same period; panel C.2b reports the (log) number of patents issued to inventors residing in the given state over the same period normalized by the average state population. The red line superimposes a linear regression. Each graph reports the coefficient of the associated regression along with its standard error clustered at the state level and the adjusted $R^2$. The labels highlight the dots of the DC area (Washington, DC, Virginia, and Maryland), the three most innovative states (New York, Illinois, and Pennsylvania), and the three least innovative states outside of the West (Mississippi, Arkansas, and South Carolina).

FIGURE C.3. Correlation Between CPC-Based and Text-Based Division Similarities




*Notes.* This Figure reports the correlation between the baseline division similarity obtained through CPC technologies ($y$-axis) and the text-based division similarity described in Appendix section A.II ($x$-axis). Each gray dot refers to a division-division pair. The figure overlays the binned means in blue and a linear fit in red. In addition, we report the regression coefficient between the two measures without any controls, including fixed effects for the initial ($y$-axis) division, and including division-division dyadic fixed effects.

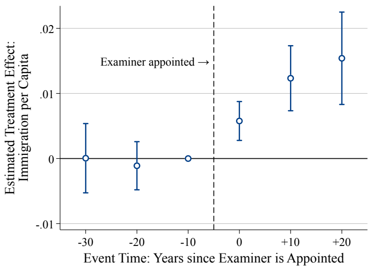
FIGURE C.4. Local Examiner and Growth in Levels: Fully Flexible Specification



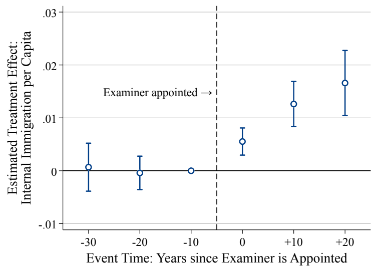

*Notes.* This figure reports the effect of newly appointed examiners on proximate indicators of economic growth. The observation units are counties at a decade frequency between 1900 and 1950. The dependent variable is: population (panel C.4a), overall and internal immigration (panels C.4b and C.4c), overall, manufacturing, and high-skilled employment (panels C.4d, C.4e, and C.4f), and the occupational income score (panel C.4g). The treatment codes the number of decades since an examiner to whom the county is exposed is appointed. A county is exposed to examiners who are born in a county within 50 kilometers. All variables are expressed in logs. All regressions include county and state-by-year fixed effects. Standard errors are clustered at the county level; bands report 95% confidence intervals.

FIGURE C.5. Local Examiner and Growth as Shares of the Population: Fully Flexible Specification

