Online Appendix for "Human Capital Transfers and Sub-national Development: Armenian and Greek Legacy in Post-expulsion Turkey"

A Data Appendix

A.1 Variable Definitions and Sources

Population Density: Natural logarithm of district population per square kilometer. It is computed using the Turkish Population Censuses and the surface area of each district as reported by the National Mapping Agency of Turkey under the Ministry of National Defense. Census results can be accessed through TurkStat's web application. As for the 1893 population density, where it is necessary, we use a rough proxy. 1893 census figures are reported for Ottoman districts (*kazas*), and information about their boundaries are not available. This makes it impossible to compute population density as the areas of *kazas* are not known. Therefore, we use the sum of areas of all modern districts that were name-matched to a given *kaza* as a proxy for total area of that *kaza*. Using this proxy, we construct population density figures in 1893 for each *kaza*.

Urbanization Rate: The share of district population who lives within the municipal boundaries that define the district centers. It is computed using data from the Turkish Population Census on the distribution of the population within and outside the district centers.

Average Luminosity: The variable measures for a given year the density of time-stable nighttime lights at the district level. The information on light density comes from the Defense Meteorological Satellite Program's (DMSP) Operational Linescan System. DMSP reports images of the earth at night captured from 20:30 to 22:00 local time. The satellites detect lights from human settlements, fires, gas flares, lightning, and the aurora. Light density measure is a six-bit number (ranging from 0 to 63) calculated for every 30-second area (approximately 1 square kilometer). Overlaying all images captured during a calendar year, dropping images where lights are shrouded by cloud or overpowered by the aurora or solar glare (near the poles), and removing ephemeral lights like fires and lightning, an annual composite image of time-stable lights are created. We compute district level luminosity by

averaging across all light density pixels that fall within the district boundaries. Figure A.1 depicts the cross-district distribution of average luminosity in 2000 along with the historical representation of Armenians and Greeks in the Ottoman population. Darker shades show districts with lower economic activity proxied by luminosity. In regression tables where we control for contemporary variables measured in year 2000, we employ the average luminosity for the period 2007-2013 as our outcome measure instead of luminosity in year 2000. For years in which luminosity is recorded by multiple satellites, we use the arithmetic average of the values in these rasters.

Historical Minority Population Shares: There are two potential sources of data on historical population of minorities. The first one is the Ottoman General Census of 1881/82-1893. The second one is the Population Statistics of the Ottoman State in 1914, i.e. the year before the mass deportations of Armenians started.

The 1914 statistics were prepared using the figures from the 1905/1906 census, and adjusting for births and deaths registered in the subsequent years. Various tribes in Eastern Anatolia could not be counted, and hence, the information on the population size of these tribes was based on estimates. More importantly for the purpose of our analysis, the major problem with the 1914 population figures is that, in several regions of the Eastern and Southeastern Anatolia, the tensions between Armenians and the state forces have intensified during the final years of the reign of Sultan Abdulhamid II. Armenian national movement also gained momentum in this period. In some regions, Armenians organized armed self-defense forces in response to attacks by Kurdish tribesmen and irregulars. Armenian revolutionary activity in the East and the ensuing violence were met with a heavily armed response by the central government. In the mid-1890s, several massacres took place against the Armenians in the eastern provinces of the Ottoman Empire. These massacres were carried out by the Hamidian Regiments –irregular corps armed by the state and named after the sultan. They led to 200,000 to 300,000 dead according to some estimates (Akçam, 2006). During this period, several regions in the East of Anatolia have been the stage of the Armenian uprisings, and the clashes between Armenian militia and Ottoman Empire's forces including the Sasun Rebellion of 1894, the Zeitun Rebellion of 1895-1896 and the 1896 Defense of Van. The incidents continued in the immediate aftermath of the Young Turk Revolution of 1908. In April 1909, anti-Armenian pogroms in Adana Vilayet resulted in the deaths of as many as 20,000–30,000 Armenians (Adalian, 2010). The casualties caused by sporadic clashes between state forces and the Armenian rebels, the civilians who died during the massacres committed against Armenians over the period between 1894 and 1914, and the people who migrated elsewhere to escape from violence all make the 1914 population figures less suitable for an analysis of the long-term legacy of Armenian communities in Anatolia. In addition,

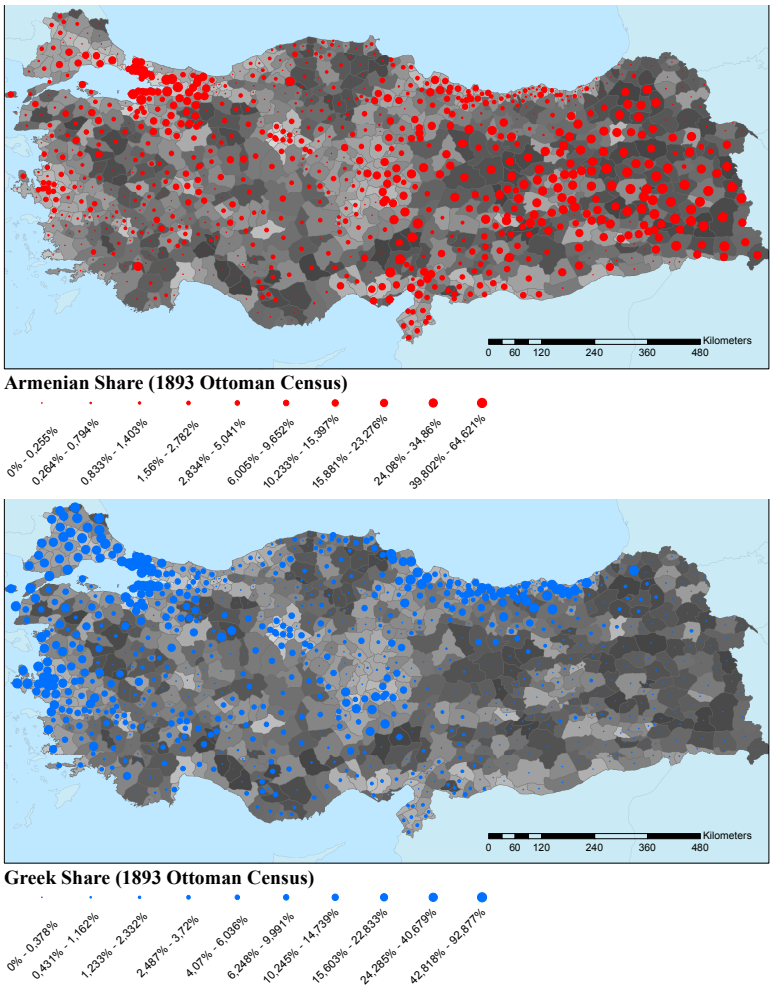


FIGURE A.1: Minority presence (1893) and average luminosity (2000) across Turkish districts

the Young Turk government, who took power in 1908 and initiated the Turkification of Anatolia, had incentives to under-report minority figures to deny them higher representation in the state and the military, which makes 1914 population numbers less reliable (Üngör and Polatel, 2011).

Therefore, in the construction of the historical population measures, we use data on Greek, Armenian and total population reported by the Population Census of the Ottoman Empire that was conducted during the period 1881-1893. The census measures were reported either at the kaza (Ottoman district) or independent sancak level (when there is no kaza designation). The variables measure for each modern district in 2000, the share in total population of Armenian and Greek inhabitants of the Ottoman location (kaza or sancak) that was matched with this district. In rare cases when a given modern district is matched

with multiple Ottoman kazas, the minority population shares reflect the overall share of these populations in the combination of these kazas. 1881-1893 Ottoman Census was the first census where females were also counted. The census used several ethnic-confessional categories for the Christian population. The 1893 population data used in this study were published for the first time by Karpat (1985). As Karpat (1985) puts it, "These population records issued in 1893 represent the most complete and reliable Ottoman population figures compiled in the nineteenth century. Unlike earlier general population statistics, these gave precise and detailed information on the population of all areas, noting the districts and regions where the census was not completed and providing estimates for the areas not subjected to individual census and registration. There were a few areas within the modern Turkish boundaries –namely, Erzurum, Bitlis, Elaziz and Van– where the census counts were known to be incomplete due to the practical difficulty of counting nomadic tribes. Although the Ottoman statistical office reported the names of the specific vilayets, sancaks, and tribes for which counts were incomplete and provide population estimates for these areas, these estimates are unreliable and are not available at the Ottoman district level. Therefore, in our main analysis, we drop all modern districts that were mapped to Ottoman locations with incomplete Census counts, rather than using unreliable estimates for uncounted tribes and making heroic assumptions about how they were distributed across Ottoman districts within a given sancak/vilayet. Otherwise, the population figures in these statistics were considered definitive and reliable, and were used as a basis for official statistics concerning the Ottoman population and for subsequent administrative measures." Armenian share in Ottoman district d is computed as

$$\text{Armenian Share}_d = \frac{\text{Armenian Population}_d}{\text{Total Population}_d} \tag{3}$$

Greek population share is calculated likewise. To each modern Turkish district i, we assigned the Ottoman district $d(i)$ that contained or largely overlapped with i. Therefore, historical Armenian/Greek share district i was exposed to is given by $\text{Armenian Share}_{d(i)}$.

Mapping Ottoman kazas (districts) listed in the 1893 Census onto modern Turkish districts is a challenging task because historical maps contain no information about the boundaries of kazas. This makes it impossible to employ spatial mapping techniques. Instead, we match Ottoman kazas with Turkish districts by name, based on the Ottoman location names listed in Sezen (2006). This source documents how the administrative status and classification of each location evolved from the early Ottoman period until the current administrative units of the Turkish Republic. This information allows us to search for the

name of modern districts (*ilçe*) and identify which Ottoman kaza they used to belong. In most of the cases, an Ottoman kaza is either matched with a single modern district or with multiple modern districts, as an Ottoman kaza is usually geographically larger than a modern district. For some modern districts, especially those that are established during the Turkish Republic in areas where there was no settlement during the Ottoman period, it was not possible to identify the kaza or sancak that contains these areas. For these districts, we relied on other sources (mainly web sites of the local state administrations and municipalities) offering information about the history of the district, including where in the Ottoman administrative hierarchy it was. Figure A.2 presents the distributions of Armenian and Greek populations across modern districts in year 2000.

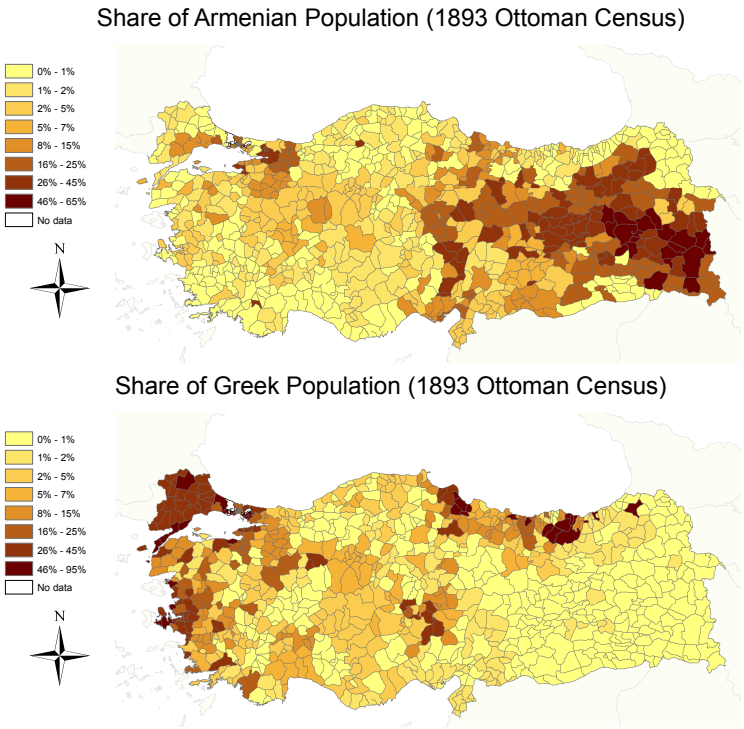


FIGURE A.2: Minority Shares in the late 19th century Ottoman Empire

Estimated Minority Population Shares: In Table B.7, we include in our sample those regions with incomplete population counts, and for Ottoman districts in these regions we impute estimated population shares as follows: For each Ottoman province or sancak with

incomplete census counts, first we take the estimated (but admittedly unreliable) figure for uncounted population as reported by the census authorities. Records suggest that almost all these uncounted groups are nomadic Muslim tribes. Therefore, we distribute the uncounted population to each Ottoman district in that given province or sancak based on what fraction of the Muslim province/sancak population lived in that particular district. Formally, the estimated Armenian share in such a district is computed as

$$\text{Est. Arm. Share}_d = \frac{\text{Arm. Pop.}_d}{\text{Total Pop.}_d + \text{Est. Uncounted Pop.}_{p(d)} \left(\frac{\text{Muslim Pop.}_d}{\text{Muslim Pop.}_{p(d)}} \right)} \tag{4}$$

where $p(d)$ denotes the Ottoman province or sancak where district d is located. Estimated Greek share is calculated likewise.

Longitude and Latitude: The latitude and longitude of the district centers in degrees. The values are retrieved via the GPS Visualizer's address locator web application and utilizing the Bing Maps database on location names and coordinates.

Average Elevation and Standard Deviation of Elevation: Average elevation and the standard deviation of elevation in a given district. Raw elevation data are downloaded from DIVA-GIS in Grid format. These data are a version of the CGIAR SRTM dataset (originally provided at 3 seconds resolution) aggregated to 30 seconds resolution. Average elevation is simply the mean of the values corresponding to those elevation grids that fall within a given district. Standard deviation of elevation is the standard deviation across the same grids. They are computed using ArcGIS®software. The original variable is given in meters. In regressions we rescale average elevation and standard deviation of elevation by dividing these measures by 1000.

Lake: A dummy which takes a value of 1 if a lake overlaps (partly or partially) with the territory of the district and 0 otherwise. The shapefile is downloaded from DIVA-GIS in vector format. The primary source is the Digital Chart of the World. The spatial computations are made using ArcGIS®software.

Sea: A dummy which takes a value of 1 if a district is adjacent (touches) a sea body, i.e., Marmara Sea, Black Sea, Mediterranean Sea or Aegean Sea, and 0 otherwise. The shapefile is downloaded from DIVA-GIS in vector format. The primary source is the GADM database of Global Administrative Areas (version 1.0). The spatial computations are made using ArcGIS®software.

Major River: A dummy which takes a value of 1 if a major river goes through any part of the district territory and 0 otherwise. The shapefile for rivers is downloaded from DIVA-GIS in vector format. The primary source is the Digital Chart of the World. Major rivers are spatially selected by cross-checking with Turkey's Map of Rivers and Lakes created by Ramazan Saygili. The spatial computations are made using ArcGIS®software.

Mean Annual Temperature and Precipitation: These variables show average annual temperature and precipitation over the period 1960-1990. Data for these climatic indicators are retrieved from GAEZ data portal and they are provided at the grid cell level. We compute averages across cells that fall within modern district boundaries using ArcGIS®software. Finally, we rescale the resulting averages by dividing by 1000.

Suitability for Cultivation of the Crop with Greatest Potential: This is a combined measure of suitability for main agricultural products in Turkey. It shows the maximum value of the indices of suitability for cultivation among the following eight crops that historically dominate agricultural production of Turkey: sugar beet, wheat, barley, olive, tobacco, potato, cotton, tea. The suitability data for these crops are borrowed from GAEZ data portal. We use crop suitability indices that are estimated for low input level rain-fed cereals. The index for each crop is provided for individual grid cells with values ranging between 0 and 10 000. We compute averages across the grid cells that fall within modern district boundaries using ArcGIS®software. Finally, we rescale the resulting averages by dividing by 1000.

Distance to Railroad in 1910: The logarithm of distance of a district (in kilometers) to the nearest railroad in 1910. The image file showing the Anatolian railroads in 1910 is downloaded here and digitized using ArcGIS®software. Distance calculations are also made using ArcGIS®.

Distance to Major 19th Century Port: The logarithm of distance of a district (in kilometers) to the nearest major 19th century port. The following were the major ports of the 19th century: The ports of Constantinople (Istanbul), Izmir (Smyrna), Samsun, Trabzon, Mersin, and Iskenderun (Alexandretta). The spatial computations are made using ArcGIS®software.

Log Distance to War Front (1919-1922): Logarithm of distance to the nearest war front during the Turkish War of Independence that took place during the period 1919-1922. The spatial computations are made using ArcGIS®software.

Log WWI Soldier Casualty: Logarithm of total number of soldiers in the Ottoman Army who died in battle during the First World War and whose birth province contains the

district in question. The casualty data are retrieved from the List of Martyrs provided by the Turkish Ministry of National Defense.

Share of immigrants in province who arrived and were settled during 1921-1929: The number of immigrants in a given province in Turkey who settled during 1921-1929 period. We divide this number by 1927 province population. *Source:* Turkish Statistical Yearbook, 1930, Vol. 3.

Shares of 1927 population born outside modern-day Turkey: These four variables measure, at the district level, the shares of 1927 Census population whose birth place is (i) Albania, (ii) Greece, (iii) Romania and (iv) other places outside modern-day Turkey. *Source:* 1927 Population Census of the Republic of Turkey.

Share of Kurdish Speakers in province in 1927: Source data give us the number of Kurdish speakers in the total population in a given province in 1927. These data come from the 1927 population Census of the Republic of Turkey.

Central kaza/sancak dummy: A dummy which takes a value of 1 if the district in question is matched either to the central kaza (the central Ottoman district) of a sancak or to the central sancak of a vilayet (Ottoman province) –the latter applies only for those vilayets which only have sancak subdivisions. The variable captures the location of historical economic centers and more urbanized places.

Distance to Istanbul: The logarithm of distance of a district (in kilometers) to Istanbul. The spatial computations are made using ArcGIS®software. The shapefile is downloaded from DIVA-GIS in vector format. The primary source is the GADM database of Global Administrative Areas (version 1.0). The spatial computations are made using ArcGIS®software.

Distance to Nearest National Border: The logarithm of distance of a district (in kilometers) to the nearest modern border of Turkey with any of its neighbors. The spatial computations are made using ArcGIS®software. The shapefile is downloaded from DIVA-GIS in vector format. The primary source is the GADM database of Global Administrative Areas (version 1.0). The spatial computations are made using ArcGIS®software.

Distance to Anatolian Silk Road: The logarithm of distance of a district (in kilometers) to the nearest Anatolian Silk Road. The spatial computations are made using ArcGIS®software. The primary source is the Old World Trade Routes (OWTRAD) Project.

