Online Appendix[34]

Uprootedness, Human Capital, and Skill Transferability

Stelios Michalopoulos
Elie Murard
Elias Papaioannou
Seyhun Orcan Sakalli

Abstract

The Online Appendix gives details on the data, descriptive and summary statistics, sensitivity checks, and further evidence. Section A gives mappings of the refugees' origin in Anatolia and settlement in Greece, summary statistics, and correlations on their settlement patterns. Section B reports additional descriptive and regression results associating forced displacement with refugees and their offspring's subsequent human capital investments. Section C gives additional results on the link between displacement and skill transferability of college degrees. Section D provides further descriptive and regression results, complementing the heterogeneity analysis and exploring additional displacement-education nexus mechanisms.

[34]Additional descriptives, examples, and details on the forced population exchange, the origin, and settlement of the Greek refugees from Anatolia can be found in https://anatolia-imprints.gr/. We thank Thanassis Stavrakoudis and his team of young researchers at the University of Ioannina for their help with the data. All errors are our responsibility.

A Descriptives and Summary Statistics

This Appendix complements the analysis in Section 3 of the paper. First, we provide a plethora of mappings and descriptions illustrating the settlement of refugees in Greece. Second, we explore differences between predominantly refugee and native rural settlements.

A.1 Descriptives and Mappings

Summary Statistics Appendix Table A1 gives summary statistics (mean; standard deviation; min, 10th percentile, median, 90th percentile, and max values) of the main variables (share of refugees, number of land grants, population in 1920 and 1928, share of Muslims in 1920) across 141 provinces (admin-3 level units) and 7,769 rural settlements.

Partial Refugee Census 1923 Appendix Figure A1 plots the share of refugees across 5,605 communities as recorded in the specialized (partial) Census in April 1923. Community (*koinotita* in Greek) is an administrative unit above the settlement level, which we use in our analysis. The specialized 1923 Census records 778,690 refugees, which is lower than the 1.2 million refugees recorded in the 1928 General Population Census, as the 1923 Refugee Census missed Greek-Orthodox refugees who arrived after the Population Exchange Convention and the Treaty of Lausanne.

Refugee Share across Provinces 1928 Appendix Figure A2 plots the share of refugees across 141 provinces, as recorded in the 1928 General Population Census. *Eparchies* (provinces), alongside *nomoi* (prefectures), have been the critical administrative units, coarse economic zones, from the establishment of the modern Greek state in the late 1820s till the 2000s. The boundaries of *eparchies* and the council's governing duties largely followed the *de facto* split during the Ottoman times, which in turn followed the church's diocese boundaries. Appendix Figure A3 plots the OLS association between the log number of refugees who arrived before and after 1922, as recorded in the 1928 General Population Census, across 141 provinces. The elasticity is 0.93, and the R^2 is 0.85.

Muslim Presence Appendix Figure A4 plots the share of Muslims (Bulgarians and Turks) who were forced to leave Greece after the 1923 Population Exchange Convention across 141 provinces, as recorded in the 1928 General Population Census. The Census suggests that 415,942 Turks and Bulgarians left. Appendix Figure A5 maps Muslim settlements and the share of Muslim populations across settlements. To impute the share of Muslims at the settlement level, we begin by computing, for each settlement, the change in its non-refugee population between 1920 and 1928. A fall in the population implies the out-migration of Muslims. Then, we scale up (or down) the settlement-level out-migration levels such that, once aggregated at the province (*eparhia*) level, it matches the province-level number of Muslims and Bulgarians left after 1920, as recorded in the 1928 census. Panel *A* plots the share of Muslims across Greece, where darker blue indicates a higher share. Panel

B zooms into Macedonia. In addition, the figure maps predominantly Muslim, Mixed, and Non-Muslim settlements, using information from Güvenç (2010) – the share of Muslims in 1915 is greater than 80% in "Predominantly Muslim" settlements; it is between 5% and 80% in "Mixed" settlements; and less than 5% in "Non-Muslim" settlements.

Ethnographic Map of Macedonia 1915 Appendix Figure A6 reproduces the digitized ethnographic map of Macedonia in 1915 (Güvenç, 2010). The map reports the main settlements alongside estimates of the Turkish, Bulgarian, and Greek-Orthodox populations.

Muslim and Refugee Settlements Appendix Figure A7 plots the density of the imputed share of Muslims departing after 1920 in Panel A and of the share of refugees in 1928 in Panel B across "Predominantly Muslim" villages (share of Muslims in 1915 equals or exceeds 80%), "Mixed" villages (share of Muslims in 1915 between 5% and 80%), and "Non-Muslim" villages (share of Muslims in 1915 equals or is less than 5%) according to the ethnographic map of Macedonia in 1915 (Güvenç, 2010). Panel A shows that the imputed share of departing Muslims proxies well the observed distribution of Muslims in 1915 across settlements in Macedonia—the imputed share of departing Muslims is skewed to the right among "Predominantly Muslim" settlements in 1915 and skewed to the left among "Non-Muslim" settlements in 1915. The correlation is high: 218 (80.7%) of 270 predominantly Muslim villages in 1915 are also predominantly Muslim in 1920 with our procedure. We split predominantly refugee settlements, where the share of refugees in 1928 exceeded 80%, into former Muslim settlements and new ones. Appendix Figure A8 plots the two types of refugee settlements across Greece (Panel A) and in Macedonia and West Thrace (Panel B), as four out of five rural refugees settled there.

Rural Refugee Catalog Examples Appendix Figure A9 illustrates the origin-destination structure of the rural refugee catalog, compiled by the Refugee Settlement Committee (RSC) in the mid-1920s. Panel A gives the settlements of origin in Anatolia of Greek-Orthodox refugees in Agios Constantinos, in Aetolia-Acarnania in West Greece. Most refugees come from Pontus, on the southern shore of the Black Sea, and from the Interior of Ionia on the West Coast of Asia Minor. Panel B maps the settlements of refugees from Panormos (Bandırma) in the Marmara Sea. Most Greek-Orthodox who left Panormos settled in West Thrace and East Macedonia.

Origin Settlements Appendix Figure A10–Panels A and B portray the Greek-Orthodox settlements in contemporary Turkey from where refugee families come from. We identify $1,850$ origin settlements in the rural refugee catalog, which we have processed in its entirety (Panel A). There are 923 settlements in the urban refugee catalog, which we have partly processed (about 43%). 923 settlements appear in both catalogs, 927 only in the rural catalog, and one only in the urban catalog. $1,735$ origin settlements are in modern Turkey; 115 are outside of Turkey (in contemporary Bulgaria, Ukraine, Russia, and Georgia).

Census Validation Appendix Figure A11 illustrates the correlation between the (log) number of refugees in the 1928 Population Census and the (log) number of families that received a land grant, as recorded in the Rural Refugee Catalog, compiled by the Refugee Settlement Committee in the mid-1920s. We focus on rural settlements, i.e., exclude settlements with greater than a population of 10,000 in 1928, province capitals, and settlements within 25 km of Athens, Piraeus, and Thessaloniki. Panel *A* gives the correlation across 1,284 rural settlements with at least one land grant. Panel *B* plots the correlation across 384 settlements with at least one land grant where the share of refugees in 1928 exceeds 80%. The correlation is strong. The number of families settled by the RSC with a land grant explain 88% of the variation in the presence of refugees in a settlement in 1928.

Baseline Sample Appendix Figure A12 maps the 405 predominantly-refugee villages that received at least one land grant, and where the refugees' population share exceeds 80% (in red), and the 6,091 native villages (in blue), where refugees' share is below 5%, and there was no recipient of a land grant. The figure also maps mixed refugee-native settlements (in low-tone gray). Panel *A* maps the country; Panel *B* maps Macedonia and West Thrace; Panel *C* zooms into the Giannitsa province in Central Macedonia, respectively.

Local Sample Appendix Figure A13 maps the local sample of predominantly-refugee and native settlements, where each native settlement is matched to the nearest predominantly-refugee settlement in the same province within 25 km. To perform this matching, first, we identify all pairs of refugee and native settlements within 25 km of each other in the same province (excluding provinces with no pairs). Second, for each native settlement, we identify the closest predominantly refugee settlement. Third, we split each province into Voronoi polygons based on the location of these refugee settlements. Fourth, we exclude refugee settlements further than 25 km from a native settlement within the same Voronoi polygon. Panel *A* maps the whole country, while Panel *B* zooms into Macedonia and West Thrace. Panel *C* presents the sub-province Voronoi polygons constructed for Giannitsa province in Central Macedonia, along with the names of native and predominantly refugee settlements nearest to them in the same province.