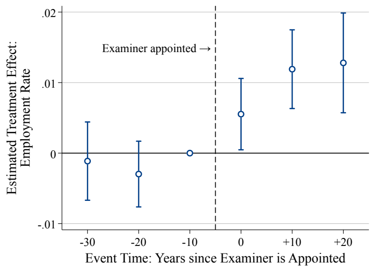
(A) Share of Immigrants



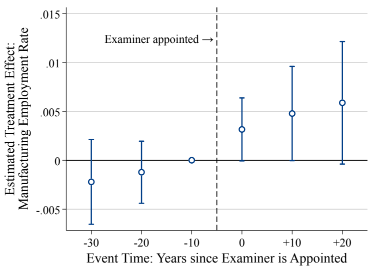

(B) Share of Internal Migrants



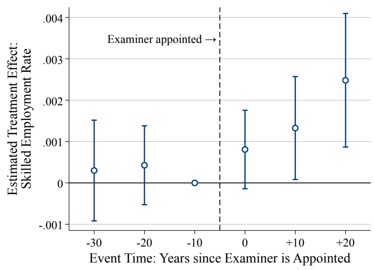

(C) Employment Rate



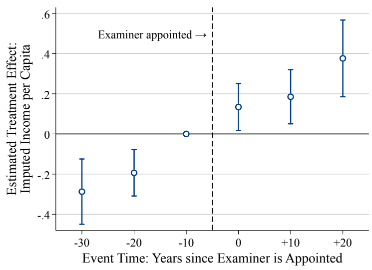

(D) Manufacturing Employment Rate




(E) Skilled Employment Rate



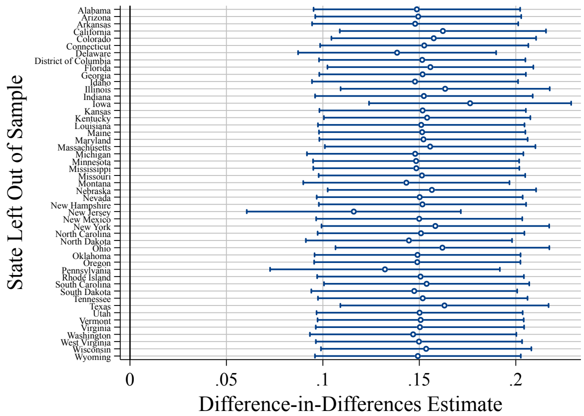

(F) Income per Capita



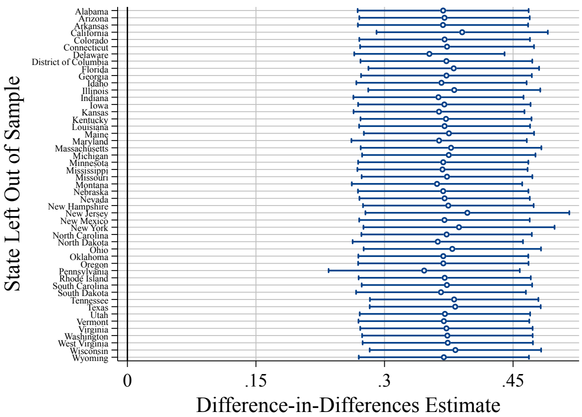

*Notes.* This figure reports the effect of newly appointed examiners on proximate indicators of economic growth. The observation units are counties at a decade frequency between 1900 and 1950. The dependent variable is: the share of immigrants (panel C.5a), the share of internal migrants (panel C.5b), overall, manufacturing, and high-skilled employment rates (panels C.5c, C.5d, and C.5e), and the occupational income score per capita (panel C.5f). The treatment codes the number of decades since an examiner to whom the county is exposed is appointed. A county is exposed to examiners who are born in a county within 50 kilometers. All variables are expressed as shares of the population. All regressions include county and state-by-year fixed effects. Standard errors are clustered at the county level; bands report 95% confidence intervals.

Figure C.6. Leave-One-Out Estimates: Examiners and Patenting

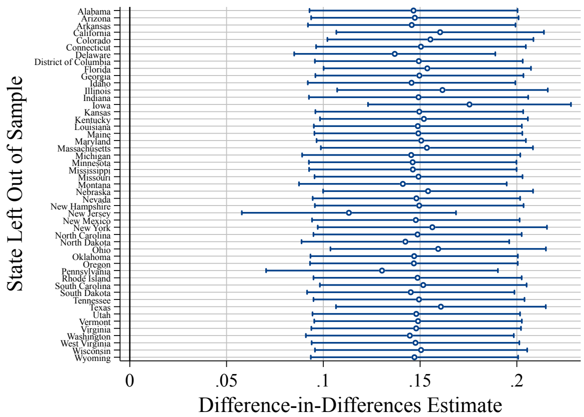
(a) All USPTO Divisions




(b) Treated USPTO Divisions



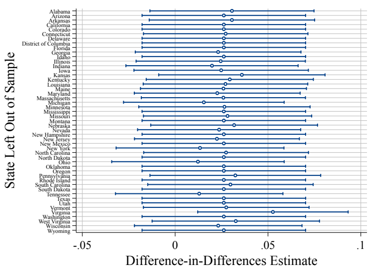

(c) Non-Treated USPTO Divisions



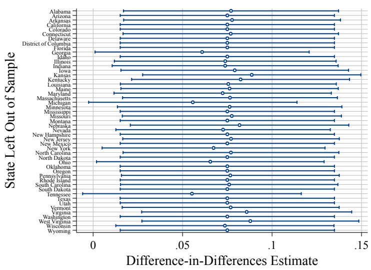

*Notes.* This figure reports the effect of newly appointed examiners on patenting. The observation units are counties at a yearly frequency between 1919 and 1938. The dependent variable is the number of patents. Each dot reports the estimated coefficient associated with a treatment term equal to one in counties that are exposed to an examiner after the examiner is appointed and zero otherwise. A county is exposed to examiners who are born in a county within 100 kilometers. In panel C.6a, patents are aggregated across USPTO divisions; in panel C.6b, we include only patents in the same division of the newly appointed examiner; in panel C.6c, we include only patents in divisions other than that of the examiner. In each panel, we report the estimated treatment effect for 49 separate regressions, each of which excludes one state from the estimation sample. Excluded states are displayed on the $y$-axis. The model is Poisson quasi-maximum likelihood. All regressions include county and year fixed effects. Standard errors are clustered at the county level and are shown in parentheses; bands report 95% confidence intervals.