Distance to Ottoman Trade Routes: The logarithm of distance of a district (in kilometers) to the nearest Ottoman Trade route. The spatial computations are made using

ArcGIS®software. The primary source is the Old World Trade Routes (OWTRAD) Project.

Settlement from Neolithic period: Dummy for the presence of at least one pre-historic settlement within the district boundary that is dated by the archaeologists to the Neolithic period in Anatolia. *Source:* Databases of the Archaeological Settlements of Turkey Project (TAY). The database for Neolithic settlements can be accessed through this link.

Ancient Greek Site: Dummy for the presence of at least one ancient Greek site within the district boundary. These sites are either dated to the Classical (480-323 BC) or the Hellenistic (323-146 BC) period. The spatial computations are made using ArcGIS®software. The primary source is the Ancient-Greece project.

Distance to the Capital Tushpa (Van) of the Ancient Armenian Kingdom of Urartu (860-590 BC): The logarithm of distance of a district (in kilometers) to Tushpa (Van) which was the capital of the ancient Armenian Kingdom of Urartu (860-590 BC). The spatial computations are made using ArcGIS®software.

Literacy rate in 1927: The share of 1927 population in the district who is recorded as literate. *Source:* 1927 Census of the Turkish Republic.

High School completion rate in 2000: We use 2000 Census figures to compute district level high school completion rates among the population aged 6 and above. When computing the completion rates we exclude from this base population those respondents whose education status is unknown. High school graduates consist of those who completed either a high school or a vocational school that is equivalent to a high school. Census results can be accessed through TurkStat's web application.

Number of Armenian and Greek School Buildings per Muslim in 1893: Ratio between the number of Armenian or Greek school buildings within a district as of 1912 and the number of Muslims in that district in 1893. The spatial computations are made using ArcGIS®software. The primary source for minority school buildings is the Cultural Heritage Map of Turkey which was created as part of the Hrant Dink Foundation's cultural heritage inventory project. The project documents and maps Armenian, Greek, Syriac, and Jewish community buildings in Turkey. Although the data is the most comprehensive inventory of minority buildings to date, it is not fully complete. For example, Greek buildings are relatively under-documented compared to Armenian buildings.

Land Concentration in 1997: Source data for our measure of land inequality come from 1997 Village Inventory published by State Institute of Statistics of the Turkish Republic. This inventory provides various statistics based on the information collected from all localities with

a village status and aggregated to the district level. We use data on land plots owned by households that are grouped into 11 land size brackets. For each district, we have information about (i) number of individual land plots, (ii) total size of land plots, and (iii) number of land owning households that fall into each land size bracket. Since we do not have household level data on the size of land holdings, we approximate a Lorenz curve for land size distribution among households under two strong but inevitable assumptions. We assume that (a) within each land size bracket, total land is distributed equally among households in that bracket, and (b) treat households owning multiple land plots as separate households. Under these assumptions, we can order households by the size of their land holdings into 11 categories (k=1,2,...,11) and compute for each district (1) the cumulative share of households X(k) that fall into a land size category of k or below, and (2) the cumulative share Y(k) of total land (in terms of size) in the district that is owned by these households such that X(11) = 1 and Y(11) = 1. Using this Lorenz curve for discrete categories of households, we approximate an index, akin to Gini coefficient, for land distribution that is equal to $G = 1 - \sum_{j=1}^{11}[X(j) - X(j-1)] * [Y(j) + Y(j-1)]$ where X(0) = Y(0) = 0.

Occupational groups in the Ottoman Population Registers of the 1840s for Ankara, Bursa and Manisa: Registers were digitized and organized by Mustafa Erdem Kabadayi, a historian of late Ottoman period. They cover the entire universe of male household members in these cities and record occupational status and the type of occupation each male household member holds (if any) along with ethnoreligious affiliation. We pool all individuals with an occupation and whose occupation could be matched to the PST classification. The original occupational codings used in the Ottoman population surveys were standardized using the PST and PSTI occupational coding scheme (https://www.campop.geog.cam.ac.uk/research/occupations/datasets/coding/). Table B.2 presents the overall occupational composition of the workforce.

Modern Occupational Structure in 2000: We use the 5% Micro Sample from the 2000 Turkish Census data. Among the several individual characteristics of the respondents in the Census micro sample, we observe which of the 80 occupation categories they belong to based on both their self-reported current occupation and self-reported actual profession. The two coincide most of the time, but not always. We use the latter to compute district level occupational distribution in the population because profession captures the type of professional skills a person has acquired through education, training or experience. As such, the representation of different groups of professions in the district population reflects the type of human capital residents of the district possesses. Using profession instead of current occupation has the additional advantage that the former can be observed even for

respondents who were unemployed or out of the labor force during the reference period of the Census. The 2000 Census classifies professions into 9 broad categories which is further divided into 80 sub-categories. For various profession categories that are relevant for our analysis, we compute the share of people in the relevant population who hold a profession in that category. The relevant population consists of citizens of Turkey with age 12 or above, who were born in the province where they currently reside, who have been resident in the same district since at least 1995 (five years prior to the Census) and who declared to have a profession.

Contemporary Trust: These data come from Konda Survey Company. The sample contains individual respondents pooled from the two waves of KONDA Barometre Survey (September 2012 and September 2015) which contain questions on trust. The pooled sample covers 33 out of 81 provinces but the sample in each wave is representative at the national level. Generalized trust is a dummy variable which is equal to 1 if the respondent answered "Most people can be trusted" to the question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?". Answers to other trust questions range between 1 and 5 on a Likert scale, where 1 stands for "Not trust at all" and 5 stands for "Full trust". Inter-ethnic marriage measures whether it is acceptable for the respondent or his/her son or daughter to marry someone from another ethnicity. It also ranges between 1 and 5 where 1 stands for "Definitely not acceptable" and 5 stands for "Definitely acceptable".

B Appendix for Supplementary Figures and Tables

(A) Armenian Homeland throughout History

(B) The Byzantine Empire by 1025

FIGURE B.1: Armenian and Greek Homelands

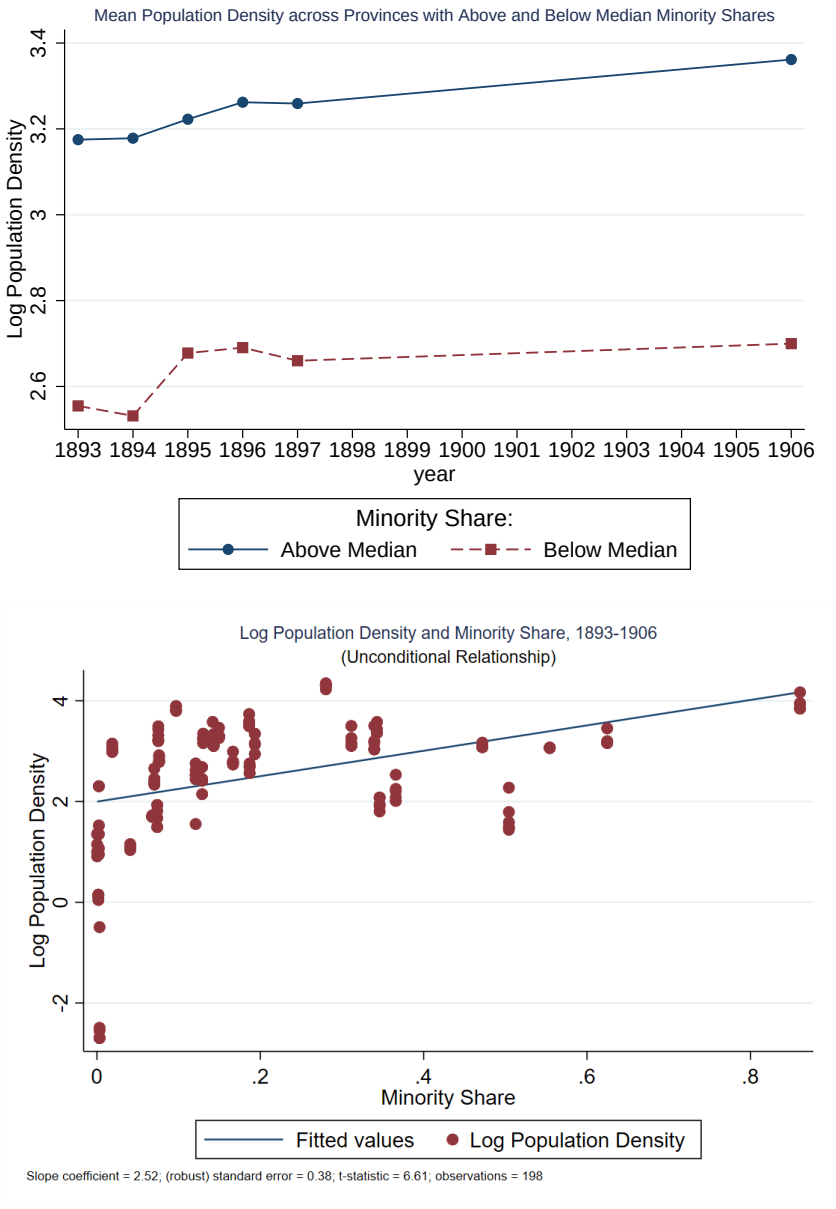


FIGURE B.2: Minority Presence and Population Density in Ottoman Provinces

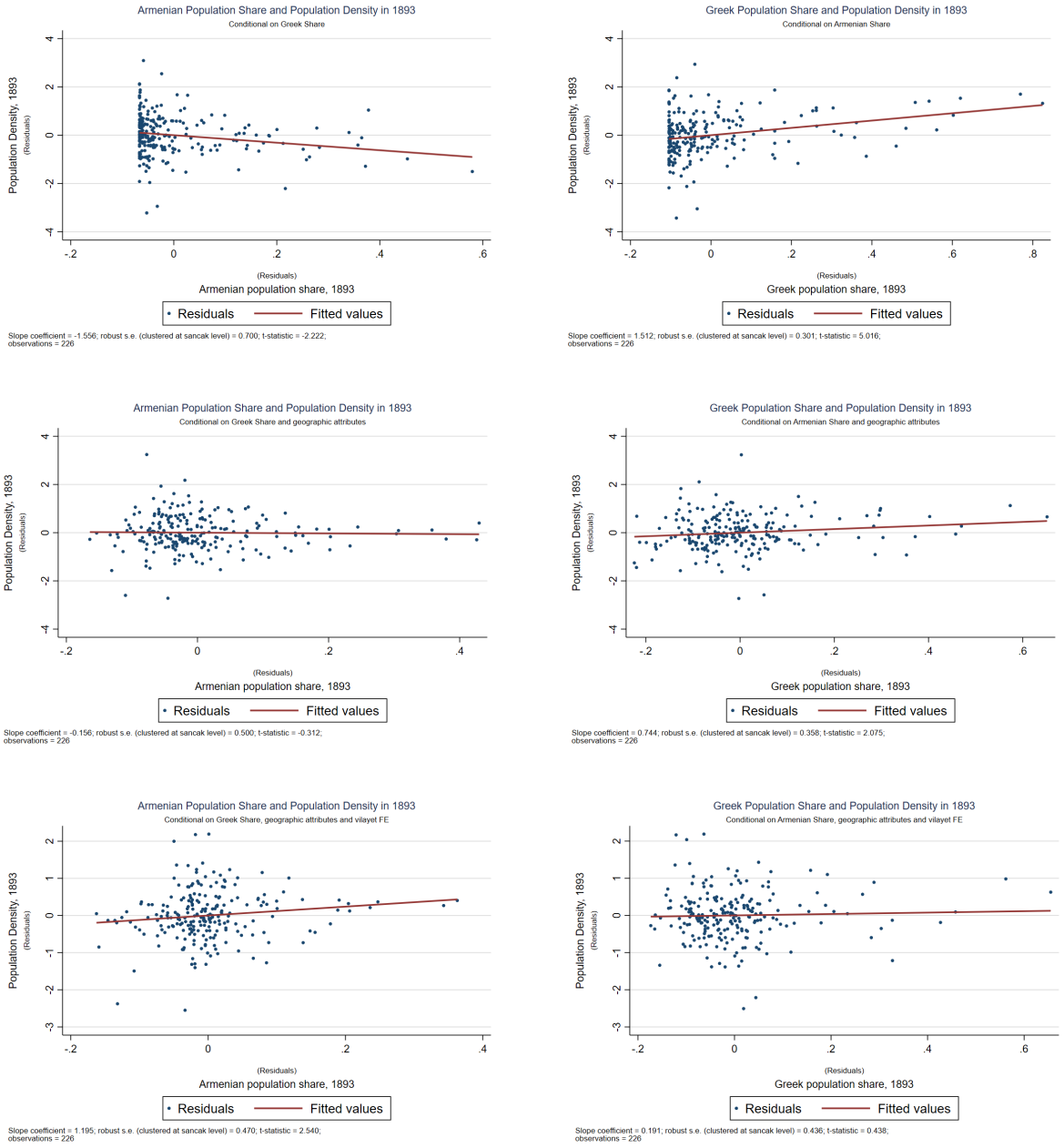


FIGURE B.3: Pre-expulsion conditions: Minority shares and population density in 1893

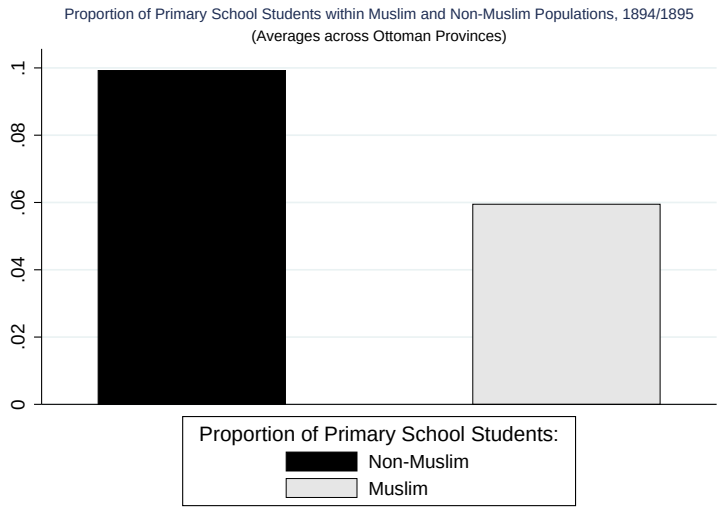


FIGURE B.4: Average Minority-Muslim Education Gap in Ottoman Provinces, 1894/1895

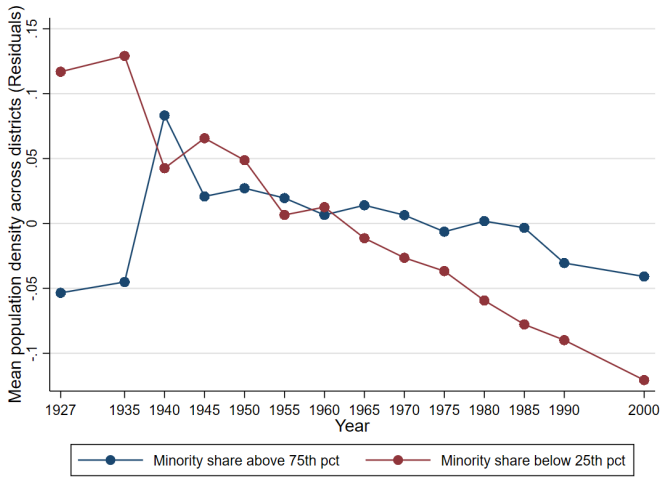
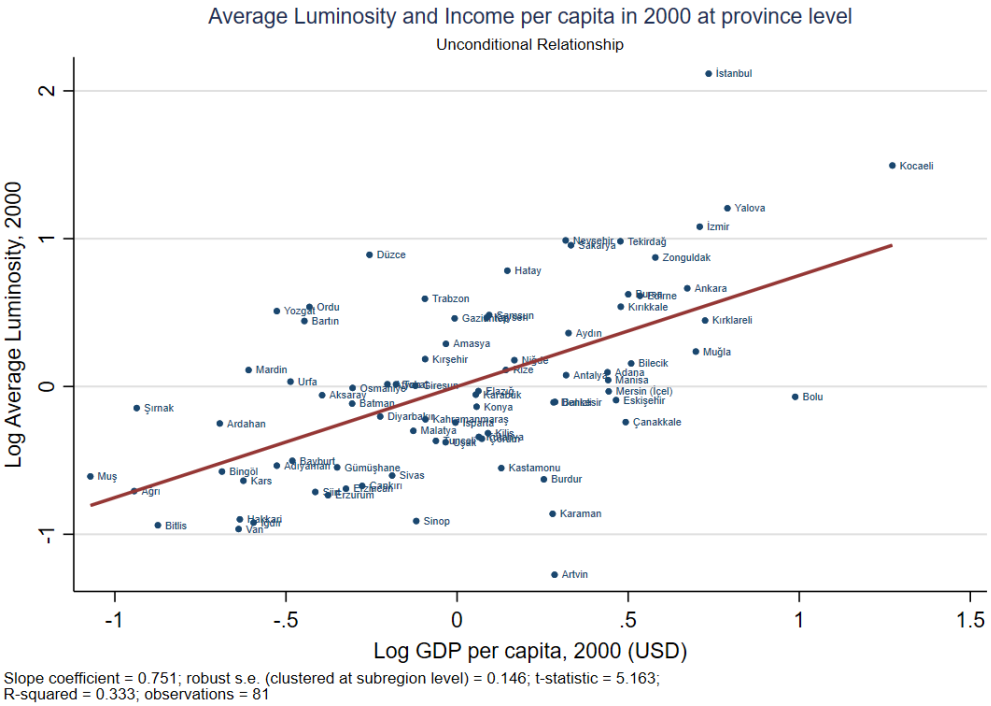


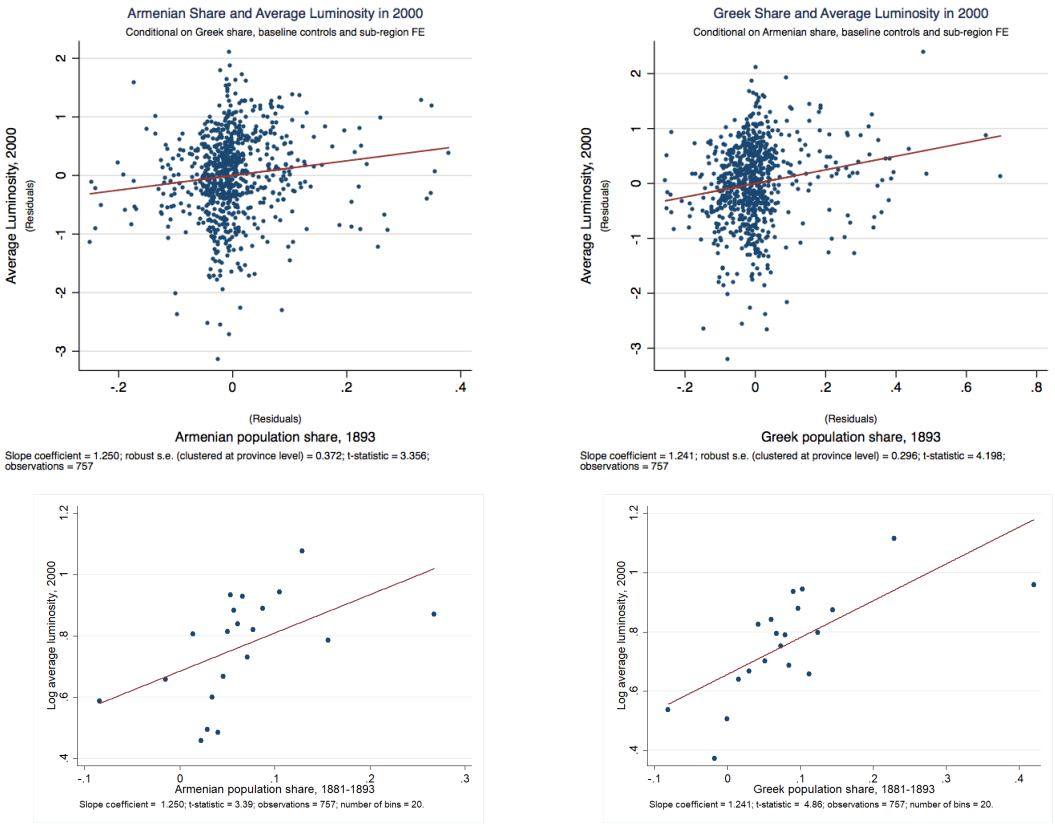
FIGURE B.5: Post-expulsion Population Trends After Accounting for District and Year Fixed Effects, 1927-2000

Notes: This figure plots for each year the average residuals from a regression of log population density in a given district-year on district and year fixed effects for two groups of modern Turkish districts: Those where the sum of historical population shares of Armenians and Greeks (i.e. minority share) is above the 75th percentile of the corresponding cross-district distribution, and those where it is below the 25th percentile. The short-term effect of expulsions is more visible when district and year specific factors are accounted for: high-minority regions were on average less densely populated in the 1930s, but there was a fast catch-up over the next two decades. Starting with the 1960s, the part of population density in high-minority districts that is unexplained by district and year fixed effects visibly diverges from that in low-minority districts. This is consistent with the idea that minority legacy has a time-varying component, and its positive effect has become more visible over the years.