Age Distribution in 1928 Appendix Figure A14 presents the age distribution by gender and refugee status as provided by the 1928 General Population Census. The share of males aged 20 to 44 is largely affected by the Balkan Wars, the Great War, and the Greco-Turkish War. However, it is much more pronounced among refugees, particularly in the 20-24 age range, as the Greek population was subject to persecution in the Ottoman Empire during this long decade of conflict. In particular, young Greek men served in the Labor Battalions (*Amele Taburları*), where mortality rates were very high during WWI (Minasidis, 2024). Notably, the lower representation of children below the age of 9 among refugees in 1928, mainly born in Greece after the exodus, reflects the adverse economic conditions, poor health, and refugees' high mortality in their first years in Greece.

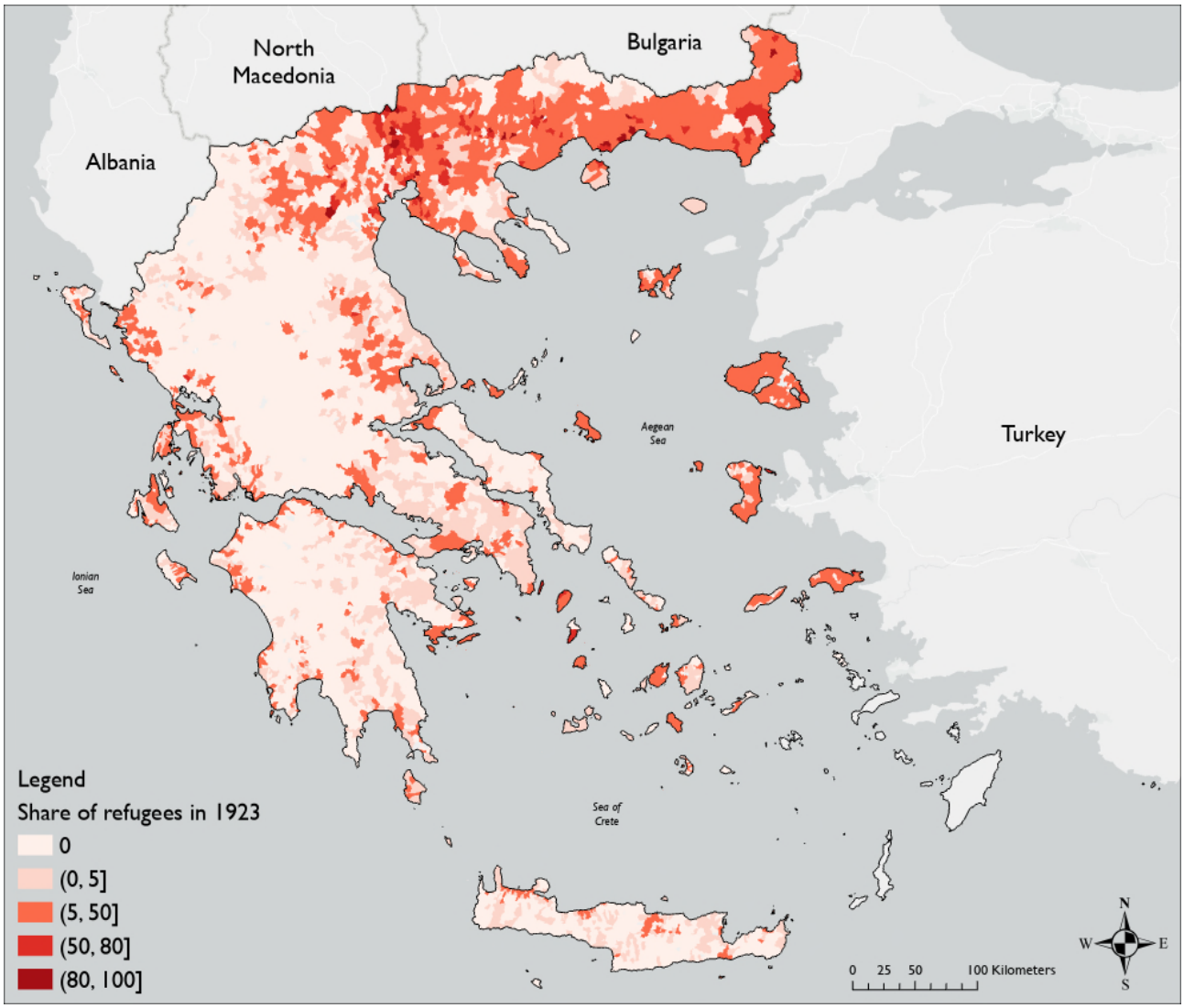
Population Dynamics Appendix Figure A15 plots the evolution of the population of predominantly refugee and native settlements from 1907 until 2011 in the baseline (Panel *A*) and the local sample (Panel *B*), using the information from the respective General Population Censuses. Three main takeaways emerge. First, our comparisons mainly entail small and mid-sized villages. Second, population differences between the two sets of settlements were minor at the time of the population exchange (as recorded in the 1928 General Population Census). Third, the population evolves similarly in both types of settlements. Despite the overall increase in Greece's population, the population in rural refugee and native settlements does not increase much.

Table A1: Summary Statistics across Provinces and Settlements

	Mean	SD	Min	10p	Median	90p	Max	Obs.
Panel A:				**Province characteristics**				
Population in 1928	43,970.4	78,119.6	3,037	10,342	30,992.0	73,403	866,924	141
Population in 1920	35,544.3	47,561.4	3,190	9,709	25,302.0	60,569	501,615	141
Share of rural population in 1928	86.3	21.3	14	57	100.0	100	100	141
Surface area	914.0	630.6	29	268	742.7	1,743	3,067	141
Share of refugees in 1928	14.0	20.2	0	0	3.5	46	90	141
Share of refugees in rural areas in 1928	13.4	21.3	0	0	2.4	46	95	141
Share of Turks and Bulgarians in 1920	8.4	18.0	0	0	0.1	33	91	141
Number of settlements with land grant	10.4	17.7	0	0	1.0	35	110	141
Number of refugee families settled with land grant	1,318.4	2,587.1	0	0	63.0	4,651	14,298	141
Share of refugees from Constantinople in 1928	0.2	0.4	0	0	0.1	0	3	141
Share of refugees from Asia Minor in 1928	5.6	7.4	0	0	2.0	16	35	141
Share of refugees from Pontus in 1928	2.7	6.6	0	0	0.1	9	49	141
Share of refugees from Thrace in 1928	3.6	7.8	0	0	0.1	16	38	141
Share of refugees from other regions in 1928	0.1	0.4	0	0	0.0	0	4	141
Panel B:				**Settlement characteristics**				
Population in 1928	485.4	565.5	0	84	329.0	1,010	10,438	7,769
Population in 1920	431.4	504.8	0	69	288.0	919	8,483	7,769
Share of refugees in 1928	10.0	24.4	0	0	0.0	44	100	7,769
Settlement with land grant (%)	17.5	38.0	0	0	0.0	100	100	7,769
Number of refugee families settled with land grant	19.9	68.6	0	0	0.0	61	1,123	7,769
Monastery in 1920 (%)	4.1	19.9	0	0	0.0	0	100	7,769
Log distance to reclaimed marshlands	2.7	1.2	0	1	2.9	4	5	7,769
Log distance to province capital	2.6	0.6	0	2	2.7	3	4	7,769
Log distance to coast	2.6	1.2	0	1	2.6	4	5	7,769
Log distance to railway stations	3.3	1.5	0	1	3.2	6	6	7,769
Log crop suitability	3.5	0.5	0	3	3.6	4	4	7,767
Log altitude	5.6	1.1	0	4	5.8	7	7	7,769
Log average precipitation	4.0	0.2	3	4	4.0	4	5	7,769
Log average temperature	2.7	0.1	2	3	2.8	3	3	7,769