Figure C.7. Leave-One-Out Estimates: Examiners and Growth, Levels

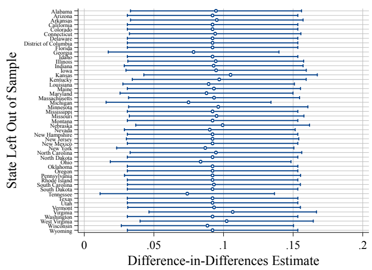
(a) Population



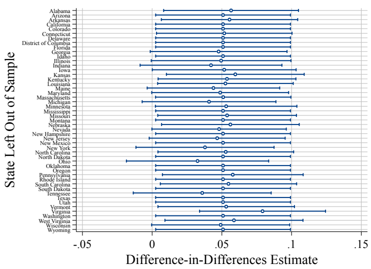

(b) Immigration



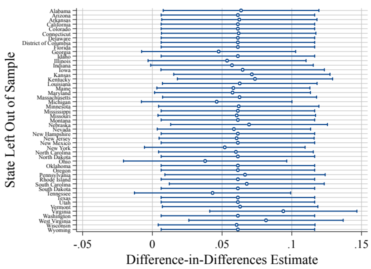

(c) Internal Migration



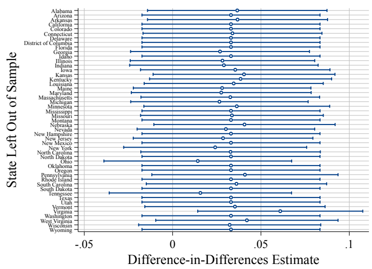

(d) Employment



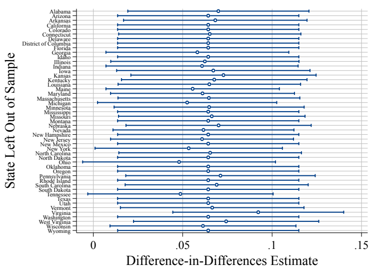

(e) Manufacturing Employment




(f) Skilled Employment



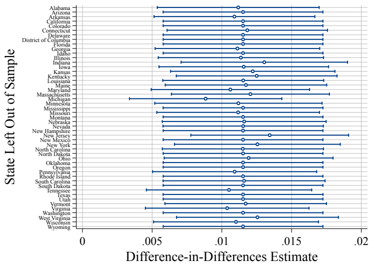

(g) Income



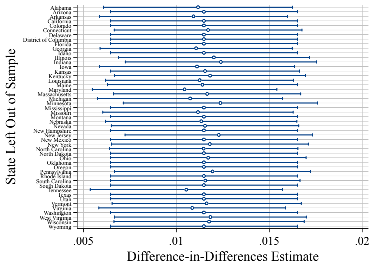

*Notes.* This figure reports the effect of newly appointed examiners on proximate indicators of economic growth. The observation units are counties at a decade frequency between 1900 and 1950. The dependent variable is: population (panel C.7a), overall and internal immigration (panels C.7b and C.7c), overall, manufacturing, and high-skilled employment (panels C.7d, C.7e, and C.7f), and the occupational income score (panel C.7g). Each dot reports the estimated coefficient associated with the years since an examiner in the proximity of a county is appointed. A county is exposed to examiners who are born in a county within 50 kilometers. In each panel, we report the estimated treatment effect for 49 separate regressions, each of which excludes one state from the estimation sample. Excluded states are displayed on the $y$-axis. The model is Poisson quasi-maximum likelihood. All variables are expressed in logs. All regressions include county and state-by-year fixed effects. Standard errors are clustered at the county level; bands report 95% confidence intervals.

FIGURE C.8. Leave-One-Out Estimates: Examiners and Growth, Share of the Population