Slope coefficient = 0.751; robust s.e. (clustered at subregion level) = 0.146; t-statistic = 5.163;
R-squared = 0.333; observations = 81

FIGURE B.6: The relationship between province income and luminosity



Notes: The figures on the top row present residual scatter plots for the conditional relationship between average intensity of nighttime lights in 2000 and minority shares. The figures on the bottom row present binned scatter plots showing the same conditional relationships. The conditioning variables are baseline geographic controls, population density in 1927, subregion FEs and the population share of the other group in 1893. The plots are based on binned averages (for each of 20 bins) of average luminosity and population share of the indicated minority group.

FIGURE B.7: Historical minority shares and Average Luminosity in 2000

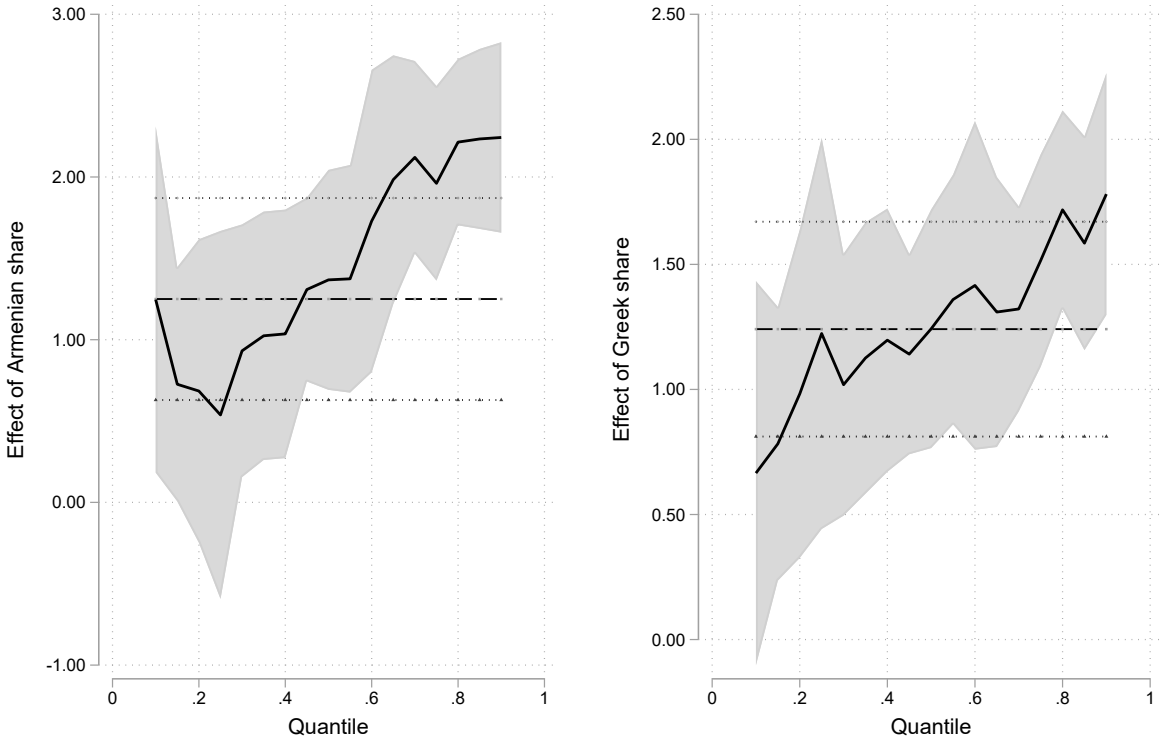


FIGURE B.8: Effect of Armenian and Greek shares on luminosity at different quantiles

Notes: This figure plots coefficients of Armenian and Greek shares at various quantiles (from 0.1 to 0.9), from quantile regressions of luminosity in 2000 based on the baseline specification (controlling for past population density, geographic characteristics, and subregion fixed effects), together with 90% confidence intervals (shaded). Solid line represents the quantile regression coefficients, whereas the dashed line is the OLS coefficient (with the dotted lines representing confidence intervals).

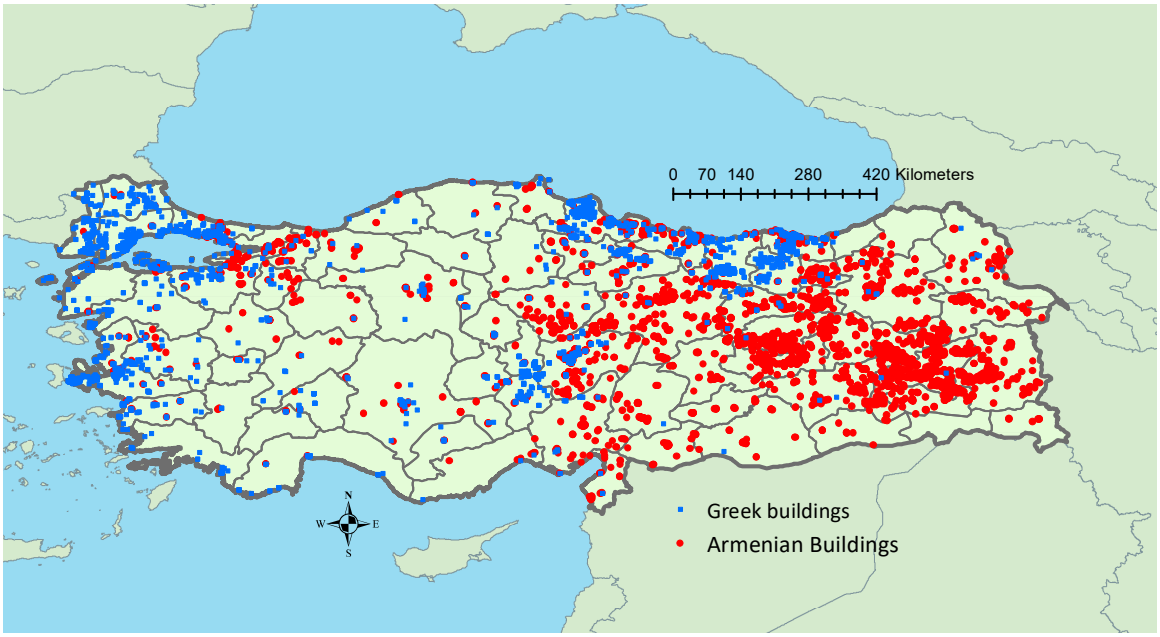


FIGURE B.9: Distribution of Old Community Buildings in Turkey

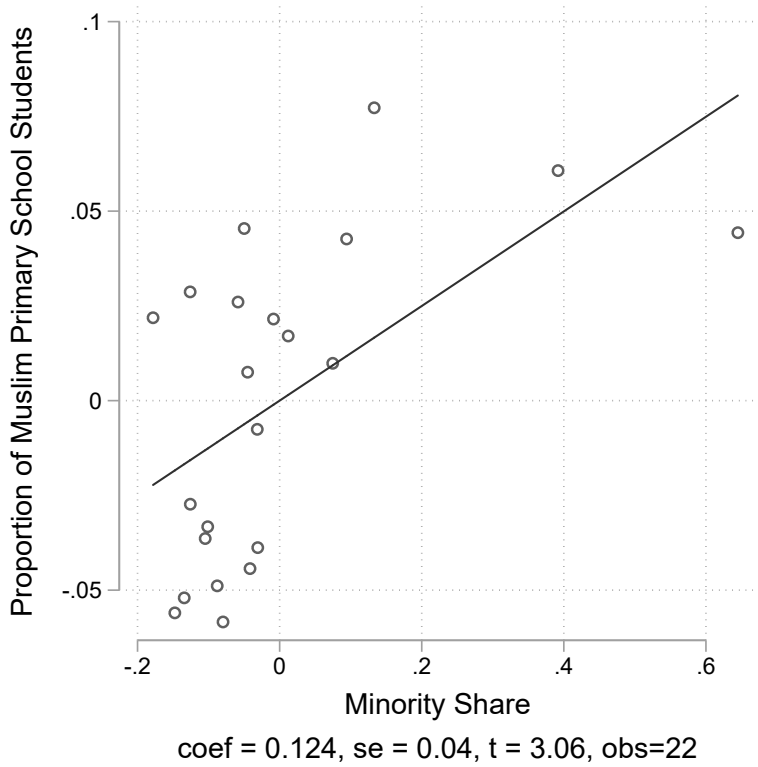


FIGURE B.10: Educational attainment among Muslims in Ottoman Provinces and Minority Shares, 1894/1895

TABLE B.1: Occupational groups and sub-groups in which minorities were over-represented according to Ottoman Population Registers of the 1840s

Panel A: Armenian over-representation	Degree of over- or under-representation of the group:				Armenian rep.> Muslim rep.	Armenians dominant
	Armenian	Jewish	Muslim	Greek		
Level of aggregation:						
(1) Occupation group:						
Dealers	23.1%	0.1%	-22.5%	-0.6%	Yes	Yes
Construction	12.4%	-2.3%	-38.5%	28.5%	Yes	No
Manufacturing	4.2%	-0.6%	-4.5%	0.8%	Yes	No
(2) Occupation sub-group:						
Construction						
Building and construction	15.5%	-2.4%	-45.6%	32.5%	Yes	No
Dealers						
Dealers in textiles and products	39.7%	-3.1%	-33.6%	-3.0%	Yes	Yes
Not further specified	14.3%	3.9%	-22.4%	4.2%	Yes	Yes
Manufacturing						
Instrument making	34.9%	-3.1%	-15.6%	-16.1%	Yes	Yes
Clothing	31.7%	1.1%	-36.7%	4.0%	Yes	Yes
Metal working	27.7%	-3.1%	-33.4%	8.9%	Yes	Yes
Machines and tools, making and operation	20.4%	0.8%	-12.8%	-10.3%	Yes	No
Precious metals and jewelry	19.2%	-2.6%	-54.0%	37.6%	Yes	No
Textiles	9.7%	-0.1%	-11.5%	2.0%	Yes	No
Footwear	0.5%	-3.0%	9.8%	-7.2%	No	No
Sellers						
Not further specified	61.0%	-3.1%	-58.4%	0.6%	Yes	Yes
Sellers of clothing and clothing accessories	52.3%	18.7%	-56.6%	-14.3%	Yes	Yes
Services and Professions						No
Medical profession	18.7%	-3.1%	-39.0%	23.5%	Yes	Yes
Miscellaneous service industries	2.5%	-2.9%	9.0%	-8.5%	No	No
Transport and Communications						No
Communications	2.7%	-3.1%	16.6%	-16.1%	No	No

Panel B: Greek over-representation	Degree of over- or under-representation of the group:				Greek rep. > Muslim rep.	Greeks dominant
	Greek	Armenian	Jewish	Muslim		
Level of aggregation:						
(1) Occupation group:						
Construction	28.5%	12.4%	-2.3%	-38.5%	Yes	Yes
Agriculture	9.1%	-10.8%	-3.1%	4.9%	Yes	No
Fishing	5.1%	-1.1%	21.1%	-25.1%	Yes	No
Manufacturing	0.8%	4.2%	-0.6%	-4.5%	Yes	No
(2) Occupation sub-group:						
Agriculture						
Agriculture	9.1%	-10.8%	-3.1%	4.9%	Yes	No
Construction						
Building and construction	32.5%	15.5%	-2.4%	-45.6%	Yes	Yes
Dealers						
Not further specified	4.2%	14.3%	3.9%	-22.4%	Yes	No
Fishing						
Not further specified	5.1%	-1.1%	21.1%	-25.1%	Yes	No
Manufacturing						
Brick and tile manufacture	67.2%	-22.3%	-3.1%	-41.7%	Yes	Yes
Precious metals and jewelry	37.6%	19.2%	-2.6%	-54.0%	Yes	Yes
Earthenware, pottery manufacture	29.4%	-22.3%	-3.1%	-3.9%	Yes	No
Non-ferrous metal manufacture and products	15.5%	-10.0%	5.6%	-11.0%	Yes	No
Metal working	8.9%	27.7%	-3.1%	-33.4%	Yes	No
Food industries	8.1%	-13.7%	0.0%	5.7%	Yes	No
Tobacco industries	7.1%	-8.4%	-1.7%	3.2%	Yes	No
Clothing	4.0%	31.7%	1.1%	-36.7%	Yes	No
Textiles	2.0%	9.7%	-0.1%	-11.5%	Yes	No
Sellers						
Sellers of food	18.6%	-17.7%	0.4%	-1.2%	Yes	No
Not further specified	0.6%	61.0%	-3.1%	-58.4%	Yes	No
Services and Professions						
Medical profession	23.5%	18.7%	-3.1%	-39.0%	Yes	Yes
Food, drink and accommodation services	13.9%	-10.2%	-1.7%	-1.8%	Yes	No

Sample: Men with a declared occupation in the cities of Ankara, Bursa and Manisa.

Notes: The sample includes the entire universe of male household members in the cities of Ankara, Bursa and Manisa in Turkey who declared an occupation. The listed occupation categories are standardized by the UrbanOccupationsOETR project using the PST and PSTI occupational coding schemes developed by the Cambridge Group for the History of Population and Social Structure (https://www.campop.geog.cam.ac.uk/research/occupations/datasets/coding).

TABLE B.2: Size of total male workforce by occupation groups in the pooled sample of three Anatolian cities (Ankara, Bursa and Manisa)

Occupation group	Size of male workforce	% of male workforce	Occupation group	Size of male workforce	% of male workforce
Manufacturing	7176	50.09%	**Dealers & Sellers**	1948	13.60%
Textiles	2476	17.28%	Dealers in textiles and products	366	2.55%
Clothing	1023	7.14%	Not further specified	355	2.48%
Footwear	866	6.04%	Dealers in live animals	33	0.23%
Food industries	686	4.79%	Dealers in clothing and clothing accessories	17	0.12%
Industries using leather, bone etc.	442	3.09%	Dealers in footwear	10	0.07%
Iron and steel manufacture and products	401	2.80%	Dealers in food	2	0.01%
Tobacco industries	216	1.51%	Dealers in iron and steel, and iron and steel products	2	0.01%
Machines and tools, making and operation	206	1.44%	Dealers in wood and wood products	2	0.01%
Precious metals and jewelry	205	1.43%	Small traders	539	3.76%
Non-ferrous metal manufacture and products	196	1.37%	Sellers of food	395	2.76%
Wood industries	195	1.36%	Sellers of chemical products	108	0.75%
Minor manufactures and trades	45	0.31%	Sellers of clothing and clothing accessories	55	0.38%
Drink industries	40	0.28%	Sellers of fuel	18	0.13%
Earthenware, pottery manufacture	33	0.23%	Sellers of printed products	17	0.12%
Chemical, soap, adhesives, manufacture	30	0.21%	Sellers of tobacco	12	0.08%
Industries producing products from fibres	27	0.19%	Sellers of iron and steel products	11	0.08%
Instrument making	21	0.15%	Not further specified	6	0.04%
Printing	20	0.14%	**Agriculture**	1105	7.71%
Manufacturing not further specified	19	0.13%			
Furnishing	18	0.13%	**Transport and Communications**	674	4.70%
Brick and tile manufacture	6	0.04%	Road transport (animal power)	349	2.44%
Metal working	4	0.03%	Transport and Communications (Not further specified)	308	2.15%
Paper industries	1	0.01%	Communications	8	0.06%
Services and Professions	2588	18.06%	Inland navigation	8	0.06%
Professions	618	4.31%	Sea transport	1	0.01%
Miscellaneous service industries	605	4.22%	**Construction**	608	4.24%
Food, drink and accommodation services	440	3.07%	Building and construction	545	3.80%
Domestic service	330	2.30%	Public Works	63	0.44%
Local government service	211	1.47%	**Forestry**	188	1.31%
Armed forces	115	0.80%	**Fishing**	33	0.23%
Commercial and administrative services	87	0.61%	**Mining and quarrying**	7	0.05%
Professional support	65	0.45%			
Entertainment	47	0.33%	**Grand Total**	14327	100.00%
Services and Professions(Not further specified)	28	0.20%			
Financial services and professions	21	0.15%			
National government service	21	0.15%			