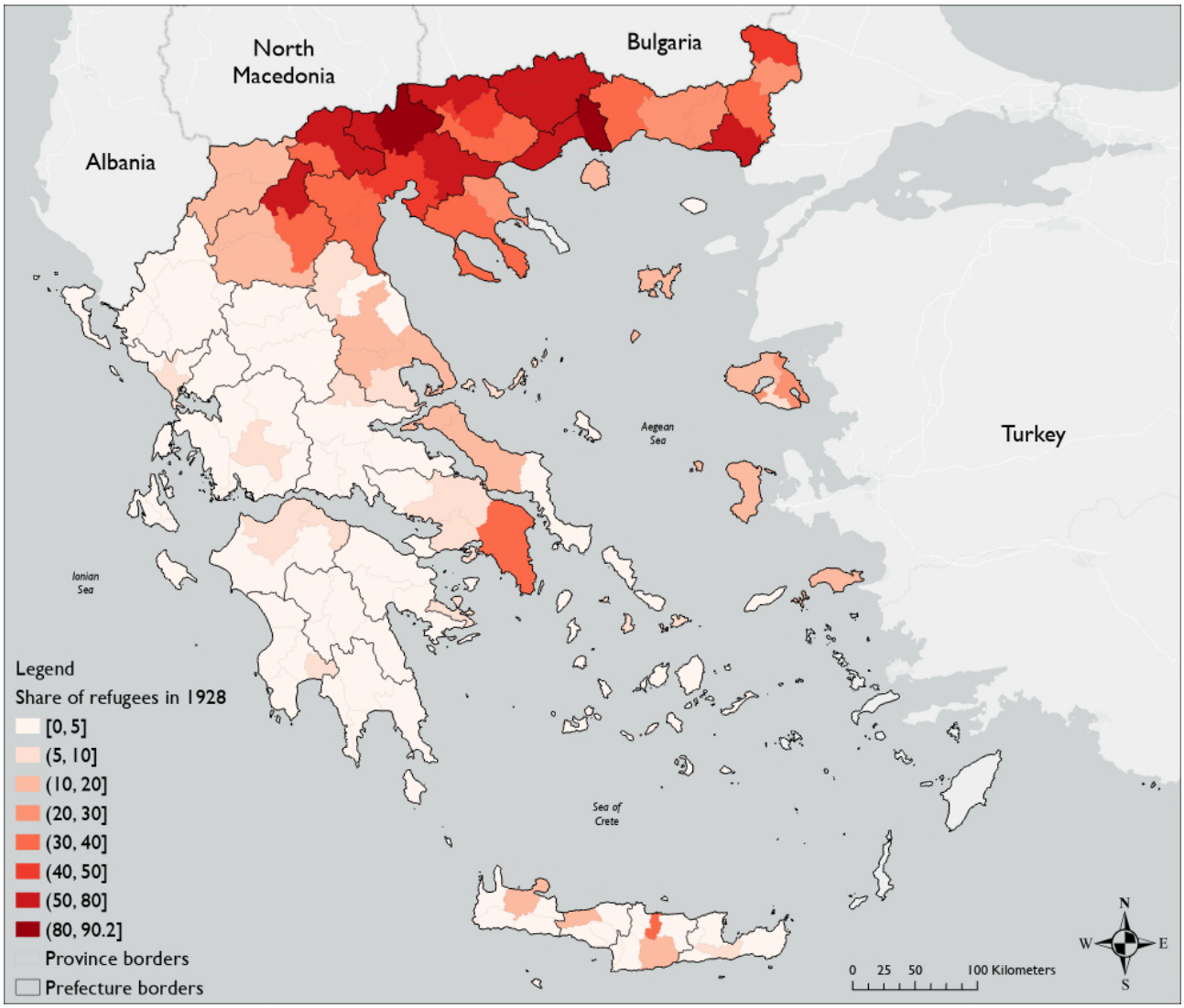
Notes. The table gives summary statistics of population and refugee shares across provinces and settlements, along with settlement geographical characteristics. Panel *A* gives statistics across 141 provinces (*eparchies*). Panel *B* gives statistics across all rural settlements, baseline sample, which excludes cities with a population exceeding 10,000 in 1928, province capitals, and settlements within 25 km distance to Athens, Piraeus, and Thessaloniki.

Figure A1: Refugee Share in 1923



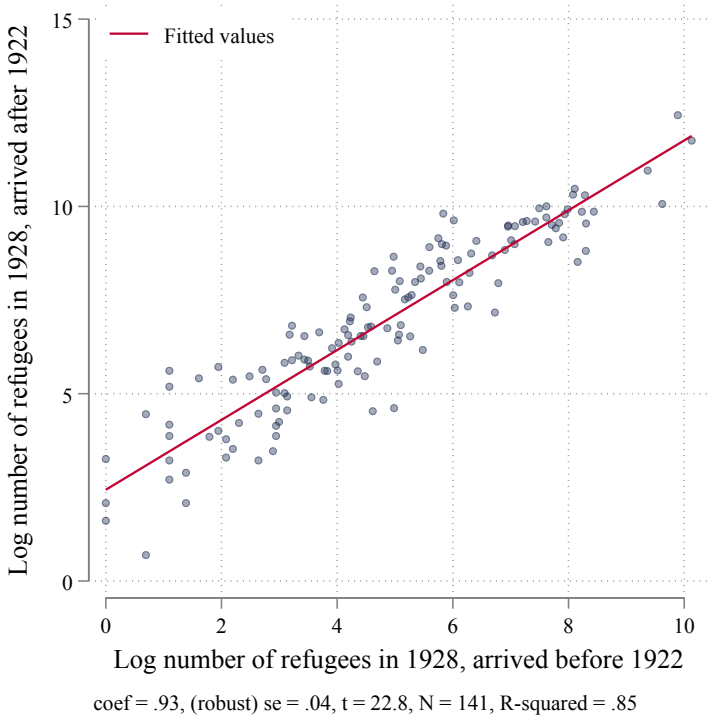
Notes. The figure maps the share of refugees in 1923 across 5,605 communities, as recorded in the specialized census of the Refugee Settlement Committee. (Community (*koinotita* in Greek) is an administrative unit above the settlement level, which we use in our analysis.) The partial Census includes 778,690 refugees.

Figure A2: Refugees in 1928 across Provinces (*Eparchies*)



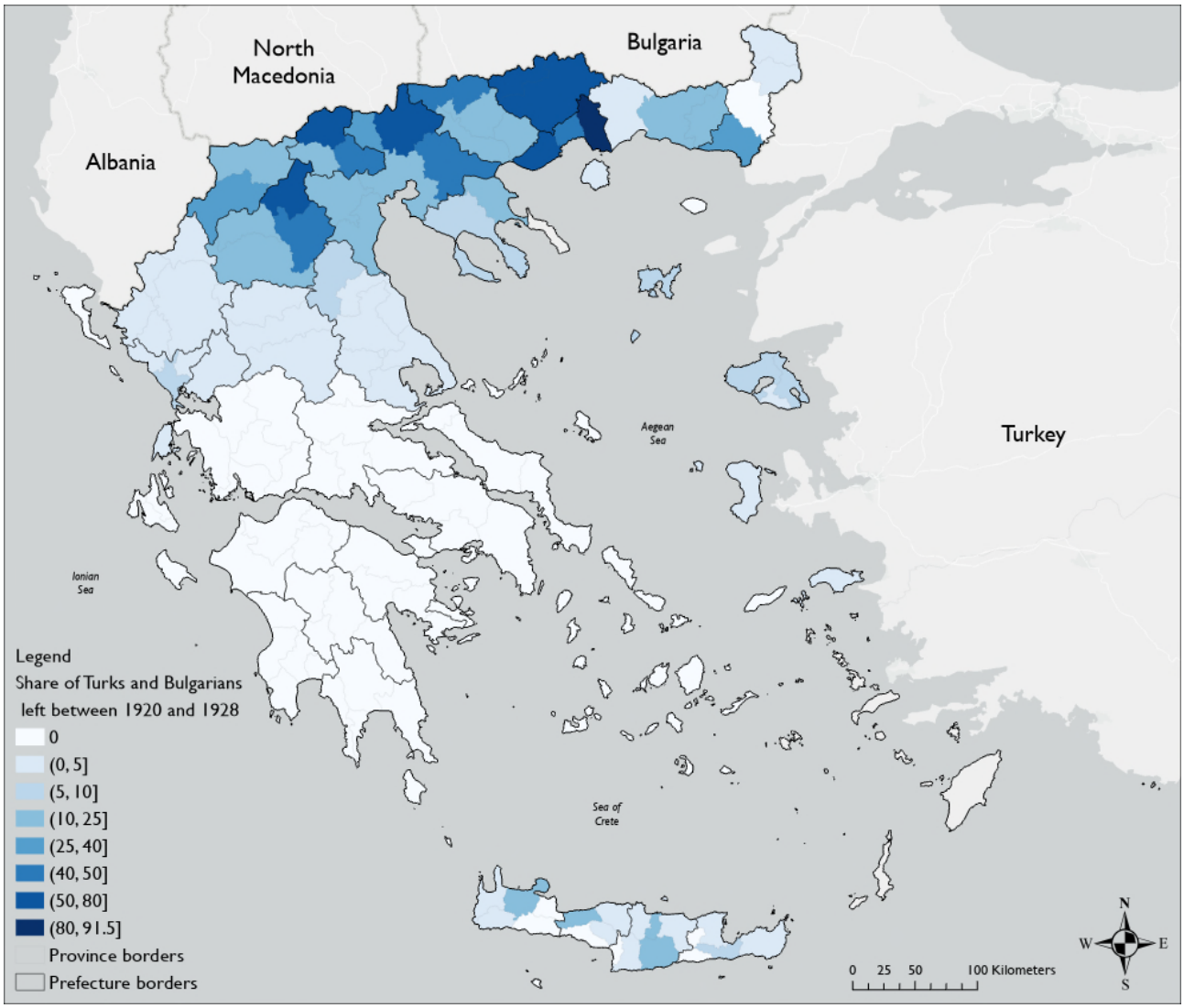
Notes. The figure maps the distribution of all refugees across 141 provinces (*eparchies*), excluding Mount Athos, as recorded in the 1928 General Population Census. There are 1,221,849 refugees in a total population of 6,204,684.