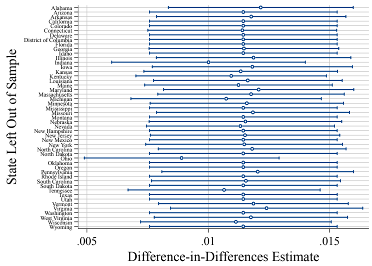
(A) Share of Immigrnats



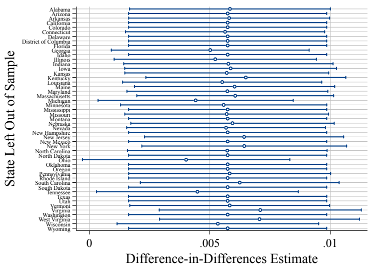

(B) Share of Internal Migrants



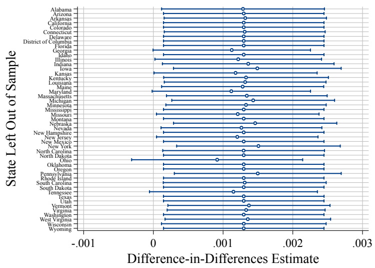

(C) Employment Rate



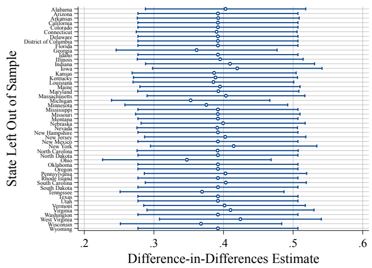

(D) Manufacturing Employment Rate




(E) Skilled EmploymentRate



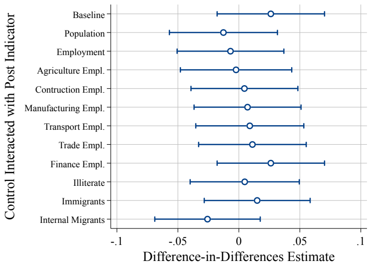

(F) Income Per Capita



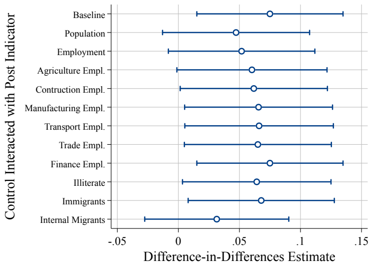

*Notes.* This figure reports the effect of newly appointed examiners on proximate indicators of economic growth. The observation units are counties at a decade frequency between 1900 and 1950. The dependent variable is: the share of immigrants (panel C.8a), the share of internal migrants (panel C.8b), overall, manufacturing, and high-skilled employment rates (panels C.8c, C.8d, and C.8e), and the occupational income score per capita (panel C.8f). Each dot reports the estimated coefficient associated with the years since an examiner in the proximity of a county is appointed. A county is exposed to examiners who are born in a county within 50 kilometers. In each panel, we report the estimated treatment effect for 49 separate regressions, each of which excludes one state from the estimation sample. Excluded states are displayed on the $y$-axis. The model is Poisson quasi-maximum likelihood. All variables are expressed as shares of the population. All regressions include county and state-by-year fixed effects. Standard errors are clustered at the county level; bands report 95% confidence intervals.

FIGURE C.9. Horserace Estimates: Examiners and Growth, Levels

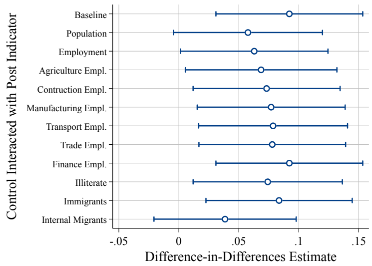
(A) Population



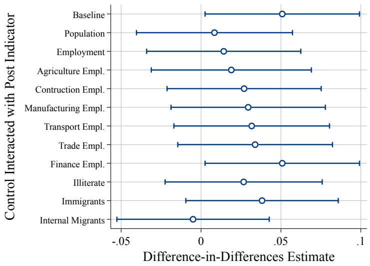

(B) Immigration



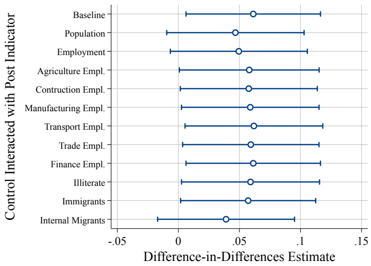

(C) Internal Migration



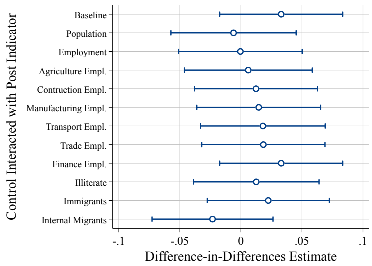

(D) Employment



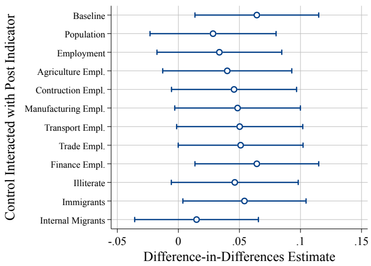

(E) Manufacturing Employment




(F) Skilled Employment



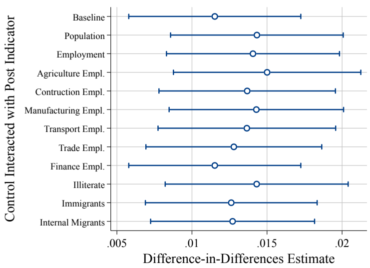

(G) Income



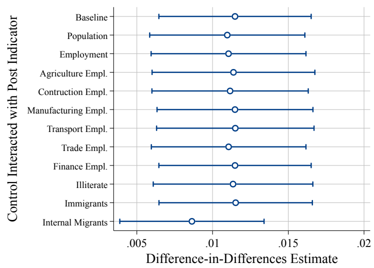

*Notes.* This figure reports the effect of newly appointed examiners on proximate indicators of economic growth. The observation units are counties at a decade frequency between 1900 and 1950. The dependent variable is: population (panel C.9a), overall and internal immigration (panels C.9b and C.9c), overall, manufacturing, and high-skilled employment (panels C.9d, C.9e, and C.9f), and the occupational income score (panel C.9g). Each dot reports the estimated coefficient associated with the years since an examiner in the proximity of a county is appointed. A county is exposed to examiners who are born in a county within 50 kilometers. In each panel, we report the estimated treatment effect for various separate regressions, each of which includes one variable measured in 1900 and interacted with the post-treatment indicator variable. The included control interaction terms are displayed on the $y$-axis. The model is Poisson quasi-maximum likelihood. All variables are expressed in logs. All regressions include county and state-by-year fixed effects. Standard errors are clustered at the county level; bands report 95% confidence intervals.