TABLE B.3: Summary statistics for variables used in district-level analysis

Variable	Obs.	Mean	Std.	Min	Max	Percentile 10th	Percentile 90th
Outcomes							
Log population density, 1881-1893	757	2.58	0.99	-0.37	6.02	1.38	3.67
Log population density, 1927	757	3.04	0.68	0.14	5.21	2.24	3.77
Log population density, 2000	757	4.11	1.06	1.71	9.34	2.97	5.39
Urbanization rate, 2000	757	0.46	0.20	0.09	1.00	0.22	0.75
Log(0.01+Average Luminosity, 2000)	757	0.76	1.05	-3.43	4.00	-0.55	1.98
Log(0.01+Average Luminosity, 2007-2013)	757	1.11	0.99	-2.83	4.07	-0.06	2.34
Literacy rate, 1927	757	0.07	0.06	0.01	0.32	0.02	0.12
High school competion rate, 2000	757	0.16	0.07	0.05	0.54	0.09	0.25
Land Concentration, 1997	751	0.46	0.11	0.02	0.95	0.34	0.60
Variables of interest							
Armenian population share, 1881-1893	757	0.06	0.11	0.00	0.65	0.00	0.20
Greek population share, 1881-1893	757	0.09	0.14	0.00	0.93	0.00	0.26
# Armenian school buildings per Muslim in 1893 (x1000)	757	0.03	0.08	0.00	0.88	0.00	0.07
# Greek school buildings per Muslim in 1893 (x1000)	757	0.17	3.24	0.00	88.89	0.00	0.09
# Armenian community buildings per 1935 population (x1000)	757	0.08	0.25	0.00	2.97	0.00	0.18
# Greek community buildings per 1935 population (x1000)	757	0.15	0.82	0.00	15.06	0.00	0.21
Spatially-weighted Muslim exposure to Armenian community (x100)	757	0.08	0.37	0.00	4.88	0.00	0.09
Spatially-weighted Muslim exposure to Greek community (x100)	757	0.12	0.77	0.00	12.67	0.00	0.09
Geographical controls							
Longitude	757	33.40	4.38	25.91	44.17	27.70	39.66
Latitude	757	39.31	1.48	36.08	42.02	37.34	41.16
Mean elevation/1000	757	0.90	0.50	0.00	2.55	0.21	1.51
Standard deviation of elevation/1000	757	0.26	0.15	0.01	0.95	0.09	0.44
Lake	757	0.29	0.45	0.00	1.00	0.00	1.00
Sea	757	0.22	0.41	0.00	1.00	0.00	1.00
Major river	757	0.28	0.45	0.00	1.00	0.00	1.00
Average annual temperature/1000	757	0.01	0.00	0.01	0.02	0.01	0.02
Average annual precipitation/1000	757	0.64	0.20	0.33	1.73	0.44	0.88
Suitability to cultivation of the crop with greatest potential/1000	757	4.19	1.66	0.24	9.34	2.04	6.44
Robustness and other controls							
Log Distance to Railroad in 1910	757	10.93	1.55	3.98	12.99	8.86	12.63
Log Distance to Major 19th century port	757	5.03	0.80	0.00	6.24	4.11	5.74
Log Distance (km) to war front, 1919-1922	757	4.15	1.28	-0.99	5.85	2.43	5.53
Log WW1 soldier casualty from province	757	7.07	1.02	1.39	8.37	5.99	8.22
Share of Kurdish speakers in 1927 province	757	7.55	18.35	0.00	88.94	0.02	41.81
Share of immigrants (1921-1929) in 1927 province	757	0.04	0.07	0.00	0.60	0.00	0.09
Share of 1927 population born in Albania	757	0.00	0.00	0.00	0.00	0.00	0.00
Share of 1927 population born in Greece	757	0.01	0.02	0.00	0.36	0.00	0.01
Share of 1927 population born in Romania	757	0.00	0.00	0.00	0.07	0.00	0.00
Share of other 1927 population born outside Turkey	757	0.02	0.06	0.00	0.64	0.00	0.06
Central kaza/sancak	757	0.42	0.49	0.00	1.00	0.00	1.00
Log Distance to Istanbul	757	6.04	0.67	3.52	7.19	5.11	6.84
Log Distance to nearest national border	757	5.19	0.87	1.66	6.27	4.00	6.07
Log Distance to Ottoman Trade Routes	757	10.28	0.91	7.15	11.95	9.01	11.40
Log Distance to Anatolian Silk Road	757	10.10	0.88	7.30	11.77	8.88	11.23
Any settlement from Neolithic period?	757	0.19	0.39	0.00	1.00	0.00	1.00
Any ancient Greek site?	757	0.17	0.37	0.00	1.00	0.00	1.00
Log Distance to Tushpa (Van)	757	6.66	0.63	0.76	7.32	5.94	7.22
Population share of age<13 in 1927	757	0.33	0.04	0.23	0.44	0.28	0.38
Female-male ratio in 1927	757	1.11	0.14	0.64	1.55	0.94	1.25

TABLE B.4: Summary statistics for variables used in village/neighborhood-level analysis

Variable	Obs.	Mean	Std.	Min	Max	Percentile 10th	Percentile 90th
Outcome variable							
Log Average Luminosity (in 2000) within 5km radius of village/neighborhood	49321	0.49	2.25	-4.61	4.14	-3.30	2.88
Variables of interest							
Armenian community building within 5km	49321	0.22	0.42	0.00	1.00	0.00	1.00
Greek community building within 5km	49321	0.16	0.37	0.00	1.00	0.00	1.00
Muslim religious or school building within 5km	49321	0.12	0.32	0.00	1.00	0.00	1.00
Armenian building between 5 to 7km	49321	0.18	0.38	0.00	1.00	0.00	1.00
Greek building between 5 to 7km	49321	0.13	0.33	0.00	1.00	0.00	1.00
Log Density of Armenian buildings within 5km	11000	-3.25	0.98	-4.36	-0.12	-4.36	-1.80
Log Density of Greek buildings within 5km	8111	-3.07	1.23	-4.36	0.46	-4.36	-1.47
Armenian school building within 5km	49321	0.14	0.34	0.00	1.00	0.00	1.00
Greek school building within 5km	49321	0.09	0.29	0.00	1.00	0.00	0.00
Log(1+# of Armenian non-school community buildings within 5km)	49321	0.27	0.60	0.00	3.87	0.00	1.10
Log(1+# of Armenian non-school community buildings within 5km)	49321	0.22	0.63	0.00	4.64	0.00	0.69
Geographical controls							
Log distance to nearest major river	49321	9.89	1.24	0.04	12.02	8.18	11.18
Log distance to sea coast	49321	11.11	1.52	0.91	13.22	9.03	12.54
Log distance to lake	49321	9.93	1.08	0.00	11.95	8.70	11.04
Log altitude	49321	4.78	2.96	0.00	13.70	0.00	7.36
Longitude	49321	34.92	4.79	26.07	44.76	28.15	41.47
Latitude	49321	39.25	1.49	35.86	42.09	37.14	41.12
Other controls							
Ottoman city within 15km	49321	0.34	0.47	0.00	1.00	0.00	1.00
Log distance to Ottoman Trade routes (circa 1300-1600 CE)	49321	10.06	1.35	0.74	12.32	8.24	11.42
Log distance to Anatolian Silk Road (circa 1200-1400 CE)	49321	9.65	1.50	0.13	11.98	7.55	11.21
Log distance to nearest modern province center	49321	10.31	1.02	0.00	11.86	9.10	11.21
Log distance to Istanbul	49321	13.17	0.69	10.58	14.21	12.17	13.95
Log distance to Izmir	49321	13.27	0.88	6.49	14.31	12.13	14.09
Log distance to Ankara	49321	12.84	0.69	6.72	13.94	12.07	13.59
Log distance to Bursa	49321	13.06	0.84	5.55	14.20	12.00	13.94
Log distance to modern railroad network	49321	9.91	1.50	0.87	12.10	7.68	11.46
Village dummy	49321	0.37	0.48	0.00	1.00	0.00	1.00

Notes: Summary statistics for log density of Armenian and Greek buildings are reported for the respective regression samples used in Table 3 which exclude localities that do not have any Armenian and Greek community building within 5km radius.

Table B.5: Historical Minority Shares and Population Densities in 1927 and 2000

PANEL A	**Log(Population density in 1927)**								
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS
Armenian population share, 1881-1893	-0.929	-0.746	-0.983	-0.719	-0.744	-0.712	-0.713	-0.739***	-0.566*
	[0.653]	[0.699]	[0.648]	[0.534]	[0.533]	[0.529]	[0.531]	[0.273]	[0.298]
Greek population share, 1881-1893	0.195	-0.344*	-0.348*	-0.548***	-0.530***	-0.500***	-0.505***	-0.517***	-0.522**
	[0.346]	[0.175]	[0.189]	[0.165]	[0.177]	[0.178]	[0.187]	[0.188]	[0.225]
Proxy for population density in 1893	0.283*	0.211	0.203	0.199	0.199	0.197	0.196	0.487***	0.439***
	[0.157]	[0.146]	[0.138]	[0.122]	[0.123]	[0.121]	[0.121]	[0.142]	[0.162]
Effect of increasing Armenian share	-18.431	-14.814	-19.510	-14.276	-14.758	-14.132	-14.155	-14.669***	-11.242*
from the 10-th to the 90-th percentile	[12.951]	[13.864]	[12.870]	[10.591]	[10.574]	[10.493]	[10.532]	[5.422]	[5.920]
Effect of increasing Greek share	6.337	-11.211**	-11.325*	-17.838***	-17.243***	-16.287***	-16.441***	-16.826***	-16.997**
from the 10-th to the 90-th percentile	[11.259]	[5.694]	[6.160]	[5.362]	[5.747]	[5.793]	[6.080]	[6.126]	[7.329]
Observations	309	309	309	309	309	309	309	309	309
Adjusted R-squared	0.246	0.366	0.378	0.461	0.458	0.457	0.455	0.610	0.626

PANEL B	**Log(Population density in 2000)**								
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS
Armenian population share, 1881-1893	1.393***	1.395**	1.481***	1.588***	1.615***	1.524***	1.518***	1.185***	1.024**
	[0.451]	[0.532]	[0.423]	[0.460]	[0.392]	[0.409]	[0.380]	[0.373]	[0.454]
Greek population share, 1881-1893	0.970***	1.325***	1.390***	0.950***	0.712***	0.714***	0.931***	0.816**	0.466
	[0.340]	[0.318]	[0.322]	[0.264]	[0.263]	[0.269]	[0.282]	[0.317]	[0.295]
Log(Population density, 1927)	0.826***	0.918***	0.945***	0.765***	0.756***	0.765***	0.775***	0.814***	0.806***
	[0.098]	[0.103]	[0.096]	[0.105]	[0.096]	[0.092]	[0.095]	[0.090]	[0.110]
Effect of increasing Armenian share	27.637***	27.686***	29.398***	31.521***	32.053***	30.256***	30.129***	23.524***	20.326**
from the 10-th to the 90-th percentile	[8.953]	[10.560]	[8.391]	[9.128]	[7.788]	[8.120]	[7.549]	[7.395]	[9.008]
Effect of increasing Greek share	24.946***	34.053***	35.743***	24.427***	18.309***	18.359***	23.937***	20.973***	11.986
from the 10-th to the 90-th percentile	[8.751]	[8.185]	[8.283]	[6.798]	[6.756]	[6.905]	[7.243]	[8.141]	[7.595]
Observations	757	757	757	757	757	757	757	757	757
Adjusted R-squared	0.304	0.325	0.332	0.421	0.440	0.441	0.454	0.480	0.521
Longitude & Latitude			×	×	×	×	×	×	×
Mean & std. of elevation				×	×	×	×	×	×
Lake, sea and major rivers					×	×	×	×	×
Temperature & Precipitation						×	×	×	×
Suitability to cultivation							×	×	×
Modern region dummies		×	×	×	×	×	×		
Modern subregion dummies								×	
Modern province dummies									×

Notes: This table presents results from the regressions of Log Population Density in 1927 and Log Population Density in 2000 on historical minority shares controlling for past population density, geographic variables, region, subregion or province fixed effects. The estimated effect associated with increasing minority shares from the tenth to the ninetieth percentile of their respective cross-district distributions is expressed in terms of % change in the level of population density. Robust standard errors, clustered at the modern Turkish province (*il*) level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table B.6: Historical Minority Shares and Urbanization Rates in 2000

	Urbanization rate in 2000								
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS
Armenian population share, 1881-1893	0.252***	0.269***	0.344***	0.345***	0.355***	0.392***	0.391***	0.466***	0.486***
	[0.095]	[0.089]	[0.094]	[0.091]	[0.090]	[0.086]	[0.083]	[0.091]	[0.096]
Greek population share, 1881-1893	0.136**	0.205***	0.208***	0.177***	0.174***	0.168***	0.206***	0.194***	0.265***
	[0.061]	[0.066]	[0.063]	[0.065]	[0.065]	[0.063]	[0.064]	[0.061]	[0.074]
Log(Population density, 1927)	0.064***	0.096***	0.099***	0.088***	0.087***	0.084***	0.085***	0.093***	0.105***
	[0.020]	[0.020]	[0.019]	[0.020]	[0.020]	[0.020]	[0.020]	[0.022]	[0.022]
Effect of increasing Armenian share	5.002***	5.346***	6.827***	6.853***	7.040***	7.781***	7.759***	9.257***	9.651***
from the 10-th to the 90-th percentile	[1.883]	[1.761]	[1.870]	[1.800]	[1.782]	[1.701]	[1.656]	[1.812]	[1.899]
Effect of increasing Greek share	3.507**	5.259***	5.342***	4.561***	4.481***	4.325***	5.292***	4.980***	6.801***
from the 10-th to the 90-th percentile	[1.556]	[1.697]	[1.626]	[1.680]	[1.678]	[1.609]	[1.651]	[1.574]	[1.907]
Observations	757	757	757	757	757	757	757	757	757
Adjusted R-squared	0.062	0.102	0.107	0.117	0.118	0.125	0.136	0.151	0.205
Longitude & Latitude			×	×	×	×	×	×	×
Mean & std. of elevation				×	×	×	×	×	×
Lake, sea and major rivers					×	×	×	×	×
Temperature & Precipitation						×	×	×	×
Suitability to cultivation							×	×	×
Modern region dummies		×	×	×	×	×	×		
Modern subregion dummies								×	
Modern province dummies									×

Notes: This table presents results from the regressions of Urbanization Rate in 2000 on historical minority shares controlling for past population density, geographic variables, region, subregion or province fixed effects. The estimated effect associated with increasing minority shares from the tenth to the ninetieth percentile of their respective cross-district distributions is expressed in terms of percentage-point change in the urbanization rates. Robust standard errors, clustered at the modern Turkish province (*il*) level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE B.7: Robustness of Main Results to Using Estimated Minority Shares for Incomplete Census Districts

	(1)	(2)	(3)
	Population Density	Urbanization	Luminosity
	OLS	OLS	OLS
Estimated Armenian share, 1881-1893	1.040***	0.438***	0.945**
	[0.353]	[0.097]	[0.394]
Estimated Greek share, 1881-1893	0.900***	0.186***	1.256***
	[0.299]	[0.060]	[0.274]
Log(Population density, 1927)	0.681***	0.070***	0.561***
	[0.094]	[0.021]	[0.086]
Observations	859	859	859
Baseline controls	Yes	Yes	Yes
Modern subregion dummies	Yes	Yes	Yes
Adjusted R-squared	0.471	0.115	0.488

Notes: This table presents results from the regressions of Log Population Density in 2000, Urbanization Rate in 2000, and Log Average Luminosity in 2000 on estimated historical minority shares for incomplete census regions, controlling for past population density, geographical variables, and subregion fixed effects. Robust standard errors, clustered at the modern Turkish province (*il*) level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Relevance of human capital and asset channels for regional development. In Table B.8, we evaluate the relevance of minority legacy on regional human capital and the potential contribution of minority assets in explaining modern development. To do so, we regress average district-level luminosity between 2007-2013 on minority shares and then introduce the intermediating variables we have proposed in the preceding sections. Our goal is to evaluate whether the predictive power of minority presence on development operates through these variables.[54] The first column just replicates our baseline specification for luminosity.

TABLE B.8: Relevance of Intermediating Channels: Human Capital, Minority Assets and Luminosity

Channel :		Minority Legacy on Human Capital				Role of Minority Assets			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
	Log Average Luminosity, 2007-2013								
Armenian population share, 1881-1893	1.047***	0.881***	0.542*	0.241	0.258		0.923***	0.890***	0.792**
	(0.305)	(0.297)	(0.299)	(0.257)	(0.267)		(0.321)	(0.289)	(0.307)
Greek population share, 1881-1893	1.301***	1.079***	0.765***	0.574**	0.597***		1.274***	1.173***	1.124***
	(0.274)	(0.291)	(0.246)	(0.223)	(0.224)		(0.322)	(0.266)	(0.320)
Literacy rate, 1927		2.538***			−0.771				
		(0.871)			(0.614)				
High school completion rate, 2000			5.108***		0.826				
			(0.530)		(0.589)				
Trade,commerce, entrepreneurs and directors				4.562**	4.433**				
				(1.883)	(1.875)				
Instrument/machine making, STEM, medicine				7.781***	7.261***				
				(0.828)	(1.004)				
Armenian buildings per 1935 population						258.269*	186.056		144.511
						(140.788)	(148.911)		(152.592)
Greek buildings per 1935 population						74.653**	4.009		12.606
						(35.026)	(35.229)		(36.312)
Land Holdings Gini, 1997								1.252***	1.236***
								(0.368)	(0.369)
Baseline geographical controls	×	×	×	×	×	×	×	×	×
Modern sub-region dummies	×	×	×	×	×	×	×	×	×
Observations	751	751	751	751	751	751	751	751	751
Adjusted R^2	0.524	0.535	0.607	0.660	0.661	0.507	0.524	0.539	0.538

Notes: This table presents results from the district level regressions of average luminosity between 2007-2013 on various potential intermediating variables through which minority legacy on regional development operates. We consider two main channels. Columns (1)–(4) employ historical and contemporary educational attainment as intermediating variables to assess the extent to which minority presence mattered for development through its influence on human capital accumulation. Columns (5)–(8) employ number historical minority community buildings per 1935 population and the Land Holdings Concentration for contemporary land holdings as intermediating variables to assess the extent to which minority presence mattered for development through its influence on asset accumulation and land concentration. In all specifications we control for past population density in 1927, geographic baselines controls, and subregion fixed effects. Robust standard errors, clustered at the level of modern Turkish province (il), are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

In columns 2–5, we assess the relevance of the human capital channel. When either literacy rates in 1927 or high school completion rates in 2000 are added on the right-hand side, the minority share coefficients are significantly reduced relative to our benchmark

[54]Educational attainment and land concentration measures on the right-hand side are for the years 2000 and 1997 respectively. Thus, unlike the previous analyses, here we do not use luminosity in 2000 to lessen reverse causality issues.

estimates in column 1. According to Sobel-Goodman mediation tests, literacy rates in 1927 or high school completion rates in 2000 individually explain 16 percent and 48 percent of the conditional correlation between Armenian share and contemporary nighttime lights. Corresponding fractions for Greek share are about 17 percent and 41 percent respectively. Next, we control for district-level prevalence of those occupation groups in which Armenians and Greeks were historically over-represented. Earlier in Table 6, we have shown that some of these occupation groups are indeed more prevalent in districts with greater historical presence of Armenians and Greeks. Now, we probe how the estimated relationship between Greek and Armenian shares change in magnitude and significance, once we account for occupation-based proxies for human capital. In particular, we divide the high-skilled occupation groups reported in Table 6 into two broader categories. First category consists of entrepreneurs, directors, and people whose profession is in trade or commerce. Second category consists of professions in STEM, medicine and those who work in machine or instrument making. Using the 5% micro sample from the 2000 Census, we compute district-level shares of these two profession groups. When we control for these shares in column 4, the sizes of the coefficient estimates for Armenian and Greek shares drop considerably in magnitude (compared to column 1). Armenian share is no longer statistically significant at conventional levels while Greek share remains significant but at 5% level instead of 1%. Sobel-Goodman mediation tests indicate that the prevalence of the first group of high-skill professions (i.e. trade, commerce, entrepreneurs and directors) mediate about 41% and 40% of the total effect of historical Armenian and Greek presence respectively. The corresponding numbers are more substantial for the second group of professions (i.e. STEM, medicine and instrument/machine making). Prevalence of this group mediates around 76% and 51% percent of the total effect of Armenian and Greek presence, respectively. In column 5 we add all mediating human capital proxies simultaneously. Coefficient for Armenian population share remains statistically insignificant. The Greek share is slightly larger and significant at 1% level.