Figure A3: Correlation Refugees Arriving before and after the Asia Minor Catastrophe (1922–23)



Notes. The figure shows the OLS association between the log number of refugees arriving after 1922 (vertical axis) and before 1922 (horizontal axis) across 141 provinces (admin-3 level units). Data come from the 1928 General Population Census.

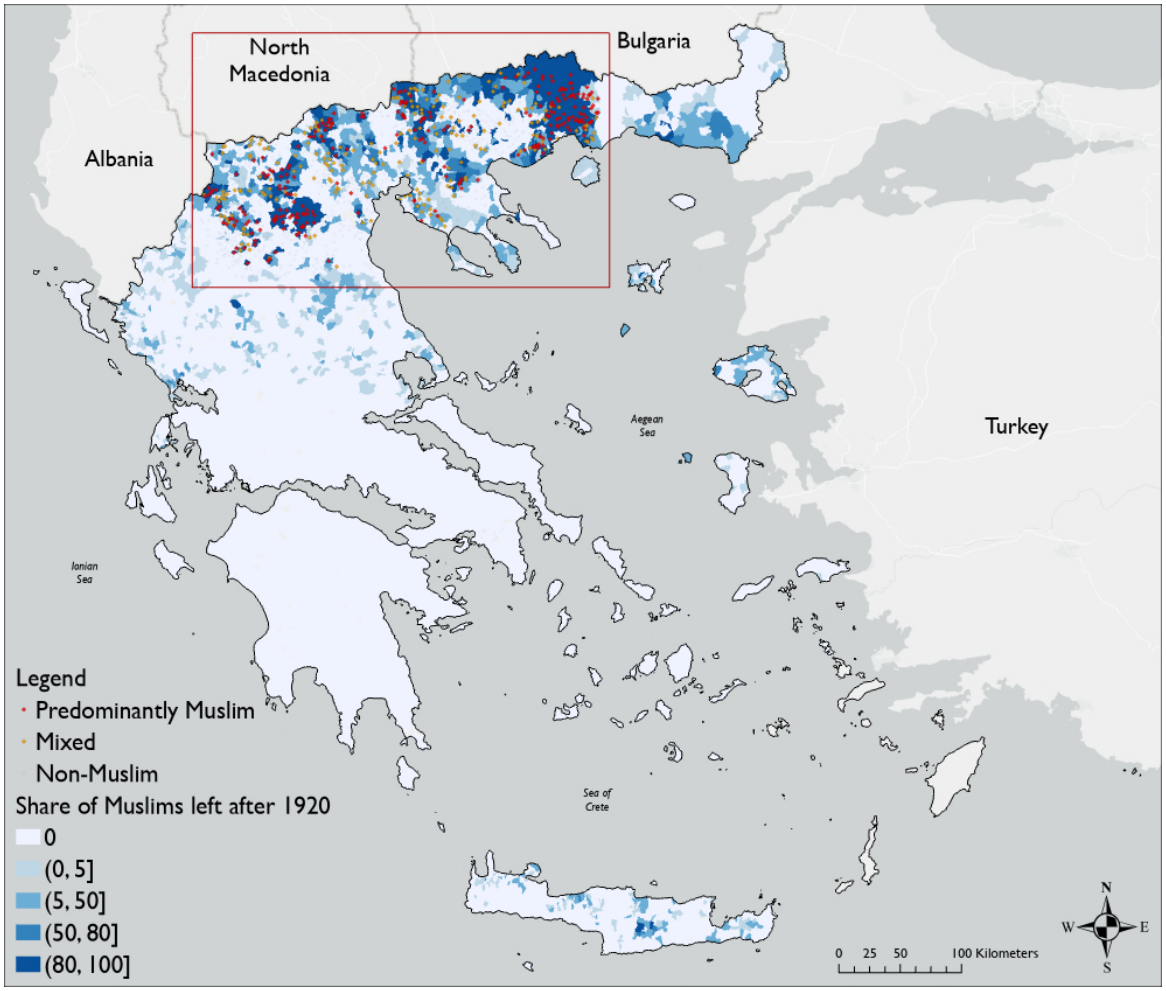
Figure A4: Muslim Population (Turks and Bulgarians) before the Population Exchange



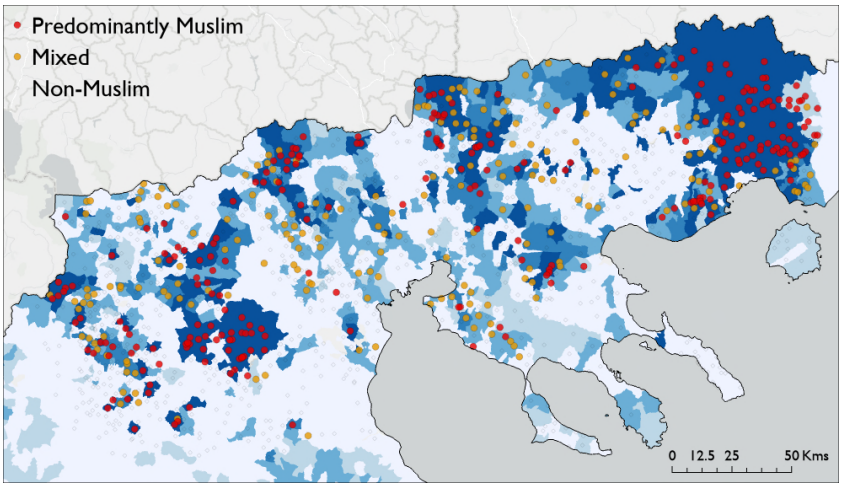
Notes. The figure maps the share of Turks and Bulgarians who left Greece between 1920 and 1928 across 141 provinces (*eparchies*), excluding Mount Athos, as recorded in the 1928 General Population Census. In total, 415,942 Turks and Bulgarians left Greece in this period.

Figure A5: Muslim Share in 1920 across Settlements

(a) All Greece

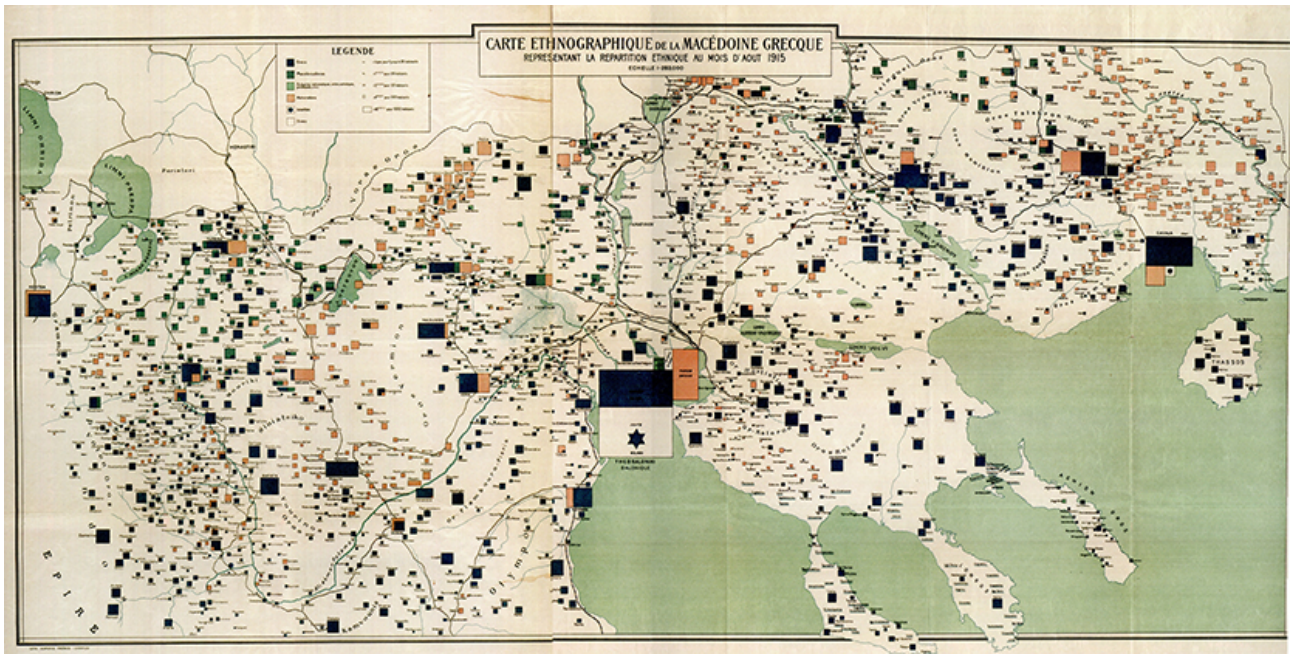


(b) Macedonia



Notes. The panels map the imputed share of Muslim (Turkish and Bulgarian) refugees who left Greece after 1920 across 5,605 communities in Greece in Panel *A*, and across communities in Macedonia in Panel *B*. (Community (*koinotita* in Greek) is an administrative unit above the settlement level, which we use in our analysis.) The red rectangle in Panel *A* marks the regions covered by the ethnographic map of 1915 (Güvenç (2010)).