Figure C.10. Horserace Estimates: Examiners and Growth, Shares of the Population

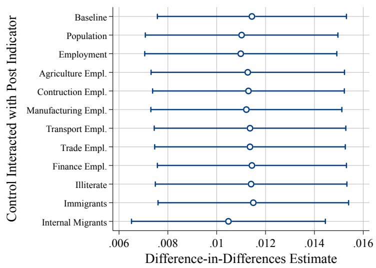
(a) Share of Immigrants



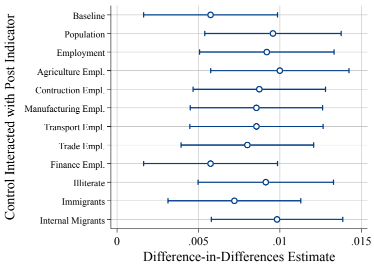

(b) Share of Internal Migrants



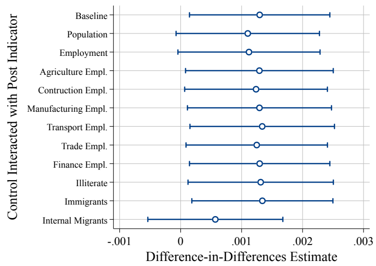

(c) Employment Rate



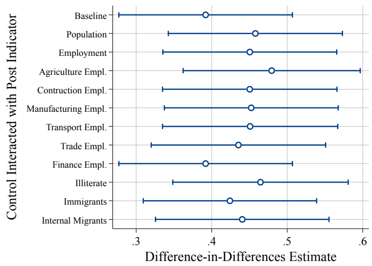

(d) Manufacturing Employment Rate




(e) Skilled Employment Rate



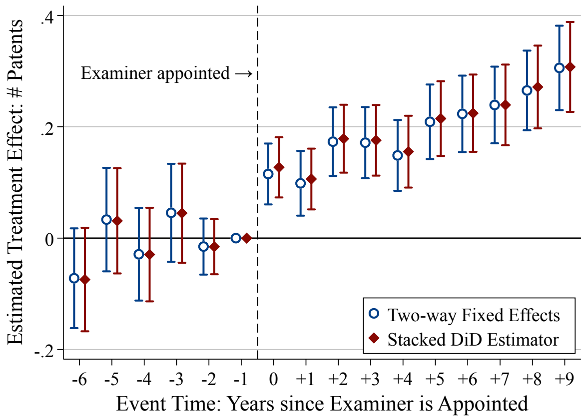

(f) Income Per Capita



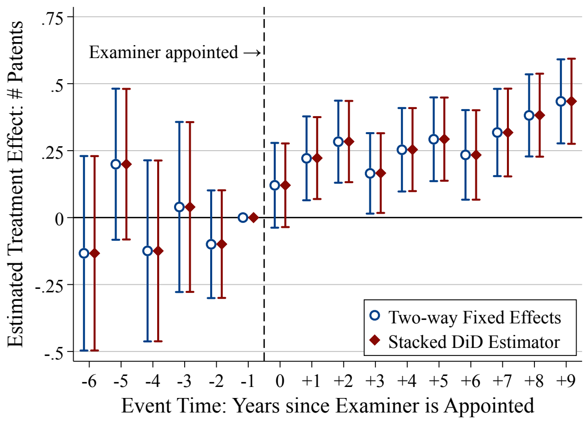

*Notes.* This figure reports the effect of newly appointed examiners on proximate indicators of economic growth. The observation units are counties at a decade frequency between 1900 and 1950. The dependent variable is: the share of immigrants (panel C.10a), the share of internal migrants (panel C.10b), overall, manufacturing, and high-skilled employment rates (panels C.10c, C.10d, and C.10e), and the occupational income score per capita (panel C.10f). Each dot reports the estimated coefficient associated with the years since an examiner in the proximity of a county is appointed. A county is exposed to examiners who are born in a county within 50 kilometers. In each panel, we report the estimated treatment effect for various separate regressions, each of which includes one variable measured in 1900 and interacted with the post-treatment indicator variable. The included control interaction terms are displayed on the $y$-axis. The model is Poisson quasi-maximum likelihood. All variables are expressed as shares of the population. All regressions include county and state-by-year fixed effects. Standard errors are clustered at the county level; bands report 95% confidence intervals.