Columns 6–9 explore the relevance of minority assets. Both historical Armenian and Greek community buildings predict significantly higher luminosity. However, this effect disappears after conditioning on minority population shares (see columns 6 and 7). Coefficients on minority shares in column 7 are virtually no different compared to the ones in column 1. At the same time, minority community buildings –as a proxy for the level of minority assets– by itself do not mediate the effect of minority shares on luminosity.

In columns 8 and 9, we assess the mediating role of land concentration, first without conditioning on the level of historical minority assets and then conditional on the latter. When we control for land concentration and minority buildings simultaneously, the estimates

for Armenian and Greek shares are reduced by 24 and 13 percents respectively. Sobel-Goodman tests suggest that the variation in contemporary land inequality mediate only about 15 percent of the total effect of Armenian share and 10 percent of the total effect of Greek share.

In sum, our district-level analysis suggests that the concentration of Armenian (but not Greek) assets and the degree of contemporary land inequality seem to account for some part of the reduced-form effects of minority shares on luminosity. However, these indirect effects are not as important in magnitude as the mediating role of historical or contemporary proxies for human capital.

In Table B.9 we explore the relevance of the human capital channel this time at the level of villages and neighborhoods. We do so by distinguishing between school and non-school community buildings and investigating whether proximity to a Greek or Armenian school building matters for local development above and beyond the legacy of other types of community buildings. Specifically, we analyze whether being situated within 5km of a minority school building translates into higher luminosity once we account for the density of other community buildings.[55] Here our implicit assumption is that conditional on the density of non-school minority buildings and presence of historical mosques and Muslim schools in the vicinity, the presence of a school building near a given locality is an appropriate proxy for the human capital of the minority community near that locality. In the first three columns, we use the entire sample of villages and neighborhoods, while in the last three columns, we limit the analysis to villages only. Regardless of the sample, presence of a school building, be it Greek or Armenian, explains differences in nighttime lights even across locations with similar minority presence, similar historical and contemporary characteristics and are located in the same district (Columns 3 and 6). The estimated coefficients for school dummies are statistically significant at conventional levels. As expected, estimated magnitudes get smaller as we add more controls and district fixed effects. Our most stringent specification suggests that predicted luminosity is about 40 and 23 percent higher around localities that are near an Armenian or Greek school than otherwise. These are economically substantive effects at the local level. The estimates are somewhat larger in the limited sample of villages, but less precisely estimated for Greek schools.

[55] Given that the distribution of the number of buildings is right-skewed, we measure building density by the log transformation of one plus the number of buildings within 5km radius around each location centroid. Since there is higher right skew and much less variation in the number of school buildings at the intensive margin –than non-school buildings– across the locations in our sample, we employ a dummy for school buildings instead of a continuous measure.

TABLE B.9: Minority Schools and Local Development – Village/Neighborhood Level Analysis

Sample:	All neighborhoods and villages			Villages only		
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
	Log Average Luminosity (in 2000) within 5km					
Armenian school building within 5km	1.400***	0.601***	0.337***	0.659***	0.445**	0.354***
	(0.125)	(0.110)	(0.072)	(0.154)	(0.213)	(0.094)
Greek school building within 5km	1.621***	0.688***	0.207**	1.456***	0.814***	0.311*
	(0.186)	(0.116)	(0.103)	(0.124)	(0.231)	(0.184)
Numbers of non-school Armenian & Greek buildings within 5km		×	×		×	×
Historical mosque or Muslim school within 5km		×	×		×	×
Geographical controls			×			×
Near an Ottoman city			×			×
Distance to historical trade routes			×			×
Contemporary controls			×			×
Bucak/District FE			×			×
Observations	49321	49321	49321	18288	18288	18288
Adjusted R^2	0.187	0.230	0.645	0.023	0.043	0.506
Effect of having an Armenian school building within 5km	308.012***	82.870***	40.274***	94.602***	56.882*	43.037***
	(51.090)	(20.253)	(10.201)	(30.302)	(33.804)	(13.645)
Effect of having a Greek school building within 5km	408.779***	99.600***	23.121*	333.501***	127.062**	36.637
	(94.861)	(23.346)	(12.711)	(54.517)	(52.881)	(25.198)

Notes: Dependent variable is log average luminosity (measured in year 2000) within 5km radius around the village/neighborhood centroid. The first three columns include a dummy for village=1. Geographical controls are *distances to nearest major river, sea coast and lake, altitude, longitude and latitude. Near an Ottoman city*=1 if the village/neighborhood is within 15km of an Ottoman city centroid, and zero otherwise. Historical trade routes refer to Anatolian Silk Road (circa 1200-1400 CE) and Trade routes in the Ottoman Empire (circa 1300-1600 CE) as reported in OWTRAD database. Contemporary controls include distance to nearest province center, distance to contemporary railroad network, distances to biggest commercial/industrial centers (Istanbul, Izmir, Ankara and Bursa). Like in district-level analysis, Istanbul Province is omitted from the sample, leaving 80 provinces. Standard errors are clustered at the province level. Marginal effects reported in the bottom two rows show how much (in percentage terms) higher average luminosity within 5km is expected to be if there is at least one minority school building within 5km to the locality as opposed to none.

Table B.10: Historical Minority Presence and Contemporary Trust

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS
	Generalized trust		Trust to someone you just met		Trust to religious outgroups		Trust to ethnic outgroups		Inter-ethnic marriage	
Armenian population share, 1881-1893	−0.102	−0.121	−0.297	−0.560	−0.567	−0.698	−1.164*	−1.198*	0.264	0.722
	(0.088)	(0.093)	(0.360)	(0.363)	(0.813)	(0.954)	(0.601)	(0.697)	(0.717)	(0.734)
Greek population share, 1881-1893	−0.014	0.008	0.268	0.248	−0.017	0.172	−0.029	0.009	−0.349	−0.074
	(0.069)	(0.081)	(0.243)	(0.277)	(0.322)	(0.315)	(0.263)	(0.258)	(0.431)	(0.449)
Log(Population density, 1927)	0.006	0.009	0.050	0.066	−0.054	−0.054	0.025	0.017	−0.030	−0.076
	(0.011)	(0.011)	(0.070)	(0.074)	(0.065)	(0.078)	(0.051)	(0.058)	(0.085)	(0.101)
Baseline geographical controls	×	×	×	×	×	×	×	×	×	×
Modern sub-region dummies	×	×	×	×	×	×	×	×	×	×
Individual characteristics	×	×	×	×	×	×	×	×	×	×
Historical demography in early republic	×	×	×	×	×	×	×	×	×	×
Observations	4272	3002	4241	2975	4200	2945	4217	2961	4178	2934
Adjusted R^2	0.020	0.021	0.064	0.059	0.054	0.055	0.053	0.053	0.095	0.109
Mean of the outcome	0.069	0.071	1.995	1.972	2.392	2.373	2.396	2.374	3.583	3.557
Effect of 10th–90th %ile move	−0.093	−0.125	−0.065	−0.143	−0.117	−0.166	−0.246*	−0.294*	0.054	0.166
in Armenian share	(0.080)	(0.096)	(0.079)	(0.093)	(0.167)	(0.227)	(0.127)	(0.171)	(0.146)	(0.169)
Effect of 10th–90th %ile move	−0.022	0.011	0.092	0.087	−0.006	0.056	−0.010	0.003	−0.111	−0.023
in Greek share	(0.104)	(0.115)	(0.084)	(0.097)	(0.111)	(0.103)	(0.092)	(0.086)	(0.137)	(0.141)

Notes: The sample contains individual respondents pooled from the two waves of KONDA Barometre Survey (September 2012 and September 2015) which contain questions on trust. The pooled sample covers only 33 out of 81 provinces but the sample in each wave is representative at the national level. Generalized trust is a dummy variable which is equal 1 if the respondent answered "Most people can be trusted" to the question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?". Answers to other trust questions range between 1 and 5 on a Likert scale, where 1 stands for "Not trust at all" and 5 stands for "Full trust". Inter-ethnic marriage measures whether it is acceptable for the respondent or his/her son or daughter to marry someone from another ethnicity. It also ranges between 1 and 5 where 1 stands for "Definitely not acceptable" and 5 stands for "Definitely acceptable". All models are estimated using OLS. The control variables are the same in all columns. Armenian and Greek shares, population density in 1927 and the geographical controls (same as the ones we use in our baseline analysis) are measured at the level of district where the respondent was interviewed. The control for historical demography is the share of Kurdish-speaking people in the 1927 Census measured at the province level. Individual characteristics include age, gender, fixed effects for education level, dummies for ethnic and religious identity and dummies for urban and metropolitan areas. For each outcome measure, the odd numbered column reports the results from the full sample of respondents while the even numbered column is based on a subsample that only includes respondents who were born in the province where they currently reside. Standard errors are clustered at the province level. The marginal effects associated with a move in the historical minority shares from 10th to the 90th percentile of their respective cross-district distributions are expressed in units of standard deviation of the outcome measure.

C Robustness Appendix

In this section, we provide the details of several robustness analyses to demonstrate that the significant baseline results are not simply due to the omission of some observable factors that could potentially affect economic prosperity independently of minority presence.

C.1 Threats to Identification

In our baseline specification, we control for a wide range of geographic and climatic factors as well as suitability to cultivation that could affect agricultural productivity. We have also shown in Table 2 that our results remain qualitatively robust when we control for fixed effects for 21 subregions or fixed effects for 81 provinces in our baseline sample. Nonetheless, we cannot preclude the possibility that minority shares were subject to selection on other factors not accounted by our baseline controls. Here, we first summarize various sources of endogeneity bias, and then, give a brief overview of the strategies we employ in the subsequent sections to mitigate these identification problems.

The main threat to our identification is the potential selection of minorities based on local drivers of development that we might have omitted from our regressions. If minority selection was positive, this would bias our estimates upwards. An alternative source of bias would be selective Muslim penetration across regions. If Muslim groups sorted into areas that offered greater economic potential regardless of minority presence, the concentration of non-Muslim population should be diluted over time. Such positive Muslim selection could bias our estimates on Greek and Armenian shares downwards. It is also possible that more productive people among Muslims self-selected into high minority regions. This selection could be due to locational fundamentals, other economic opportunities these locations offered to high skilled workers or certain attributes of minority groups living in these locations. Such high-skilled selection might explain faster development in regions with higher minority shares even after the expulsions. When we discuss minority influence on regional human capital accumulation (see Section 5), we argue that this type of sorting actually reflects an indirect minority legacy when, for example, the market forces and economic opportunities that attract skilled Muslims were themselves an outcome of long-lasting minority presence. When, in one way or the other, sorting of high skilled Muslims was directly or indirectly driven by minority presence, we interpret this as a legacy effect of minority presence.

Our first strategy to address these endogeneity concerns is to control for various correlates of historical and contemporary development that may confound our minority share coefficients.[56]

Importantly, we take into account early settlement patterns that might have been driven by development potential. To do that, we account for neolithic settlements across districts as well as ancient Greek and Armenian settlement patterns.[57]

We additionally run Oster (2019) test for coefficient stability to show that selection on unobservables is unlikely to explain away the positive coefficients of interest.

Lastly, we employ matching estimators as an alternative to OLS and compare locations with high and low minority presence that are similar with respect to exogenous geographic characteristics as well as possibly endogenous drivers of selective minority settlements. Matching allows us to more flexibly account for observable characteristics and it also improves identification by focusing on a smaller sample where treated and untreated districts are more similar.

In what follows, we show that our baseline conclusions qualitatively survive all of these alternative strategies. Therefore, taken together, the evidence we provide makes a strong case for a causal legacy effect of Greeks and Armenians on current development.

C.2 Robustness to Alternative Correlates of Development

We check the robustness of the baseline OLS estimates to several potential correlates of historical and contemporary development. In the following discussion, we motivate each set of controls we introduce to our baseline model. The results of all robustness checks are presented in Table C.1.

Access to Railroads and Ports: Regions with greater access to railroads or major port infrastructure in the past might have developed earlier than others (**??**). Since railroad network might well be an outcome of the economic activity spurred by minorities, we control for access to railroads as further in the past as possible. Using maps of historical rail network around 1910, we calculate distances of districts to the nearest railroad. Similarly, to measure

[56] Also, comparing column (1) of Table 2 with column (8) of Table 2, Altonji et al. (2005) ratios are 4.3 for Armenians and 3.05 for Greeks. Therefore, these values are reasonably large for us to have confidence that selection on unobservables is unlikely to explain away our estimates.

[57] In principle, we can conduct an instrumental variables analysis in which group-specific ancient settlement patterns (e.g. locations of capitals to old Armenian kingdoms and locations of ancient Greek sites in Anatolia) are used as instruments that predict 19th century Armenian and Greek shares, respectively. In fact, such an empirical strategy results both in a strong first-stage and significantly positive coefficient estimates for minority shares in the second stage. However, in view of a possible violation of the exclusion restriction assumption, we do not pursue this strategy.

access to sea trade, we compute distances to the nearest major 19th century port.[58] Column (2) shows that controlling for these two variables does not alter our findings.

Exposure to War: Early 20th century was a period of constant warfare. The Ottoman Empire participated in WWI which in turn led to the Turkish Independence War (1919-1923). These wars were devastating for Anatolia both in terms of human casualties and material destruction. Thus, one concern is that regions that were more heavily affected by the destruction of war might have fallen behind others. If exposure to war correlates with minority presence, this would lead to biased estimates. Controlling for two measures of war in column (3), number of Ottoman soldiers who died in battle during the WWI and distance to the nearest war front in the Turkish Independence War, we show that our results are qualitatively unaltered.

Settlement of Migrants and Kurdish Presence: Late 19th and early 20th centuries were a period of involuntary migrations. To disentangle the long-run impact of minorities on current outcomes from the potential effect of incoming migrants, we account for the number of immigrants who settled in Turkish provinces over the period 1921-1929 as a fraction of provincial population in year 1927 as well as the shares of 1927 population born outside modern-day Turkey (by country of birth). Also, historically, Kurdish population was highly concentrated in the eastern and the southeastern part of Anatolia in the areas inhabited by Armenians. Even after accounting for region fixed effects and geographic factors, Kurdish presence may still influence local development or it could proxy for confounding factors such as the low intensity civil conflict that plagued the predominantly Kurdish south-east of Turkey since the 1980s. Accordingly, we control for the share of Kurdish speakers in 1927 population. Column (4) suggests that none alters our findings.

Historical Regional Centers: One may also think our results are biased due to self-selection of minorities into historically more central and urbanized locations. To mitigate this problem, in column (5) of Table C.1, we control for an indicator for modern districts that were assigned to the central district (merkez *kaza*) of a given Ottoman *sancak* (akin to a county). Yet, Armenian and Greek shares remain highly significant.

Distance to Istanbul and to the National Borders: Proximity to Istanbul as the historical capital of the Ottoman Empire and the most important economic center of Turkey (both now and in the past) is a strong predictor of economic activity today. In addition, proximity to national borders may influence the degree and types of economic activity. To address these concerns, in column (6), we control for distances to Istanbul and to the nearest national border. Results do not change.

[58] These ports were situated in Istanbul, Trabzon, Mersin, Iskenderun, Samsun and Izmir.

Distances to Old Trade Routes: Minorities might have selected into areas with greater trading potential, and those locations with greater trading advantage might have developed independently of minorities' contribution to economic activity. To explicitly address this problem, in column (7) of Table C.1, we control for the distance to the Anatolian Silk Road as well as the distance to the Ottoman trade routes. Coefficients on Armenian and Greek shares remain highly significant, but reduced in size. However, it would be wrong to attribute the whole decline in estimated magnitudes to selective settlement. After all, it is highly likely that trade potential, and hence, the location of trade routes were themselves influenced by the persistent presence of these indigenous communities.