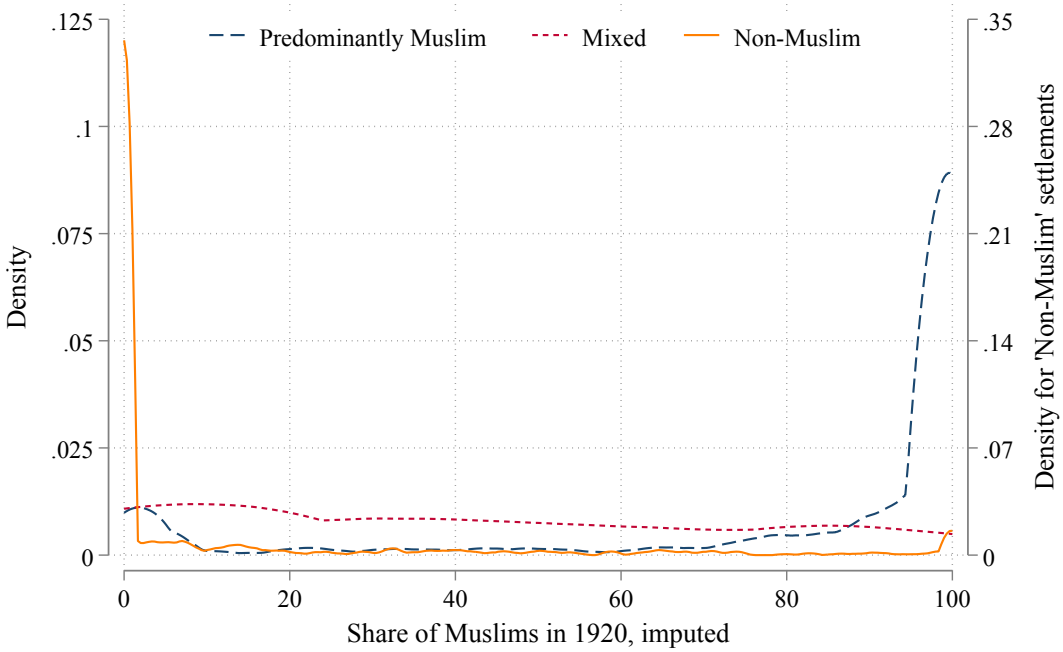
Figure A6: Ethnographic Map of Macedonia (1915)



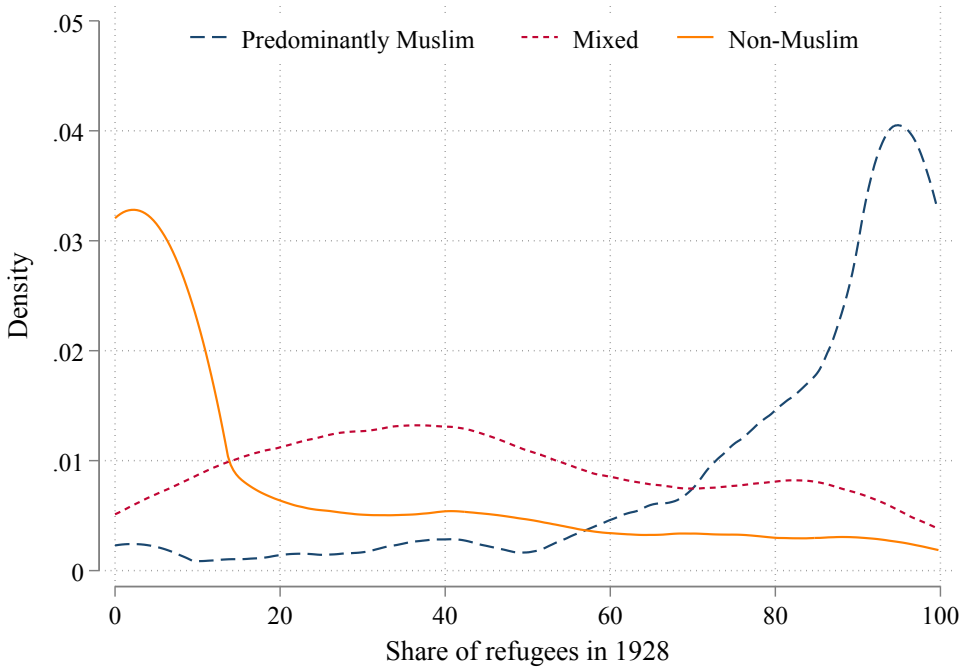
Notes. The figure presents the population distribution by ethno-religious groups at the settlement level in Macedonia in 1915. The title of the map reads: *"Carte ethnographique de la Macedoine Grecque: presentant la repartition ethnique au mois d'aout 1915"* [Ethnographic map of Greek Macedonia: presenting the ethnic distribution in August 1915]. The map gives ethnoreligious groups for localities in Drama, Florina, Halhidiki, Kavala, Kozani, Pella, Thessaloniki, and Serres prefectures and some localities in Elassona province of Larissa prefecture. Data source: Güvenç (2010).

Figure A7: Distribution of Greek-Orthodox and Muslim Refugees across Settlements

(a) Departing Muslims (after 1920)



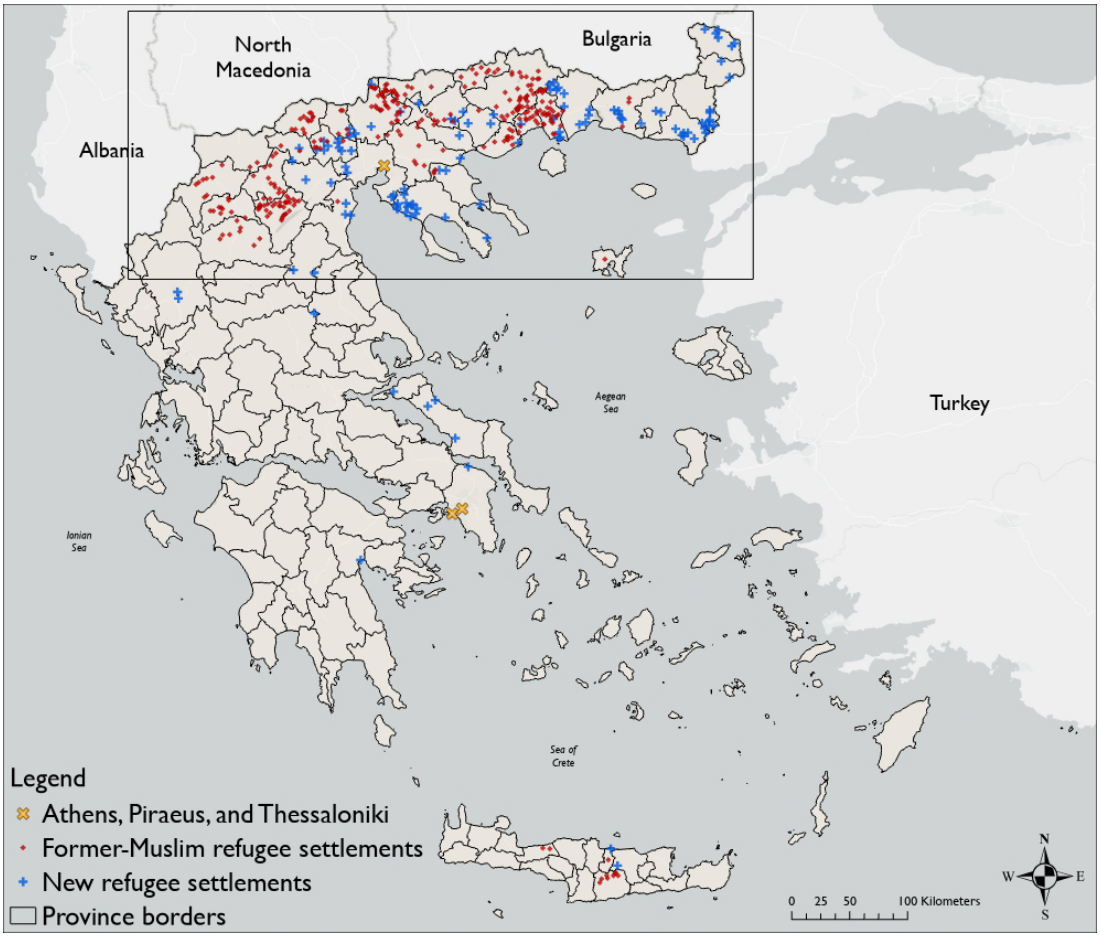
(b) Greek-Orthodox Refugees (in 1928)