Figure C.11. Local Examiners and Innovation: Alternative Estimator

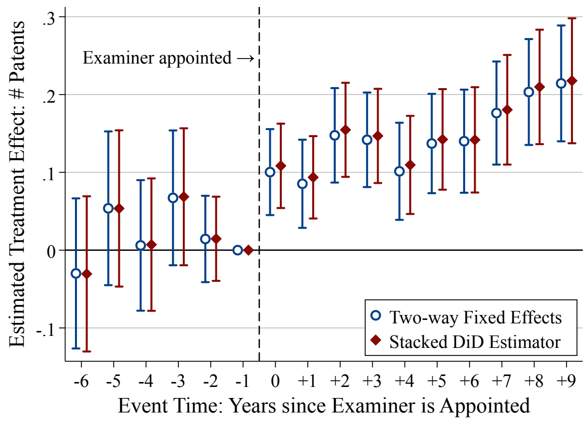
(a) All USPTO Divisions




(b) Treated USPTO Divisions



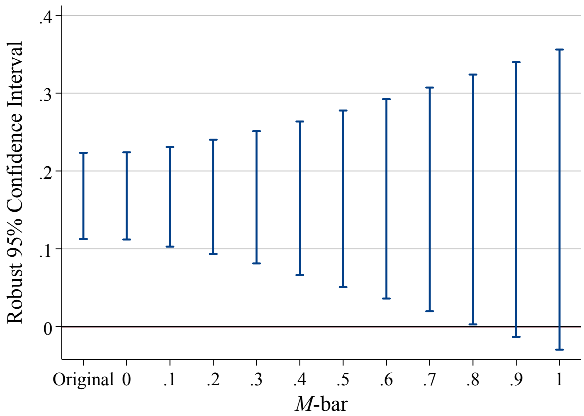

(c) Non-Treated USPTO Divisions



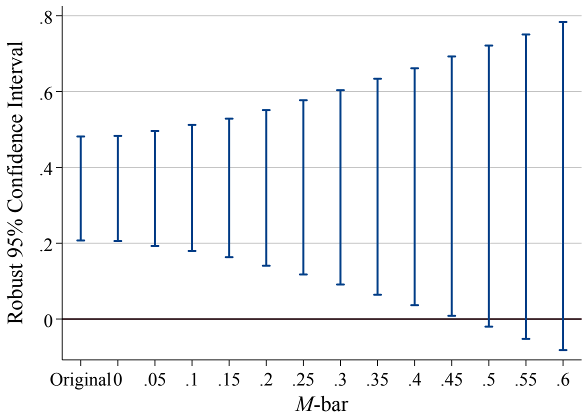

*Notes.* This figure reports the effect of the appointment of an examiner on patenting activity in their area of origin. The observation units are counties at a yearly frequency between 1919 and 1938. The dependent variable is the number of patents. Each dot reports the estimated coefficient associated with the years since an examiner in the proximity of a county is appointed. A county is exposed to examiners who are born in a county within 100 kilometers. In panel C.11a, patents are aggregated across USPTO divisions; in panel C.11b, we include only patents in the same division of the newly appointed examiner; in panel C.11c, we include only patents in divisions other than that of the examiner. In each panel, we report the estimated treatment effects obtained through the baseline two-way fixed effects estimator (blue dots) and those obtained using the stacked difference-in-differences estimator described in Cengiz *et al.* (2019). The model is Poisson quasi-maximum likelihood. All regressions include county and year fixed effects. Standard errors are clustered at the county level and are shown in parentheses; bands report 95% confidence intervals.

FIGURE C.12. Robustness of Parallel Trends Assumption: Local Examiner and Patenting