Finally, in column (8) of Table C.1, all robustness controls are added to the baseline model simultaneously. The results suggest that, even in the presence of a fairly exhaustive list of potential confounders, our main conclusions about the minority legacy remain unchanged. We consider the estimates under this very stringent specification as the respective lower bounds. They suggest that a district at the 90th percentile of the Armenian population share distribution prior to the deportations had, on average, 12 percent higher luminosity in year 2000 than a district without Armenians, while the corresponding difference due to Greek presence is about 19 percent.[59]

[59]We also check robustness to agro-ecological diversity. We compute agro-ecological fractionalization and polarization indices for a given district i as follows:

$$\text{Agro-ecological fractionalization}_i = 1 - \sum_{c=1}^{49+1} (s_i^c)^2$$

$$\text{Agro-ecological polarization}_i = 1 - \sum_{c=1}^{49+1} \left(\frac{0.5 - s_i^c}{0.5} \right)^2 s_i^c$$

where s_i^c is the fraction of cells within district i for which crop c has the greatest agro-ecological suitability among all other crops. Agro-ecological suitability index is downloaded from the FAO-GAEZ data portal. We normalize the original index values for each crop so they lie in the $[0, 1]$ interval. The agro-ecological suitability index is available for 49 crops. There is one additional category $c = 50$ that accounts for those cells where none of the 49 crops are cultivable. A cell belongs to this category if none of the crops have a suitability index that exceeds 0.4, a threshold adopted by some earlier studies to distinguish crops that are not cultivable in a grid cell. As another robustness check, instead of the suitability measure for the crop with the highest cultivation potential, we control for the average (across cells in a district) of the share of crops (among the 49 crops) that are cultivable (i.e. with a suitability index above 0.4) in a given cell. Finally, we control for average ruggedness as an alternative to the standard deviation of elevation. Due to the addition of new suitability measures, we lose 8 districts from our baseline sample which are too small to contain enough pixels for the computations. Table C.2 presents the results. Our previous conclusions are robust to this exercise.

C.3 Accounting for Early Selection and Settlement Patterns

The most important challenge to disentangling the minority legacy from other location-specific characteristics is early selection and settlements based on economic potential. To address this concern, we account for three deep-rooted factors that potentially shaped pre-historic human settlement patterns in Anatolia over the course of history.

Based on archaeological data, we first take into account the density of Neolithic settlements as a proxy for prehistoric conditions that were favorable to agriculture and amenable to permanent settlement.[60] Second, we locate ancient Greek sites dating to the Classical (480-323 BC) and the Hellenistic (323-146 BC) periods.[61] Third, we compute distances of district centers to Tushpa (the city of Van in modern-day Turkey) which was the capital of the ancient Urartu (Van) Kingdom (860-590 BC). Kingdom of Urartu was the first unified Armenian State and Tushpa was the economic, cultural and administrative center of attraction for Armenians.[62]

The results are presented in Table C.3. Column (1) reproduces the baseline results, whereas column (2) introduces neolithic sites into the regression. We observe that while neolithic sites positively predict modern economic development, the change in Armenian and Greek share coefficients are negligible. Thus, conditional on our baseline geographic controls, there is no evidence that Greek and Armenian legacy on current development is driven by the persistent and possibly positive influence of neolithic settlements. In column (3), we include a dummy for ancient Greek sites and the distance to Tushpa to account for ancient Greek and Armenian settlement patterns, respectively. Both variables are insignificant and hardly affect the minority coefficients. In column (4), we control for all three variables together, and the coefficients of interest remain remarkably stable. Finally in column (5), in addition to three early selection and settlement variables, we also control for all other robustness variables as in column 8 of Table C.1. Even in this very stringent specification, Armenian and Greek shares maintain their positive and significant influence on modern development, albeit the magnitudes are naturally reduced.

These results demonstrate that the geographic distribution of Greek and Armenian communities in the late 19th century Ottoman Empire explains a significant part of the

[60] Indeed, our regressions –not shown here– indicate that presence of neolithic sites predict higher agricultural suitability, one of the controls in our baseline model.

[61] While these sites predict geographic distribution of Ottoman Greeks in the late 19th century, a priori it is hard to rule out any influence they might have on modern development outcomes through other channels. Therefore, using ancient Greek sites as an instrument for 19th century Greek settlement patterns might not be a viable option.

[62] Distance to Tushpa might not only predict ancient Armenian settlement patterns but also early development and urbanization at the regional level. Therefore, using it as an instrument for the 19th century Armenian settlement patterns might not be a viable option.

contemporary cross-district variations in development even after taking into account early selection and settlement patterns.

C.4 Evaluating the Possible Influence of Omitted Variables

To test for the possible importance of omitted variable bias, we employ a method recently developed by Oster (2019), which builds on the work by Altonji et al. (2005). The basic idea is that if the inclusion of controls does not affect the magnitude of coefficient estimates, we can be more confident in suggesting a causal interpretation for the estimated relationship. In Appendix Table C.4 we report the statistic δ proposed in Oster (2019) both for Armenian and Greek population shares in the late 19th century Ottoman Empire under two different controlled models. Overall, our findings maintain that omitted variable bias should be less of a concern for our coefficient estimates for the Armenian share than for the Greek share.[63]

Coefficient stability analysis. One can compute a statistic δ which measures the degree of correlation between the unobservables and the variables of interest, relative to the correlation of the observables and the same variables of interest, that would be necessary to make the coefficients of interest equal to zero. An absolute value of 1 (or larger) for δ implies that the unobservables would need to be at least as important as the observables to produce a treatment effect of zero. Since researchers typically choose the controls they believe ex-ante to be the most important, absolute values of δ greater than 1 imply that it is unlikely that omitted variables are driving the results. However, as argued in Oster (2019), the coefficient movements due to additional controls should be scaled by the resulting change in R-squared. This adjustment in δ accounts for the fact that coefficient movements in response to less relevant controls are not as informative as the movements due to more relevant controls.

Table C.4 below reports the statistic δ proposed in Oster (2019) both for Armenian and Greek population shares in the late 19th century Ottoman Empire under two different controlled models.[64] As our restricted model we use a specification without any controls except subregion fixed effects. We first consider our baseline specification as the fully controlled model (column 7 in Table 2). The resulting proportionality of selection δ that is necessary to explain away the role of Armenian share is negative and sizable (-10.38). This finding implies that adding all relevant unobservables would need to (i) move the coefficient estimate in the opposite direction as the set of included observables does and (ii) selection

[63] A detailed discussion of this stability analysis in Table C.4 can be found in the text preceding that table.

[64] Following the rule of thumb suggested in Oster (2019), we compute δ under the assumption that if we could add all relevant unobservables in our model, the resulting R-squared would be 1.3 times as high as the R-squared of the fully controlled model (i.e. $R_{max} = 1.3 \times R^2$)

on these unobservables have to be more than 10 times as important as the selection on observables to produce a treatment effect of zero. The corresponding δ for Greek share is more moderate, 2.67. Nevertheless, both δ values are high enough to give us more confidence about a causal interpretation.

When we adopt the most stringent specification as our fully controlled model (column 5 in Table C.3), we compute $\delta = 2.16$ for Armenian share. For Greek share, $\delta = 0.76$. Since the latter value is lower than unity, it raises a reasonable concern for zero treatment effect for historical Greek presence. Yet, one also needs to keep in mind that some of the robustness controls are likely to be endogenous to minority presence, such as Ottoman trade routes or railroad network, which are likely shaped by the urban and commercial centers minorities established in the first place.

C.5 Matching Estimates

In this subsection, we carry out covariate and propensity score matching analyses. Matching estimators offer an appropriate comparison across treatment and control districts. Thus, better counter-factual control districts for the treated ones help mitigate endogeneity concerns. Also, restricting the analysis to the common support where treated and untreated units are more comparable improves upon identification.

We first construct Armenian and Greek treatment indicators with a value of one if the share of Armenians (Greeks) in that district is larger than the median Armenian (Greek) share across districts.[65] Given that the economic incentives of the time were mainly driven by agricultural activities, geographic and climatic characteristics, and suitability to cultivation must have influenced selection patterns. Therefore, we carry out our main matching exercise based on the exogenous geographic and climatic characteristics of the baseline model, namely, longitude, latitude, elevation, standard deviation of elevation, lake, sea, river dummies, temperature, precipitation, and suitability to cultivation.[66] We apply a two-sided common support to the propensity score, and restrict the sample to those treatment and control

[65] We first filter out subregion fixed effects from both minority shares and luminosity. This provides us with more comparable treatment and control units. In Panel A of the Online Appendix Table C.8, we reproduce the matching exercise with treatment indicators generated from the median of the raw data without filtering out subregion fixed effects. The results are similar. In addition, in Panel B of Table C.8, we provide results with another alternative treatment indicator, where treatment is set to one for districts with more than 1% minority share. The results are qualitatively the same. Therefore, our results are not sensitive to the definition of the treatment indicator.

[66] In the pre-treatment set of selection variables, only those variables that are unaffected by the treatment itself should be included. See **?** for a discussion on the choice of variables to include in the matching procedure.

districts with overlapping propensity scores. In addition, we ensure that the balancing property is satisfied throughout.[67]

Panels A and B of Table C.6 present the covariate (**??**) and propensity score matching estimates (**?**) of the average treatment effect of the Armenian and Greek treated districts (ATT) on luminosity with nearest neighbor, radius, kernel or stratification algorithms.[68] For comparison purposes, columns (1) and (2) show the OLS coefficients of minority treatments on the entire sample and on the common support.

In Panel A, the ATT of the Armenian presence under various matching methods ranges between 32.4% to 38.7%, always significant at the 1% level. The positive effect of the Armenian treatment under OLS in columns (1) and (2) is not very different from matching estimates, either on the entire sample or on the common support. In Panel B, the ATT of Greek presence on luminosity ranges from 47.7 to 63.6 percent. Moreover, OLS estimate on the entire sample seems underestimated compared to the coefficient of the regression on the common support.

Importantly, in Table C.9 of the Online Appendix, we confirm the robustness of our main results to matching on additional covariates of development, such as population density in 1893, distance to Istanbul, distances to railroads in 1910 and major ports, and distances to Anatolian silk road and Ottoman trade routes. Overall, our matching estimates suggest that districts with high historical minority concentration are more developed today even though they are no different in terms of other observable characteristics.

C.6 Robustness to the selection of benchmark year

So far we have used population density from the first census year *after* the expulsions (1927) –rather than the last Ottoman Census *before* the expulsions– as our baseline control. We have chosen a post-expulsion benchmark to remove any long-term effect of minority presence on modern outcomes that result from the contemporaneous influence of minority human capital while Greeks and Armenians were still living in these regions. We have chosen the *first* census year after the expulsions in order to capture the entire *legacy effect* of past

[67] We ensure that the balancing property is satisfied. We run the propensity score algorithm for the Armenian Treatment indicator and the balancing property is satisfied right away. This could be gauged already from Panel A of Table C.5, where most of the covariates are balanced across treated and control districts. When we run the propensity score algorithm for the Greek Treatment indicator, however, the balancing property does not hold. This could be expected from the poor covariate balance in Panel B of Table C.5. We, then, trim the sample to the propensity score interval of [0.2,0.8] over which the overlap between treatment and control is better as Panel C of Table C.5 indicates. Under this restricted sample balancing property is satisfied. Then, we move on to carry out our matching exercises (**??**).

[68] Table C.7 in the Online Appendix shows the robustness of the results to alternative propensity score matching methods.

minority presence on economic performance since the expulsions. Yet, there is one potential drawback to using 1927 figures for population density as the benchmark vis-a-which we assess the long-term economic performance. If in some regions, which were more affected by the expulsions and other adverse shocks, population size did not fully return to its post-expulsion long-run trend by 1927, our estimates may partly be confounded by a Malthusian convergence effect. To mitigate this concern, in Table C.10 we show that using a proxy for pre-expulsion population density based on 1893 figures (measured at the level of Ottoman districts) as an alternative control for initial conditions does not change our results about luminosity, urbanization and population density in a qualitative way.[69] Compared with the earlier baseline estimates from luminosity regressions (column 8 in Table 2), we see that both Armenian and Greek coefficients are still precisely estimated and are somewhat larger in magnitude (column 2 in Table C.10). This moderate difference in magnitudes may in part reflect the direct minority contribution to regional development before the expulsions –which we were originally removing from our baseline estimates. Controlling for 1893 and 1927 population figures at the same time to account for any change in economic conditions in-between, does not alter our main findings either. Finally, in Table C.11 we use population density in 1935 –the second census year after the expulsions– instead of 1927 figures as a benchmark. Given that any rebound effect in population size and economic recovery associated with it in the aftermath of the expulsions is less likely to influence growth after 1935, population density in 1935 can be viewed as a more conservative benchmark. Yet, our main findings are very similar under this alternative control.

C.7 Robustness to spatial correlation

To address concerns about spatial correlation, we carry out multiple exercises, largely following Kelly (2019). In short, the conclusion is that our results are not simply driven by spatial correlation. We elaborate below.

First of all, we generate spatially correlated noise at various correlation ranges and evaluated how well spatial noise can explain our dependent variable, luminosity, compared to our variables of interest, Armenian and Greek shares. In our simulations, spatial noise of a given district i is correlated with the noise of all other districts within a given correlation range, where the correlation weights of districts are inversely related to their distance from district i. Kelly (2019) uses correlation ranges of about 300 to 500 kilometers for European

[69] We should note that population density in 1893 is a rough proxy that is prone to additional measurement error. That is because the 1893 population figures from the Ottoman Census are reported for administrative regions that do not coincide with modern districts we use as our unit of analysis. Therefore, we had to make assumptions to construct measures for modern districts. For more details see the Data Appendix.

countries like Germany (with an area of 357,022 km^2) or France (with an area of 640,679 km^2). For Turkey (with an area of 783,356 km^2), we use spatial correlation ranges of about 300km, 500km, 800km, and 1800km, which, respectively, correspond to the 1st quartile, median, 3rd quartile, and the maximum distances between districts in Turkey.

Figure C.1 below presents the distribution of the standardized effect of spatially correlated noise on luminosity from 100 simulations in a specification akin to our baseline specification (in a regression of luminosity on spatial noise, controlling for population density in 1927, longitude, latitude, mean and standard deviation of elevation, lake, sea, major river, temperature, precipitation, suitability to cultivation, and subregion fixed effects). The distribution of the spatial noise effects is centered around zero. Notice also that, in the baseline specification, standardized Armenian share and Greek share coefficients correspond to 0.127 and 0.163 (with 1.25 and 1.24 unstandardized coefficients), respectively. Therefore, the coefficient estimates of spatial noise plotted in Figure C.1 are nowhere near the Armenian and Greek share effects, with a maximum effect of spatial noise around 0.06.

Moreover, Table C.12 assesses the robustness of the baseline results to spatial correlation.[70] First, for various baseline specifications, Panel A presents spatial correlation-adjusted Conley standard errors at various correlation ranges, assuming a linearly declining spatial weighting kernel (**?**). Correlation ranges are again the 1st quartile, median, 3rd quartile, and the maximum distances between districts in Turkey. We see in Panel A of Table C.12 that our results are robust to spatial correlation-adjusted Conley standard errors.

Second, Panel B presents the p-values of Moran statistic for spatial autocorrelation in regression residuals at various distances, where neighbors within a certain distance range are given equal weights following Kelly (2019). Results from Panel B indicate that, by and large, spatial correlation is limited. For example, in our baseline specification of column 7 with sub-region fixed effects, the Moran statistic is never significant at the 5% level. And, it is never significant at the 10% level in the more stringent specification of column 8 with province fixed effects (with p-values of at least 0.32).

Third, in Panels C and D, in 1000 simulations, we employ spatially correlated noise variables simulated similarly to the spatial noise described above. Spatial noise of a given district i is correlated with the noise of all other districts within a given correlation range, where the weights of districts are inversely related to their distance from district i. Panel C presents how often a spatial noise explanatory variable, used in place of the original variables of interest, is significant at p = 0.001 in luminosity regressions. For example, in the first row of Panel C when all districts are allowed to be correlated with one another (maximum correlation range), in the baseline specification of column 7, 0.1% of the time the spatial

[70] We mostly rely on the user written Stata command of *acreg* by **?**.

noise variable is significant. Alternatively, this fraction is 0% in column 8 specification when the correlation range is 300km. In comparison, among two thirds of the studies Kelly (2019) considers, spatial noise is significant at 0.001 level in 10% of the cases.

Fourth, in Panel D we employ spatial noise as a dependent variable. Panel D presents how often a spatial noise dependent variable, used in place of the original dependent variable, is explained by the variables of interest, Armenian and Greek shares, at a joint significance level of $p = 0.001$. For example, at a median correlation range (row 3), in the baseline specification of column 7, variables of interest are jointly significant in a spatial noise regression 0.5% of the time, while this fraction ranges between 0.4% and 0.9% (depending on the correlation range) in column 8 specification with province fixed effects. In comparison, in more than two thirds of the studies Kelly (2019) considers, when spatial noise is a dependent variable, variable of interest is significant at 0.001 level in 10% of the cases.

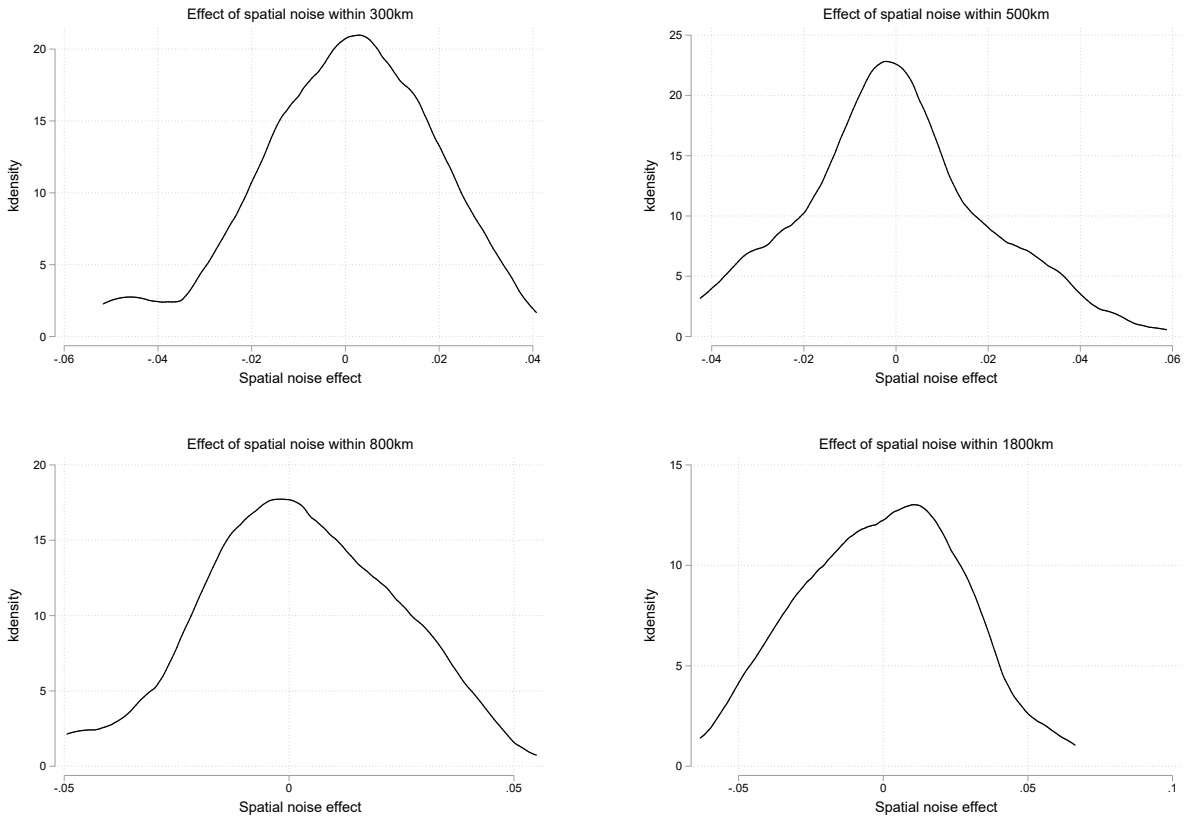


FIGURE C.1: Distribution of spatially correlated noise effects in 100 simulations

Notes: This figure presents the distribution of the standardized effect of spatially correlated noise on luminosity from 100 simulations in a specification akin to our baseline specification (a regression of luminosity on spatial noise, controlling for population density in 1927, longitude, latitude, mean and standard deviation of elevation, lake, sea, major river, temperature, precipitation, suitability to cultivation, and subregion fixed effects). Spatial noise of a given district i is correlated with the noise of all other districts within a given correlation range, where the weights of districts are inversely related to their distance from district i. Spatial correlation ranges are 300km, 500km, 800km, and 1800km. Note that in the baseline specification standardized Armenian share and Greek share effects correspond to 0.127 and 0.163 (with 1.25 and 1.24 unstandardized coefficients), respectively.