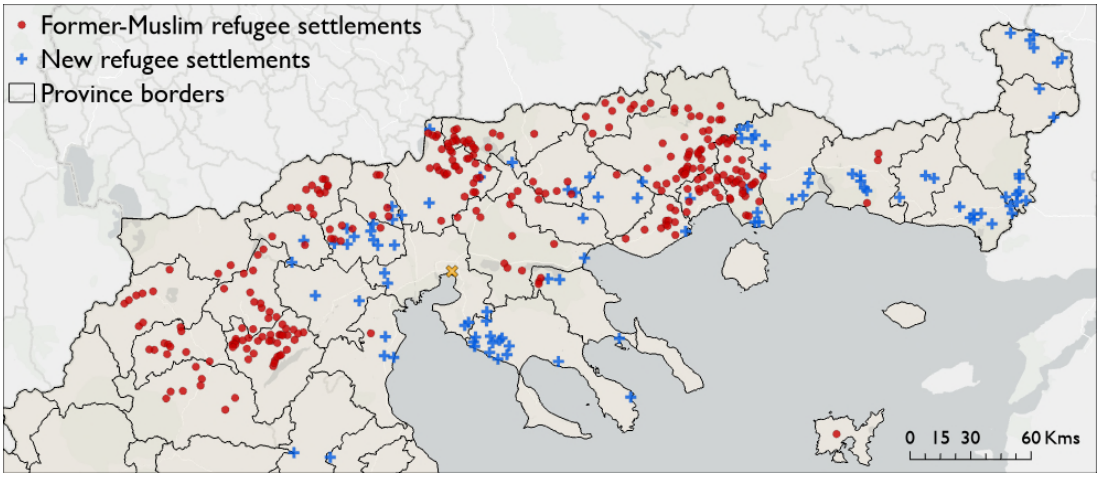
Notes. The figures plot the density of the imputed share of Muslims departing after 1920 in Panel *A* and the share of refugees in 1928 in panel *B* across "Predominantly Muslim" villages (share of Muslims in 1915 exceeds 80%), "Mixed" villages (share of Muslims in 1915 between 5% and 80%), and "Non-Muslim" villages (share of Muslims in 1915 less than 5%) according to the ethnographic map of Macedonia in 1915 (Güvenç (2010)).

Figure A8: Rural Refugee Settlements by Type in 1928

(a) Greece



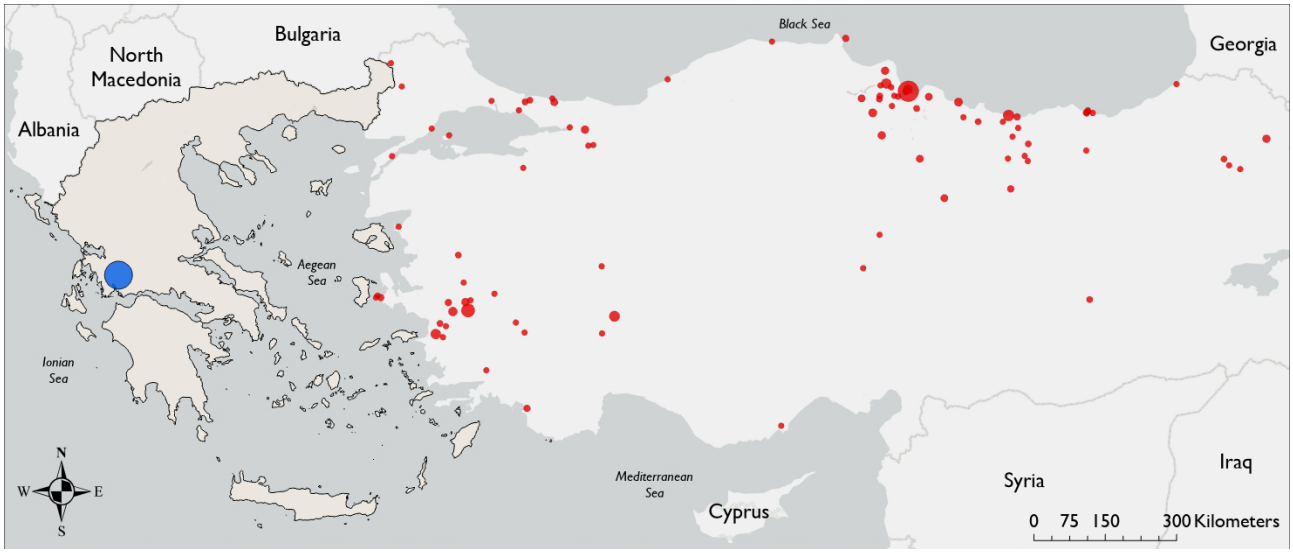
(b) Macedonia and West Trace



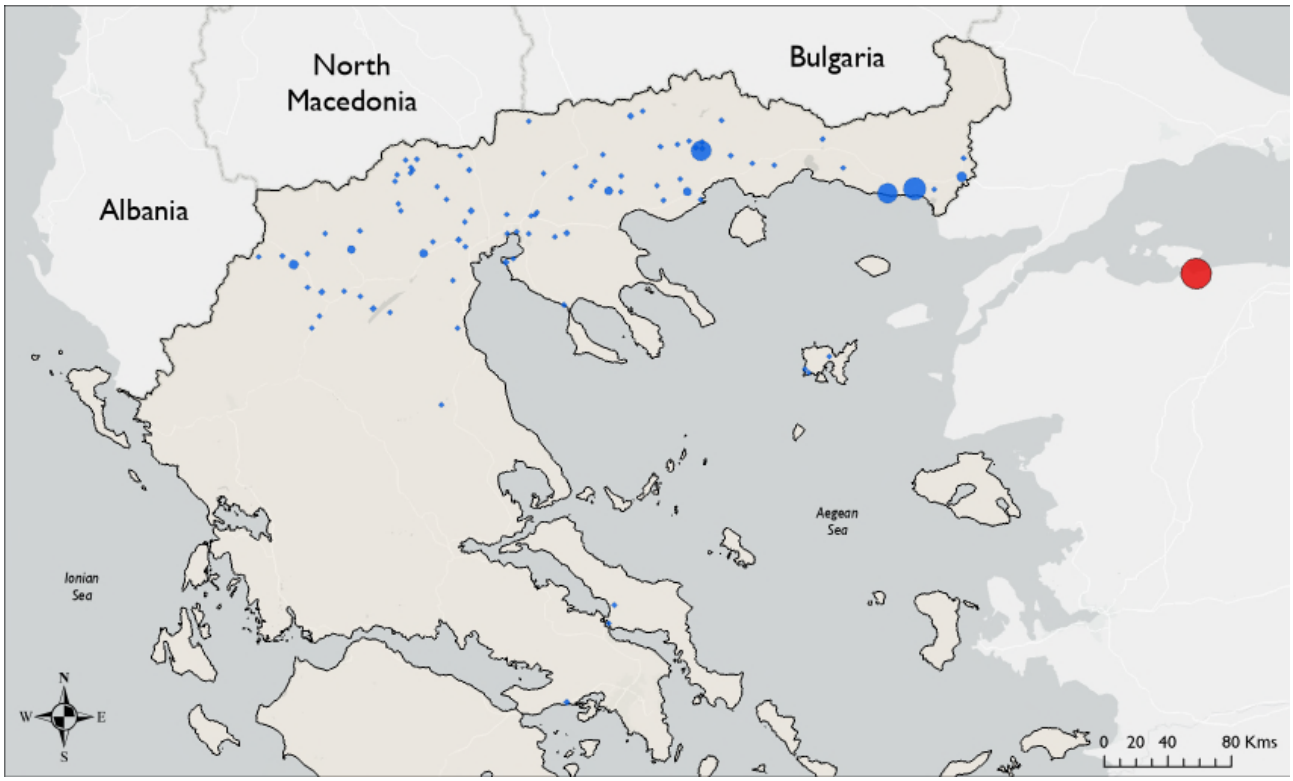
Notes. The panels map 405 predominantly-refugee rural settlements, where the share of refugees in 1928 exceeds 80% by settlement type: 291 former-Muslim refugee settlements, where the imputed share of Muslims in 1920 exceeds 80% (in red dots), and 114 new refugee settlements, where the imputed share of Muslims in 1920 is below 80% (in blue plus signs). Panel *A* plots settlements across Greece. Panel *B* zooms into Macedonia and West Thrace.