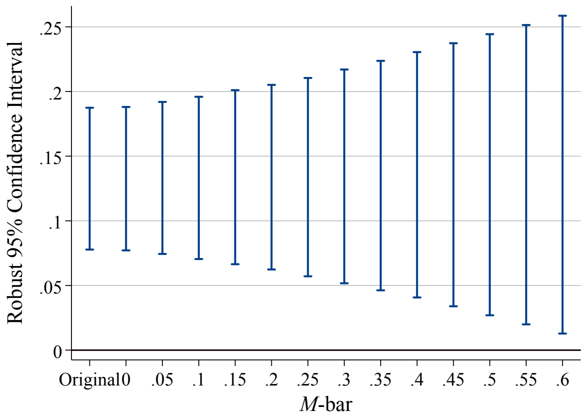
(A) All USPTO Divisions




(B) Treated USPTO Divisions

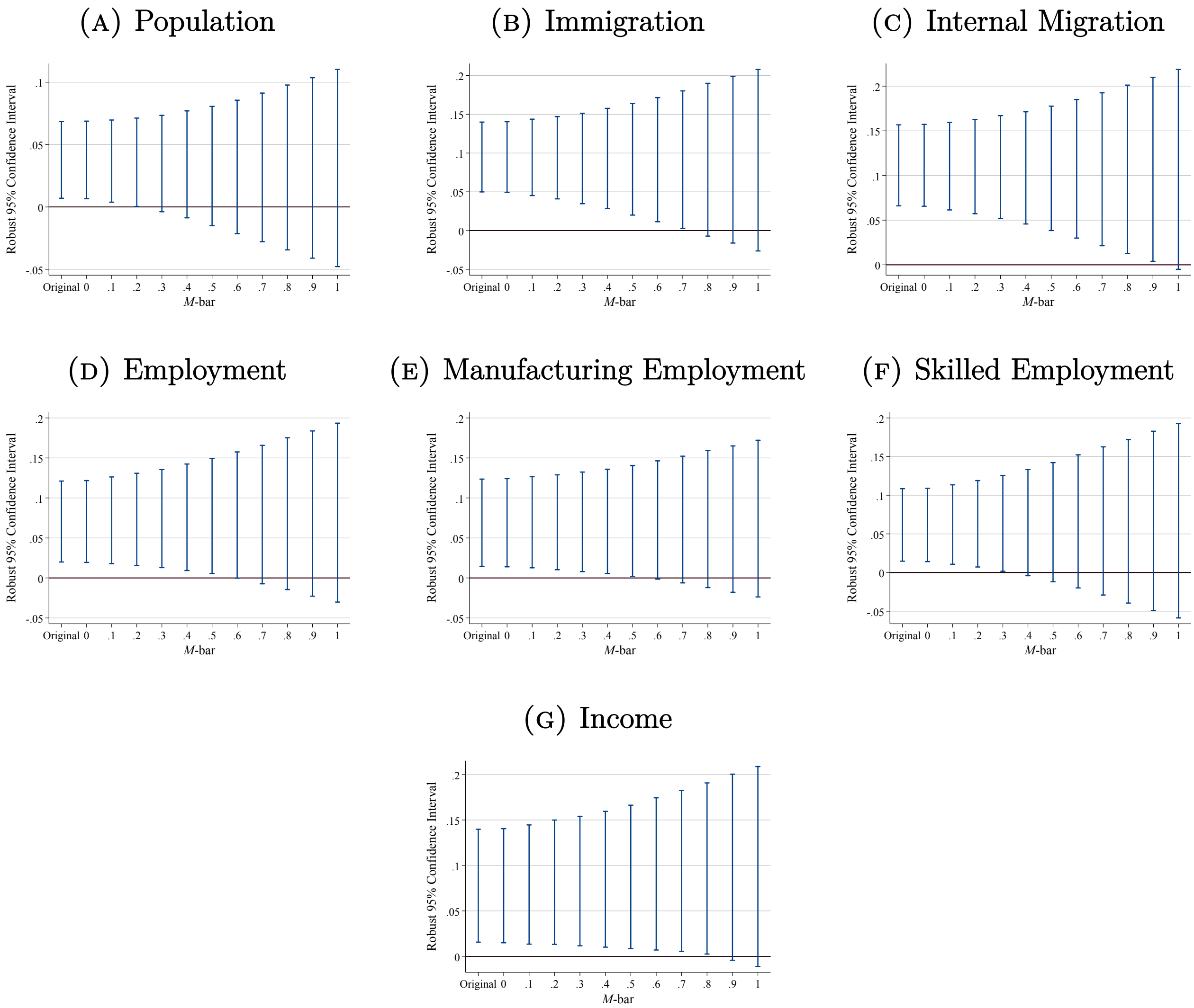



(C) Non-Treated USPTO Divisions




*Notes.* This figure reports the effect of the appointment of an examiner on patenting activity in their area of origin. The observation units are counties at a yearly frequency between 1919 and 1938. The dependent variable is the number of patents. The treatment is an indicator variable equal to one in counties that are exposed to an examiner after the examiner is appointed and zero otherwise. A county is exposed to examiners who are born in a county within 100 kilometers. In panel C.12a, patents are aggregated across USPTO divisions; in panel C.12b, we include only patents in the same division of the newly appointed examiner; in panel C.12c, we include only patents in divisions other than that of the examiner. In each panel, The bands report 95% robust confidence intervals constructed following the method developed by Rambachan and Roth (2023). The model is Poisson quasi-maximum likelihood. All regressions include county and year fixed effects.