TABLE C.1: Minority Shares and Average Luminosity, Robustness to Historical Correlates of Economic Development

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS
	Log Average Luminosity in 2000							
Armenian population share, 1881-1893	1.250***	1.033***	1.240***	1.164***	1.103***	1.075***	0.846**	0.587*
	(0.372)	(0.347)	(0.371)	(0.347)	(0.385)	(0.355)	(0.339)	(0.315)
Greek population share, 1881-1893	1.241***	1.067***	1.238***	0.957***	1.179***	1.123***	0.797***	0.729**
	(0.296)	(0.281)	(0.295)	(0.296)	(0.312)	(0.253)	(0.270)	(0.293)
Log(Population density, 1927)	0.678***	0.577***	0.667***	0.652***	0.667***	0.685***	0.474***	0.446***
	(0.084)	(0.085)	(0.080)	(0.077)	(0.081)	(0.084)	(0.065)	(0.068)
Modern sub-region dummies	×	×	×	×	×	×	×	×
Baseline geographical controls	×	×	×	×	×	×	×	×
Distances to Railroad in 1910 & Major 19th century port		×						×
Distance to war front, 1919-1922			×					×
WW1 casualties in province			×					×
Share of settled immigrants in province, 1921-1929				×				×
Shares of 1927 population born outside modern-day Turkey (by country of birth)				×				×
Share of Kurdish speakers in province, 1927				×				×
In central kaza or sancak					×			×
Distances to Istanbul & nearest national border						×		×
Distance to Anatolian Silk Road							×	×
Distance to Ottoman Trade Routes							×	×
Observations	757	757	757	757	757	757	757	757
Adjusted R^2	0.513		0.515	0.522	0.515	0.523	0.587	0.600
Effect of 10th–90th %ile move in Armenian share	24.874***	20.565***	24.677***	23.154***	21.942***	21.386***	16.827**	11.680*
	(7.412)	(6.899)	(7.390)	(6.903)	(7.660)	(7.073)	(6.752)	(6.263)
Effect of 10th–90th %ile move in Greek share	31.995***	27.506***	31.902***	24.660***	30.385***	28.947***	20.542***	18.796**
	(7.622)	(7.235)	(7.603)	(7.626)	(8.047)	(6.522)	(6.962)	(7.547)

Notes: This table presents results from the regressions of Log Average Luminosity in 2000 on historical minority shares controlling for past population density, geographical variables, historical correlates of economic development, and subregion fixed effects. The estimated effect associated with increasing minority shares from the tenth to the ninetieth percentile of their respective cross-district distributions is expressed in terms of % change in the level of average luminosity in district. Robust standard errors, clustered at the modern Turkish province (*il*) level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE C.2: Minority Shares and Average Luminosity, Robustness to Alternative Geographical Controls

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
	Log Average Luminosity in 2000			
Armenian population share, 1881-1893	1.187***	1.240***	0.581*	0.631*
	(0.366)	(0.409)	(0.312)	(0.330)
Greek population share, 1881-1893	1.228***	1.044***	0.744**	0.680**
	(0.271)	(0.259)	(0.286)	(0.290)
Log(Population density, 1927)	0.666***	0.661***	0.446***	0.459***
	(0.082)	(0.079)	(0.068)	(0.064)
Agro-ecological fractionalization		0.318		0.068
		(0.296)		(0.283)
Agro-ecological polarization		−0.309*		−0.171
		(0.178)		(0.180)
Mean ruggedness		−0.012***		−0.008***
		(0.002)		(0.002)
Average fraction of crops suitable for cultivation		0.890*		0.957**
		(0.489)		(0.373)
Longitude, latitude	×	×	×	×
Lake, sea, major river	×	×	×	×
Temperature, precipitation	×	×	×	×
Mean elevation	×	×	×	×
Std. elevation	×		×	
Suitability to cultivation of the crop with greatest potential	×		×	
Distances to Railroad in 1910 & Major 19th century port			×	×
Distance to war front, 1919-1922			×	×
WW1 casualties in province			×	×
Share of settled immigrants, 1921-1929			×	×
Share of Kurdish speakers in province, 1927			×	×
In central kaza or sancak			×	×
Distances to Istanbul & nearest national border			×	×
Distance to Anatolian Silk Road			×	×
Distance to Ottoman Trade Routes			×	×
Modern sub-region dummies	×	×	×	×
Observations	749	749	749	749
Adjusted R^2	0.501	0.515	0.586	0.590
Effect of 10th–90th %ile move in Armenian share	23.632***	24.673***	11.563*	12.556*
	(7.276)	(8.132)	(6.212)	(6.576)
Effect of 10th–90th %ile move in Greek share	31.503***	26.780***	19.073***	17.441**
	(6.940)	(6.636)	(7.335)	(7.440)

TABLE C.3: Minority Shares and Average Luminosity. Accounting for Neolithic and Ancient Settlement Patterns

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
	Log Average Luminosity in 2000				
Armenian population share, 1881-1893	1.250***	1.273***	1.189***	1.213***	0.553*
	(0.372)	(0.388)	(0.371)	(0.386)	(0.319)
Greek population share, 1881-1893	1.241***	1.218***	1.185***	1.168***	0.659**
	(0.296)	(0.297)	(0.302)	(0.303)	(0.300)
Log(Population density, 1927)	0.678***	0.673***	0.678***	0.674***	0.447***
	(0.084)	(0.087)	(0.085)	(0.087)	(0.068)
Any settlement from Neolithic period		0.178**		0.165**	0.086
		(0.076)		(0.076)	(0.071)
Any ancient Greek site			0.182	0.165	0.142
			(0.114)	(0.114)	(0.097)
Log distance to Tushpa (Van)			−0.160	−0.155	−0.148*
			(0.103)	(0.100)	(0.083)
Modern sub-region dummies	×	×	×	×	×
Baseline controls	×	×	×	×	×
All other robustness controls					×
Observations	757	757	757	757	757
Adjusted R^2	0.513	0.516	0.515	0.518	0.601
Effect of 10th–90th %ile move in Armenian share	24.874***	25.340***	23.666***	24.144***	10.994*
	(7.412)	(7.722)	(7.380)	(7.682)	(6.354)
Effect of 10th–90th %ile move in Greek share	31.995***	31.401***	30.540***	30.117***	16.980**
	(7.622)	(7.655)	(7.792)	(7.805)	(7.727)

Notes: This table demonstrates that geographic distribution of Greek and Armenian communities in the late 19th century Ottoman Empire explains a significant part of the contemporary cross-district variations in nighttime light density even after we account for three deep-rooted factors that potentially shaped pre-historic human settlement patterns in Anatolia over the course of history. The first factor is the location of Neolithic settlements documented in TAY database of archaeological settlements in Turkey. These settlements likely capture the locations that were relatively more conducive to domestication of animals and plants, and indicate greater pre-historic agricultural potential. The second factor is the location of ancient Greek sites dating to the Classical (480-323 BC) and the Hellenistic (323-146 BC) periods. While these sites predict geographic distribution of Ottoman Greeks in the late 19th century, a priori it is hard to rule out any influence they might have on modern development outcomes through other channels. The third factor is the distance from Tushpa (the city of Van in modern-day Turkey) which was the capital of the ancient Van (Urartu) Kingdom (860-590 BC). Kingdom of Urartu was the first unified Armenian State. Distance to Tushpa could therefore not only predict ancient Armenian settlement patterns but also early development and urbanization at the regional level. Robust standard errors, clustered at the level of modern Turkish province (*il*), are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table C.4: Oster (2019) coefficient stability analysis

Full Model:	(1) **Baseline specification (column 7 of Table 2)**	(2) **All controls (column 5 of Table C.3)**
	R_{max}**=0.69**	R_{max} **= 0.82**
δ (Armenian share)	-10.38	2.16
δ (Greek share)	2.67	0.76

Notes: This table applies Oster (2019) technique to assess how strong the correlation between the unobservables and our main variables of interest (Armenian and Greek shares in the late 19th century Ottoman Empire), relative to the correlation of the observables with the same variables of interest, has to be in order to explain away the estimated relationships in the full models considered. δ statistic captures this relative selection ratio. All calculations have been performed using the software *psacalc* provided by the author. In Column (1), the full model corresponds to column 7 in Table 2. In Column (2), the full model corresponds to column 5 in Table C.3. Restricted model contains only subregion FE and no other control variables. R_{max} is the maximum R-squared that could be attained if all the relevant controls would be introduced in the regression. Following the advice in Oster (2019), we set R_{max} equal to 1.3 times the value of the R^2 obtained from the full model, which in our case equals 0.69 in column (1) and 0.82 in column (2). See the main text and Oster (2019) for details.

TABLE C.5: Mean Difference Tests for Treated and Control Districts

Panel A: Armenian Treatment	**Mean Treated**	**Mean Control**	**Mean Difference**	**P-Value**
Longitude	33.46	33.32	-0.14	0.66
Latitude	39.39	39.24	-0.15	0.17
Average Elevation	912.62	891.55	-21.07	0.56
Std. Elevation	241.69	270.04	28.35**	0.01
Lake	0.30	0.28	-0.02	0.57
Sea	0.20	0.23	0.02	0.48
Major Rivers	0.26	0.30	0.04	0.23
Temperature	0.01	0.01	0.00	0.11
Precipitation	0.63	0.66	0.03*	0.07
Suitability to Cultivation	4.26	4.12	-0.14	0.24
N	758	758	758	758

Panel B: Greek Treatment	**Mean Treated**	**Mean Control**	**Mean Difference**	**P-Value**
Longitude	34.42	32.38	-2.04***	0.00
Latitude	39.06	39.56	0.50***	0.00
Average Elevation	950.71	853.84	-96.87**	0.01
Std. Elevation	252.82	259.38	6.56	0.54
Lake	0.33	0.25	-0.08**	0.01
Sea	0.25	0.18	-0.07**	0.02
Major Rivers	0.26	0.31	0.06*	0.09
Temperature	0.01	0.01	-0.00**	0.04
Precipitation	0.61	0.68	0.07***	0.00
Suitability to Cultivation	4.13	4.24	0.12	0.33
N	758	758	758	758

Panel C: Greek Treatment (Trimmed)	**Mean Treated**	**Mean Control**	**Mean Difference**	**P-Value**
Longitude	32.72	32.59	-0.13	0.68
Latitude	39.29	39.52	0.23*	0.06
Average Elevation	885.06	913.28	28.23	0.47
Std. Elevation	251.44	252.05	0.61	0.96
Lake	0.28	0.26	-0.03	0.47
Sea	0.31	0.19	-0.12***	0.00
Major Rivers	0.24	0.28	0.04	0.32
Temperature	0.01	0.01	-0.00	0.10
Precipitation	0.63	0.64	0.02	0.33
Suitability to Cultivation	4.24	4.32	0.08	0.54
N	602	602	602	602

Notes: This table reports mean difference tests for exogenous covariates by minority treatment indicators. Minority treatment indicators are set to one for districts with above median minority shares after having filtered out the subregion fixed effects. Panel A reports mean difference tests by Armenian Treatment Indicator. Panel B reports mean difference tests by Greek Treatment indicator. Panel C reports mean difference tests by Greek Treatment indicator after having trimmed the sample to the [0.2, 0.8] interval of the propensity score. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table C.6: Minority Presence and Luminosity: Matching Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Covariate Matching	Propensity Score Matching			
	OLS	OLS	Nearest Neighbour	Nearest Neighbour	Radius (r=0.05)	Kernel (bw=0.02)	Stratification
	Panel A: Armenian Treatment Estimates (Dep.Var.: Log Average Luminosity in 2000)						
Armenian Treatment	0.302	0.298	0.348	0.292	0.328	0.316	0.304
Armenian Treatment (Bias Adjusted)	-	-	0.308	0.281	0.327	0.319	0.300
Bootstrapped Standard Errors	[0.084]***	[0.084]***	-	[0.093]***	[0.070]***	[0.071]***	[0.069]***
Analytical Standard Errors	(0.084)***	(0.085)***	(0.079)***	(0.091)***	(0.072)***	-	(0.068)***
Treatment effect (%)	35.3	34.7	36.1	32.4	38.7	37.6	35.0
Treatment Districts	-	-	372	372	372	372	372
Control Districts	-	-	379	201	379	379	379
Common Support	No	Yes	Yes	Yes	Yes	Yes	Yes
Balancing Property Satisfied	-	-	-	Yes	Yes	Yes	Yes
Number of observations	758	751	751	573	751	751	751
	Panel B: Greek Treatment Estimates (Dep.Var.: Log Average Luminosity in 2000)						
Greek Treatment	0.378	0.405	0.414	0.382	0.495	0.399	0.416
Greek Treatment (Bias Adjusted)	-	-	0.397	0.390	0.492	0.403	0.416
Bootstrapped Standard Errors	[0.090]***	[0.077]***	-	[0.094]***	[0.076]***	[0.071]***	[0.072]***
Analytical Standard Errors	(0.086)***	(0.082)***	(0.086)***	(0.098)***	(0.078)***	-	(0.073)***
Treatment effect (%)	45.9	49.9	48.7	47.7	63.6	49.6	51.6
Treatment Districts	-	-	279	279	278	279	278
Control Districts	-	-	322	152	322	322	323
Common Support	No	Yes	Yes	Yes	Yes	Yes	Yes
Balancing Property Satisfied	-	-	-	Yes	Yes	Yes	Yes
Number of observations	758	601	601	431	600	601	601

Notes: This table presents the covariate and propensity score matching estimates of the average treatment effect on the Armenian and Greek treated districts (ATT), in Panels A and B respectively. Armenian and Greek Treatment indicators are equal to 1 for above median shares of respective distributions after having filtered out subregion fixed effects. Subregion fixed effects are also filtered out of the dependent variable, luminosity. The baseline exogenous variables that are used in the matching procedure are longitude, latitude, elevation, standard deviation of elevation, lake, sea, river dummies, temperature, precipitation, and suitability to cultivation. To ensure that balancing property is satisfied, we trim the sample for the Greek treatment to the propensity score interval of [0.2,0.8]. We show two sets of ATT estimates that are either uncorrected or corrected for small sample bias due to non-exact matches. Nearest neighbor matching with random draw is applied in column (4). The Epanechnikov kernel (bandwidth=0.02) is applied in column (6). Bootstrapped standard errors (1000 replications) are given in brackets, while analytical standard errors are given in parentheses (both type of standard errors are clustered at the province level for OLS). *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE C.7: Minority Presence and Average Luminosity, Matching Estimates: Robustness of the results in Table C.6 to alternative propensity score matching methods

	(1)	(2)	(3)	(4)	(5)
	Propensity Score Matching				
	Nearest Neighbour (Equal Weights)	Radius (r=0.1)	Gaussian Kernel	Epanechnikov Kernel (bw=0.06)	Epanechnikov Kernel (bw=0.04)
Panel A: Armenian Treatment Estimates (Dep.Var.: Log Average Luminosity in 2000)					
Armenian Treatment	0.292	0.318	0.313	0.311	0.314
Armenian Treatment (Bias Adjusted)	0.273	0.307	0.315	0.301	0.312
Bootstrapped Standard Errors	[0.094]***	[0.064]***	[0.069]***	[0.065]***	[0.063]***
Analytical Standard Errors	(0.091)***	(0.070)***	-	-	-
Treatment Districts	372	372	372	372	372
Control Districts	201	379	379	379	379
Common Support	Yes	Yes	Yes	Yes	Yes
Balancing Property Satisfied	Yes	Yes	Yes	Yes	Yes
Observations	573	751	751	751	751
Panel B: Greek Treatment Estimates (Dep.Var.: Log Average Luminosity in 2000)					
Greek Treatment	0.382	0.500	0.429	0.400	0.399
Greek Treatment (Bias Adjusted)	0.392	0.500	0.430	0.396	0.394
Bootstrapped Standard Errors	[0.083]***	[0.074]***	[0.076]***	[0.076]***	[0.069]***
Analytical Standard Errors	(0.098)***	(0.078)***	-	-	-
Treatment Districts	279	279	279	279	279
Control Districts	152	322	322	322	322
Common Support	Yes	Yes	Yes	Yes	Yes
Balancing Property Satisfied	Yes	Yes	Yes	Yes	Yes
Observations	431	601	601	601	601

Notes: This table presents the propensity score matching estimates of the average treatment effect of the Armenian and Greek treated districts (ATT), in Panels A and B respectively. Armenian and Greek Treatment indicators are equal to one for above median shares of respective distributions after having filtered out subregion fixed effects. Subregion fixed effects are also filtered out of the dependent variable, luminosity. The baseline exogenous variables that are used in the matching procedure are longitude, latitude, elevation, standard deviation of elevation, lake, sea, river dummies, temperature, precipitation, and suitability to cultivation. To ensure that balancing property is satisfied, we trim the sample for the Greek treatment to the propensity score interval of [0.2,0.8]. We show two sets of ATT estimates that are either uncorrected or corrected for small sample bias due to non-exact matches. Nearest neighbor matching with equal weights is applied in column (1). The Epanechnikov kernel is applied in columns (4) (bandwidth=0.06) and (5) (bandwidth=0.04). Bootstrapped standard errors (1000 replications) are given in brackets, while analytical standard errors are given in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE C.8: Minority Presence and Average Luminosity: Further robustness checks for matching estimates, Robustness to alternative treatment definitions