Figure A9: Rural Refugee Catalogue Examples: Settlement of Destination and Origin

(a) Origin of Refugees in Agios Konstantinos (Aetolia-Acarnania)



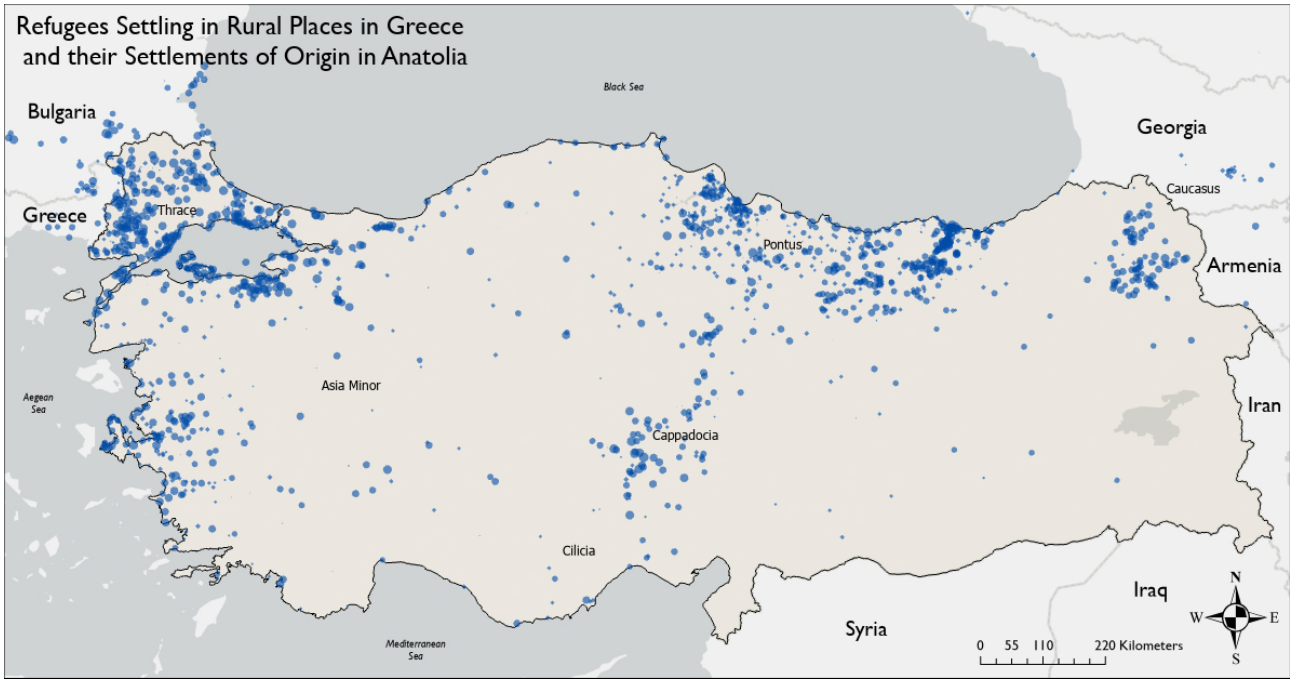
(b) Settlements of Refugees from Panormos (Bandırma, Balikesir)



Notes. The two panels illustrate the origin-destination information of the Rural Refugee Catalog compiled by RSC in the mid-1920s. Panel *A* gives the origin (in contemporary Turkey) of refugees settling in Agios Constantinos (Aetolia-Akarnania). Panel *B* gives the destination settlements of all refugees from Panormos (Band). The dot size presents the (log) number of refugee households from (in) that location.

Figure A10: Settlements of Greek-Orthodox Refugees in Turkey before the Population Exchange

(a) Rural Refugee Catalog

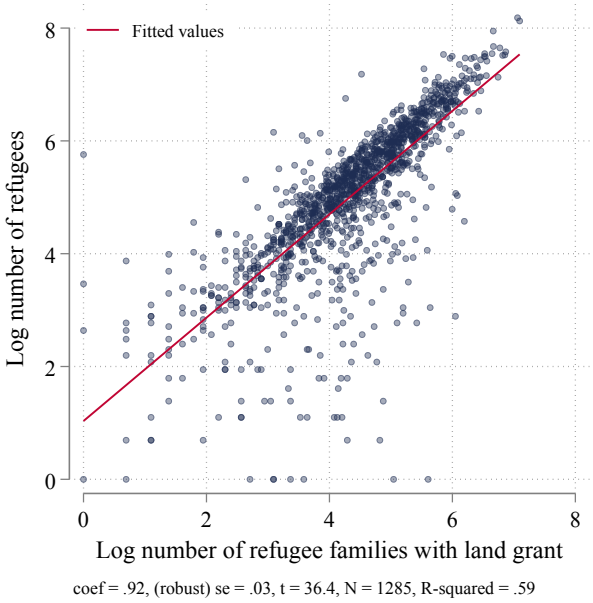


(b) Urban Refugee Catalog

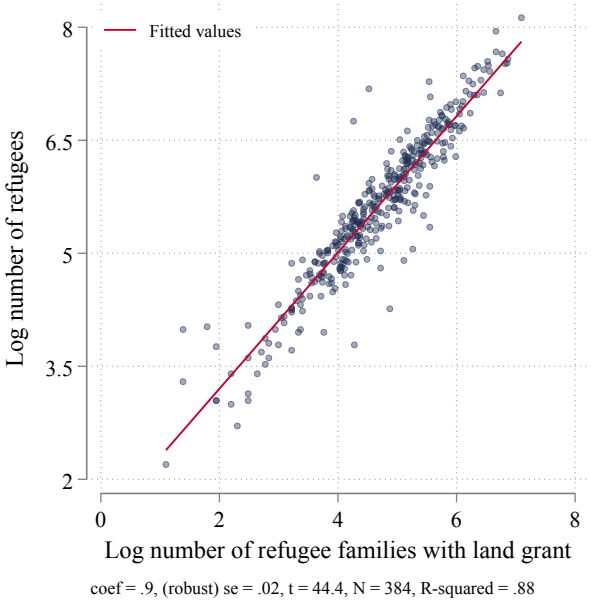
Notes. The figure plots the origin of all Greek refugees in the rural (agricultural) and urban refugee catalogs compiled by the Refugee Settlement Committee in the mid-1920s. The information is based on 211,056 household entries in the rural catalog across 1,850 settlements and 164,617 household entries in the urban catalog across 923 settlements). 1,735 origin settlements are in contemporary Turkey and 115 are outside Turkey. Blue (red) dots indicate the origin settlements of refugees primarily settling in rural (urban) areas in Greece in the mid-1920s. The dot size presents the (log) number of refugee households from that location.

Figure A11: Refugees in the 1928 Population Census and Refugee Families Receiving Land Grants in the Rural Refugee Catalog