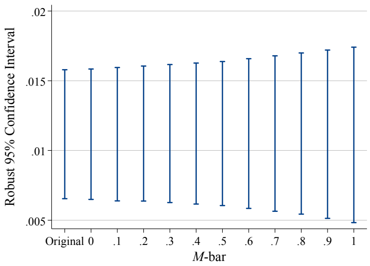
FIGURE C.13. Robustness of Parallel Trends: Local Examiner and Growth in Levels



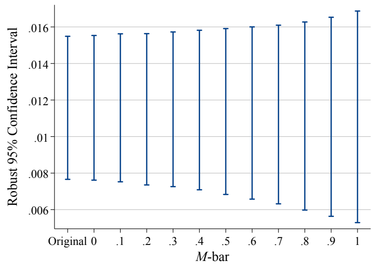

*Notes.* This figure reports the effect of newly appointed examiners on proximate indicators of economic growth. The observation units are counties at a decade frequency between 1900 and 1950. The dependent variable is: population (panel C.13a), overall and internal immigration (panels C.13b and C.13c), overall, manufacturing, and high-skilled employment (panels C.13d, C.13e, and C.13f), and the occupational income score (panel C.13g). Each dot reports the estimated coefficient associated with the years since an examiner in the proximity of a county is appointed. A county is exposed to examiners who are born in a county within 50 kilometers. The bands report 95% robust confidence intervals constructed following the method developed by Rambachan and Roth (2023). The model is Poisson quasi-maximum likelihood. All variables are expressed in logs. All regressions include county and state-by-year fixed effects.

Figure C.14. Robustness of Parallel Trends: Local Examiner and Growth as Shares of the Population

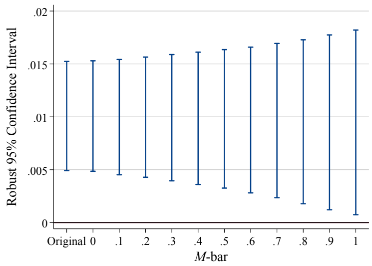
(a) Share of Immigrants



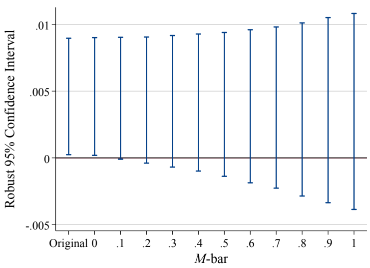

(b) Share of Internal Migrants



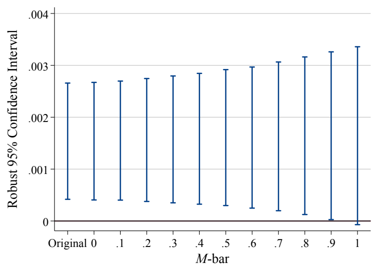

(c) Employment Rate



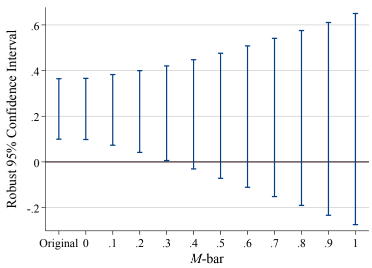

(d) Manufacturing Employment Rate




(e) Skilled Employment Rate




(f) Income per Capita




*Notes.* This figure reports the effect of newly appointed examiners on proximate indicators of economic growth. The observation units are counties at a decade frequency between 1900 and 1950. The dependent variable is: the share of immigrants (panel C.14a), the share of internal migrants (panel C.14b), overall, manufacturing, and high-skilled employment rates (panels C.14c, C.14d, and C.14e), and the occupational income score per capita (panel C.14f). Each dot reports the estimated coefficient associated with the years since an examiner in the proximity of a county is appointed. A county is exposed to examiners who are born in a county within 50 kilometers. The bands report 95% robust confidence intervals constructed following the method developed by Rambachan and Roth (2023). The model is Poisson quasi-maximum likelihood. All variables are expressed as shares of the population. All regressions include county and state-by-year fixed effects.

# Appendix References


Cengiz, D., A. Dube, A. Lindner and B. Zipperer (2019). "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics*, 134(3): 1405–1454.

Coluccia, D. M. and E. Patacchini (2024). "Uniting Diversity: Urban Infrastucture and Innovation in the United States." *Working Paper.*

Eckert, F., A. Gvirtz, J. Liang and M. Peters (2020). "A Method to Construct Geographical Crosswalks with an Application to US Counties Since 1790." *NBER Working Paper*, (No. w26770).

Kelly, B., D. Papanikolaou, A. Seru and M. Taddy (2021). "Measuring Technological Innovation Over the Long Run." *American Economic Review: Insights*, 3(3): 303–320.

Rambachan, A. and J. Roth (2023). "A More Credible Approach to Parallel Trends." *Review of Economic Studies*, 90(5): 2555–2591.