Panel A: Robustness to alternative treatment definitions
Treatment = 1 if minority share is above the sample median of the raw data

	(1)	(2)	(3) Covariate Matching	(4)	(5) Propensity Score Matching	(6)	(7)
	OLS	OLS	Nearest Neighbour	Nearest Neighbour	Radius (r=0.05)	Kernel (bw=0.02)	Stratification
	Armenian Treatment Estimates						
Armenian Treatment	0.358	0.338	0.299	0.219	0.312	0.293	0.29
Armenian Treatment (Bias Adjusted)			0.232	0.179	0.307	0.278	0.28
Bootstrapped Standard Errors	[0.116]***	[0.110]***	-	[0.115]*	[0.088]***	[0.089]***	[0.086]***
Analytical Standard Errors	(0.112)***	(0.107)***	(0.097)***	(0.109)**	(0.086)***	-	-
Treatment Districts			313	313	313	313	313
Control Districts			290	161	290	290	290
Common Support	No	Yes	Yes	Yes	Yes	Yes	Yes
Balancing Property Satisfied	-	-	-	Yes	Yes	Yes	Yes
Observations	758	603	603	474	603	603	603
	Greek Treatment Estimates						
Greek Treatment	0.482	0.552	0.44	0.582	0.565	0.539	0.567
Greek Treatment (Bias Adjusted)			0.44	0.598	0.574	0.55	0.56
Bootstrapped Standard Errors	[0.107]***	[0.136]***		[0.172]***	[0.108]***	[0.141]***	[0.119]***
Analytical Standard Errors	(0.107)***	(0.139)***	(0.135)***	(0.147)***	(0.110)***	-	(0.116)***
Treatment Districts			154	154	153	154	154
Control Districts			153	83	153	153	153
Common Support	No	Yes	Yes	Yes	Yes	Yes	Yes
Balancing Property Satisfied	-	-	-	Yes	Yes	Yes	Yes
Observations	758	307	307	237	306	307	307

Panel B: Robustness to alternative treatment definitions
Treatment = 1 if minority share is above 1%

	(1)	(2)	(3) Covariate Matching	(4)	(5) Propensity Score Matching	(6)	(7)
	OLS	OLS	Nearest Neighbour	Nearest Neighbour	Radius (r=0.05)	Kernel (bw=0.02)	Stratification
	Armenian Treatment Estimates						
Armenian Treatment	0.269	0.27	0.189	0.202	0.279	0.202	0.236
Armenian Treatment (Bias Adjusted)			0.149	0.191	0.28	0.201	0.236
Bootstrapped Standard Errors	[0.110]**	[0.107]**		[0.094]**	[0.080]***	[0.079]***	[0.086]***
Analytical Standard Errors	(0.108)**	(0.108)**	(0.089)*	(0.109)*	(0.083)***	-	(0.084)***
Treatment Districts			477	477	477	477	477
Control Districts			275	175	275	275	275
Common Support	No	Yes	Yes	Yes	Yes	Yes	Yes
Balancing Property Satisfied	-	-	-	Yes	Yes	Yes	Yes
Observations	758	752	752	652	752	752	752
	Greek Treatment Estimates						
Greek Treatment	0.429	0.393	0.317	0.444	0.496	0.488	0.509
Greek Treatment (Bias Adjusted)			0.36	0.415	0.497	0.489	0.506
Bootstrapped Standard Errors	[0.099]***	[0.116]***		[0.174]**	[0.118]***	[0.145]***	[0.139]***
Analytical Standard Errors	(0.096)***	(0.117)***	(0.137)**	(0.150)***	(0.117)***	-	-
Treatment Districts			185	185	184	185	184
Control Districts			145	89	144	145	146
Common Support	No	Yes	Yes	Yes	Yes	Yes	Yes
Balancing Property Satisfied	-	-	-	Yes	Yes	Yes	Yes
Observations	758	330	330	274	328	330	330

Notes: This table presents the covariate and propensity score matching estimates of the average treatment effect of the Armenian and Greek treated districts (ATT). In Panel A, Armenian and Greek Treatment indicators are equal to one for above median shares of respective distributions of the raw data. In Panel B, Armenian and Greek Treatment indicators are equal to one for above 1% shares of respective distributions of the raw data. The baseline exogenous variables that are used in the matching procedure are longitude, latitude, elevation, standard deviation of elevation, lake, sea, river dummies, temperature, precipitation, and suitability to cultivation. To ensure that balancing property is satisfied, we trim the sample for the Armenian treatment to the propensity score interval of [0.3,0.7] in Panel A, and for the Greek treatment to the propensity score interval of [0.4,0.6] in Panel A and to [0.3,0.7] in Panel B. We show two sets of ATT estimates that are either uncorrected or corrected for small sample bias due to non-exact matches. Nearest neighbor matching with random draw is applied in column (4). The Epanechnikov kernel (bandwidth=0.02) is applied in column (6). Bootstrapped standard errors (1000 replications) are given in brackets, while analytical standard errors are given in parentheses (both type of standard errors are clustered at the province level for OLS). *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE C.9: Minority Presence and Average Luminosity, Matching Estimates: Robustness to matching on additional variables

	(1)	(2)	(3)	(4)	(5)
	Propensity Score Matching (Radius Matching with r=0.05)				
	Panel A: Armenian Treatment Estimates (Dep.Var.: Log Average Luminosity in 2000)				
Armenian Treatment	0.323	0.321	0.266	0.314	0.308
Armenian Treatment (Bias Adjusted)	0.320	0.321	0.266	0.315	0.321
Bootstrapped Standard Errors	[0.072]***	[0.070]***	[0.070]***	[0.071]***	[0.078]***
Analytical Standard Errors	(0.071)***	(0.072)***	(0.072)***	(0.072)***	(0.072)***
Treatment Districts	372	372	348	372	366
Control Districts	380	378	351	380	385
Common Support	Yes	Yes	Yes	Yes	Yes
Balancing Property Satisfied	Yes	Yes	Yes	Yes	Yes
Baseline Exogenous Variables	Yes	Yes	Yes	Yes	Yes
Observations	752	750	699	752	751
Additional Matching Variable	Population Density in 1893	Distance to Istanbul	Distances to Railroad in 1910 & Major 19th century port	Distances to Anatolian Silk Road & to Ottoman Trade Routes	All of the previous
	Panel B: Greek Treatment Estimates (Dep.Var.: Log Average Luminosity in 2000)				
Greek Treatment	0.493	0.504	0.450	0.446	0.367
Greek Treatment (Bias Adjusted)	0.492	0.506	0.462	0.452	0.371
Bootstrapped Standard Errors	[0.080]***	[0.075]***	[0.076]***	[0.080]***	[0.078]***
Analytical Standard Errors	(0.082)***	(0.079)***	(0.080)***	(0.078)***	(0.080)***
Treatment Districts	250	277	274	280	241
Control Districts	305	325	312	319	308
Common Support	Yes	Yes	Yes	Yes	Yes
Balancing Property Satisfied	Yes	Yes	Yes	Yes	Yes
Baseline Exogenous Variables	Yes	Yes	Yes	Yes	Yes
Observations	555	602	699	599	549
Additional Matching Variable	Population Density in 1893	Distance to Istanbul	Distances to Railroad in 1910 & Major 19th century port	Distances to Anatolian Silk Road & to Ottoman Trade Routes	All of the previous

Notes: This table presents the robustness of the propensity score matching estimates of the average treatment effect of the Armenian and Greek treated districts (ATT) to matching on additional variables, in Panels A and B respectively. Armenian and Greek Treatment indicators are equal to one for above median shares of respective distributions after having filtered out subregion fixed effects. Subregion fixed effects are also filtered out of the dependent variable, luminosity. The baseline exogenous variables that are used in every column for the matching procedure are longitude, latitude, elevation, standard deviation of elevation, lake, sea, river dummies, temperature, precipitation, and suitability to cultivation. Additional variables on which we match are highlighted in row "Additional Matching Variable." To ensure that balancing property is satisfied, we trim the sample for the Armenian treatment to the propensity score interval of [0.25,0.75] in columns (1), and (3). To ensure that balancing property is satisfied, we trim the sample for the Greek treatment to the propensity score interval of [0.2,0.8] in columns (1), (2), (3), (4), and (5). We show two sets of ATT estimates that are either uncorrected or corrected for small sample bias due to non-exact matches. Bootstrapped standard errors (1000 replications) are given in brackets, while analytical standard errors are given in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE C.10: Sensitivity of baseline OLS results to the omission and selection of population density controls – I

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS
	Log Average Luminosity, 2000			Log population density, 2000			Urbanization rate, 2000		
Armenian population share, 1881-1893	1.149**	1.338***	1.466***	1.064**	1.339***	1.493***	0.453***	0.505***	0.523***
	(0.463)	(0.444)	(0.364)	(0.465)	(0.452)	(0.372)	(0.100)	(0.102)	(0.095)
Greek population share, 1881-1893	1.354***	1.340***	1.224***	0.951***	0.930***	0.791**	0.209***	0.205***	0.189***
	(0.282)	(0.303)	(0.311)	(0.317)	(0.349)	(0.341)	(0.066)	(0.070)	(0.065)
Log(Population density, 1927)			0.684***			0.822***			0.094***
			(0.082)			(0.087)			(0.021)
Log(Population density, 1893)		−0.099**	−0.113***		−0.145***	−0.162***		−0.028***	−0.030***
		(0.045)	(0.037)		(0.054)	(0.045)		(0.009)	(0.008)
Other baseline controls	×	×	×	×	×	×	×	×	×
Modern sub-region dummies	×	×	×	×	×	×	×	×	×
Observations	757	757	757	757	757	757	757	757	757
Adjusted R^2	0.411	0.416	0.520	0.335	0.346	0.494	0.099	0.110	0.164
Effect of 10th–90th %ile move in Armenian share	22.871**	26.619***	29.166***	21.181**	26.647***	29.709***	9.008***	10.053***	10.404***
	(9.217)	(8.832)	(7.234)	(9.258)	(9.004)	(7.407)	(1.982)	(2.035)	(1.891)
Effect of 10th–90th %ile move in Greek share	34.904***	34.531***	31.544***	24.523***	23.979***	20.386**	5.391***	5.287***	4.875***
	(7.277)	(7.804)	(8.010)	(8.165)	(9.000)	(8.791)	(1.690)	(1.810)	(1.666)

Notes: This table presents results from the regressions of Log Average Luminosity (columns 1–3), Log Population Density (columns 4–6) and Urbanization Rate (columns 7–9) all measured for the year 2000 on historical minority shares under three alternative specifications to demonstrate that the results are not sensitive to (i) omitting population density in 1927 from the list of baseline controls (columns 1, 4 and 7) (ii) replacing population density in 1927 with the pre-expulsion level of population density (measured in 1893) as an alternative control (columns 2, 5 and 8) and (iii) controlling for population density in 1893 and population density in 1927 simultaneously (columns 3, 6 and 9). The remaining right hand side variables are baseline geographic controls, and subregion fixed effects. Robust standard errors, clustered at the modern Turkish province (*il*) level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE C.11: Sensitivity of baseline OLS results to controlling for population density in 1935 instead of 1927 – II

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
	Log Average Luminosity, 2000		Log population density, 2000		Urbanization rate, 2000	
Armenian population share, 1881-1893	1.250***	1.202***	1.185***	1.128***	0.466***	0.460***
	(0.372)	(0.379)	(0.373)	(0.349)	(0.091)	(0.087)
Greek population share, 1881-1893	1.241***	1.270***	0.816**	0.849**	0.194***	0.197***
	(0.296)	(0.329)	(0.317)	(0.346)	(0.061)	(0.060)
Log(Population density, 1927)	0.678***		0.814***		0.093***	
	(0.084)		(0.090)		(0.022)	
Log(Population density, 1935)		0.665***		0.808***		0.093***
		(0.082)		(0.088)		(0.020)
Longitude & Latitude	×	×	×	×	×	×
Mean & std. of elevation	×	×	×	×	×	×
Lake, sea and major rivers	×	×	×	×	×	×
Temperature & Precipitation	×	×	×	×	×	×
Suitability to cultivation	×	×	×	×	×	×
Modern sub-region dummies	×	×	×	×	×	×
Observations	757	757	757	757	757	757
Adjusted R^2	0.513	0.517	0.480	0.490	0.151	0.156
Effect of 10th–90th %ile move in Armenian share	24.874***	23.917***	23.586***	22.451***	9.282***	9.154***
	(7.412)	(7.543)	(7.415)	(6.943)	(1.817)	(1.732)
Effect of 10th–90th %ile move in Greek share	31.995***	32.738***	21.029***	21.892**	4.993***	5.087***
	(7.622)	(8.478)	(8.163)	(8.929)	(1.578)	(1.556)

Notes: This table presents results from the regressions of Log Average Luminosity (columns 1–3), Log Population Density (columns 4–6) and Urbanization Rate (columns 7–9) all measured for the year 2000 on historical minority shares under three alternative specifications to demonstrate that the results are not sensitive to (i) omitting population density in 1927 from the list of baseline controls (columns 1, 4 and 7) (ii) replacing population density in 1927 with the pre-expulsion level of population density (measured in 1893) as an alternative control (columns 2, 5 and 8) and (iii) controlling for population density in 1893 and population density in 1927 simultaneously (columns 3, 6 and 9). The remaining right hand side variables are baseline geographic controls, and subregion fixed effects. Robust standard errors, clustered at the modern Turkish province (*il*) level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table C.12: Robustness to Spatial Correlation

Dep. Var. Log(Average Luminosity in 2000)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Spatial correlation-adjusted Conley standard errors								
Armenian population share, 1881-1893	-0.539		-0.145	0.661	1.149	1.191	1.250	1.242
Cluster-robust standard errors	[0.685]		[0.639]	[0.441]	[0.463]**	[0.349]***	[0.372]***	[0.362]***
Conley standard errors (neighborhood distance=max≈1800km)	[0.697]		[0.651]	[0.199]***	[0.257]***	[0.156]***	[0.166]***	[0.173]***
Conley standard errors (neighborhood distance=3rd quartile≈800km)	[0.948]		[0.842]	[0.263]**	[0.358]***	[0.217]***	[0.267]***	[0.232]***
Conley standard errors (neighborhood distance=median≈500km)	[0.838]		[0.727]	[0.223]***	[0.348]***	[0.237]***	[0.320]***	[0.245]***
Conley standard errors (neighborhood distance=1st quartile≈300km)	[0.778]		[0.701]	[0.367]*	[0.394]***	[0.316]***	[0.355]***	[0.276]***
Greek population share, 1881-1893		2.385	2.371	1.472	1.354	1.370	1.241	0.694
Cluster-robust standard errors		[0.446]***	[0.444]***	[0.317]***	[0.282]***	[0.298]***	[0.296]***	[0.218]***
Conley standard errors (neighborhood distance=max≈1800km)		[0.237]***	[0.249]***	[0.167]***	[0.156]***	[0.116]***	[0.132]***	[0.118]***
Conley standard errors (neighborhood distance=3rd quartile≈800km)		[0.344]***	[0.368]***	[0.284]***	[0.262]***	[0.184]***	[0.214]***	[0.170]***
Conley standard errors (neighborhood distance=median≈500km)		[0.396]***	[0.402]***	[0.293]***	[0.289]***	[0.181]***	[0.265]***	[0.192]***
Conley standard errors (neighborhood distance=1st quartile≈300km)		[0.440]***	[0.445]***	[0.327]***	[0.322]***	[0.255]***	[0.298]***	[0.220]***
Panel B: P-values of Moran tests for spatial autocorrelation in regression residuals								
P-values of Moran tests (neighborhood distance=max≈1800km)	0.87	0.60	0.66	0.71	0.56	0.62	0.79	0.92
P-values of Moran tests (neighborhood distance=3rd quartile≈800km)	0.00	0.00	0.00	0.18	0.87	0.52	0.53	0.55
P-values of Moran tests (neighborhood distance=median≈500km)	0.00	0.04	0.02	0.00	0.11	0.00	0.07	0.38
P-values of Moran tests (neighborhood distance=1st quartile≈300km)	0.00	0.00	0.00	0.16	0.32	0.18	0.09	0.32
Panel C: The fraction of 1000 simulations where a spatial noise explanatory variable is significant at p=0.001								
Fraction significant (neighborhood distance=max≈1800km)	2%	2%	2%	0%	0.3%	0%	0.1%	0.2%
Fraction significant (neighborhood distance=3rd quartile≈800km)	0.6%	0.6%	0.6%	0.2%	1.1%	0%	1.2%	0.1%
Fraction significant (neighborhood distance=median≈500km)	1%	1%	1%	0.1%	1.2%	0.1%	1.3%	0%
Fraction significant (neighborhood distance=1st quartile≈300km)	3%	3%	3%	1.2%	0.6%	0.8%	1.7%	0%
Panel D: The fraction of 1000 simulations where the variables of interest explained a spatial noise dependent variable with a significance level of 0.001								
Fraction significant (neighborhood distance=max≈1800km)	42%	29%	59%	44%	1.7%	46%	1.7%	0.5%
Fraction significant (neighborhood distance=3rd quartile≈800km)	33%	22%	48%	48%	0.9%	48%	0.9%	0.7%
Fraction significant (neighborhood distance=median≈500km)	15%	20%	33%	41%	0.5%	42%	0.5%	0.9%
Fraction significant (neighborhood distance=1st quartile≈300km)	9%	13%	22%	28%	1.7%	27%	1.4%	0.4%
Log(Population density, 1927)						×	×	×
Longitude & Latitude				×	×	×	×	×
Mean & std. of elevation				×	×	×	×	×
Lake, sea and major rivers				×	×	×	×	×
Temperature & Precipitation				×	×	×	×	×
Suitability to cultivation				×	×	×	×	×
Modern subregion dummies					×		×	
Modern province dummies								×

Notes: This table assesses the robustness of the baseline results to spatial correlation. Panel A presents spatial correlation-adjusted Conley standard errors at various correlation ranges, assuming a linearly declining spatial weighting kernel (**?**). Panel B presents the p-values of Moran statistic for spatial autocorrelation in regression residuals at various distances, where neighbors within a certain distance range are given equal weight (Kelly, 2019). In Panels C and D, we employ spatially correlated noise variables in 1000 simulations. Spatial noise of a given district i is correlated with the noise of all other districts within a given correlation range, where the weights of districts are inversely related to their distance from district i. Panel C presents how often a spatial noise explanatory variable, used in place of the original variables of interest, is significant at p = 0.001. Panel D presents how often a spatial noise dependent variable, used in place of the original dependent variable, is explained by the variables of interest at a joint significance level of p = 0.001.