(a) All rural settlements with land grant



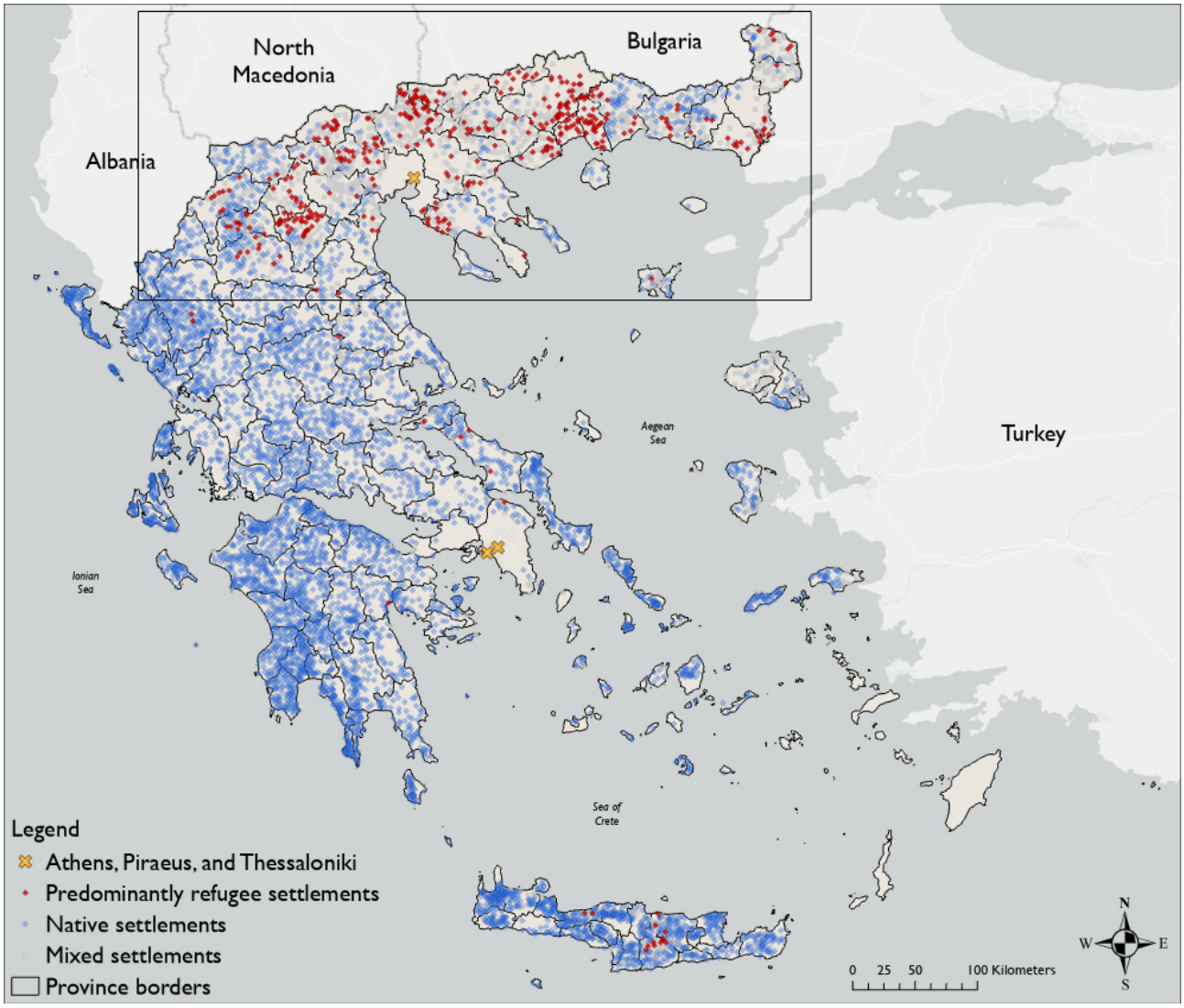
(b) Predominantly refugee settlements with land grant



Notes. The panels plot the OLS correlation between the log number of refugees based on the 1928 Population Census and the log number of refugee families that received a land grant based on the Rural Refugee Catalog for all rural settlements with at least one land grant (Panel *A*) and among rural predominantly refugee settlements, where the share of refugees in 1928 exceeds 80%, with a land grant (Panel *B*).

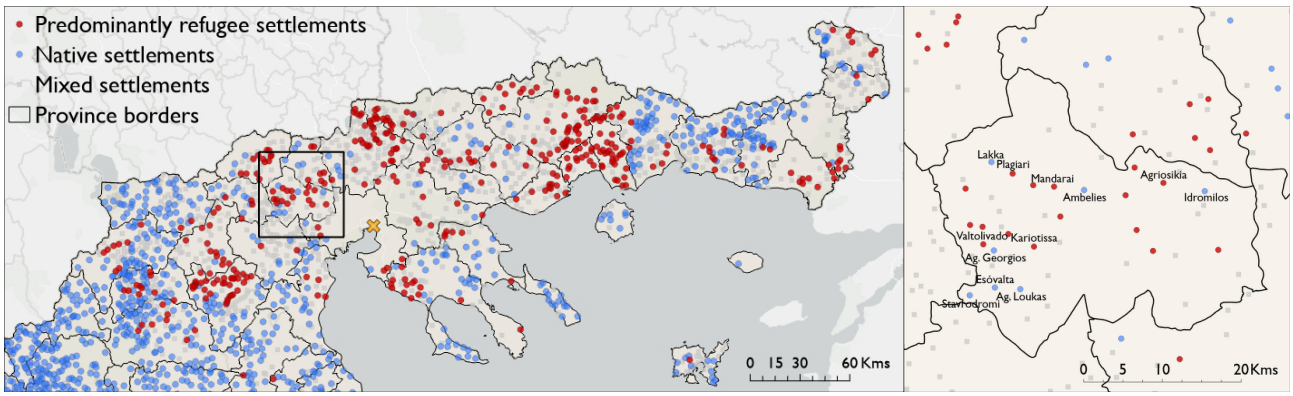
Figure A12: Rural Refugee and Native Settlements in 1928

(a) Greece



(b) Macedonia and West Trace

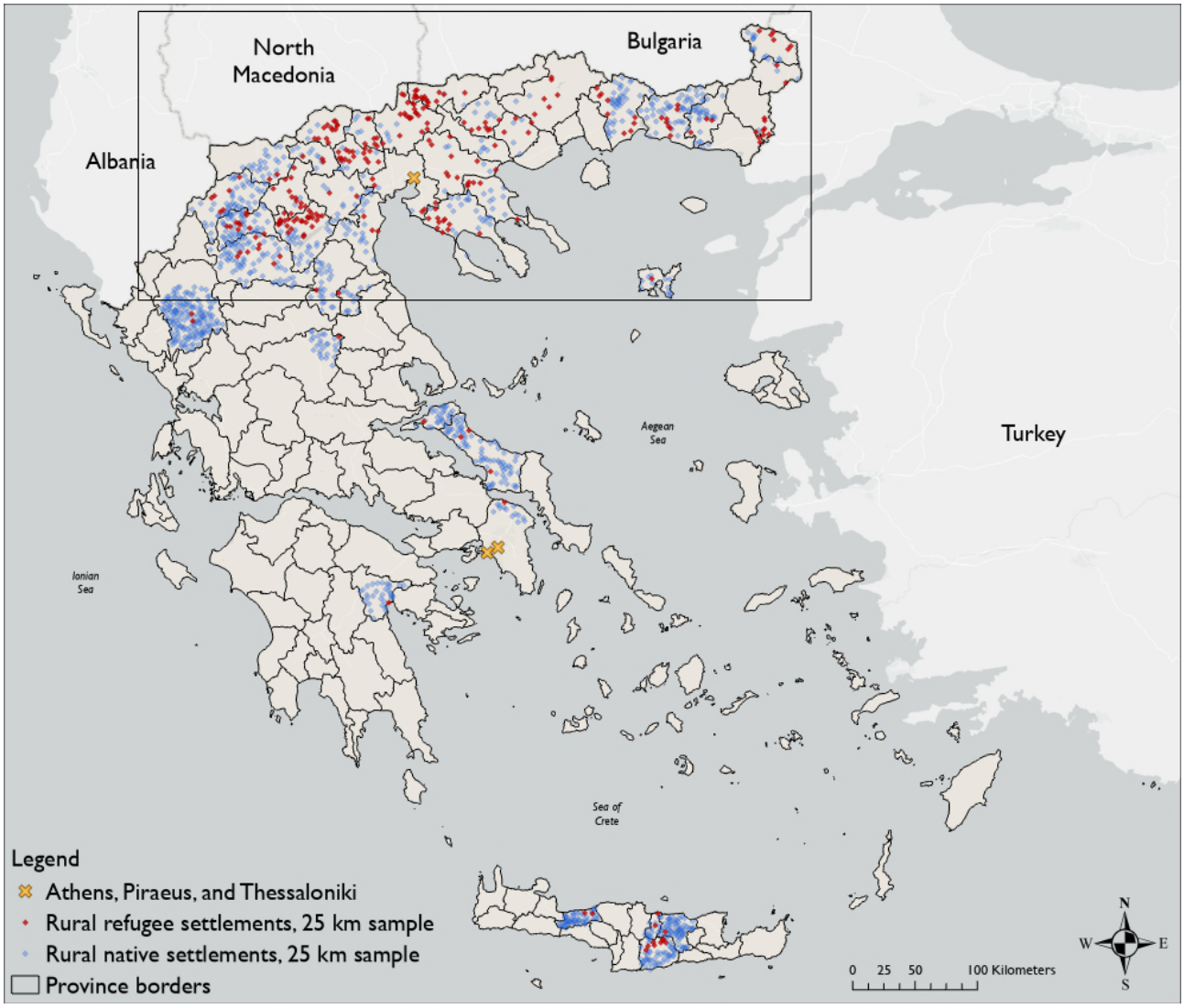
(c) Giannitsa province



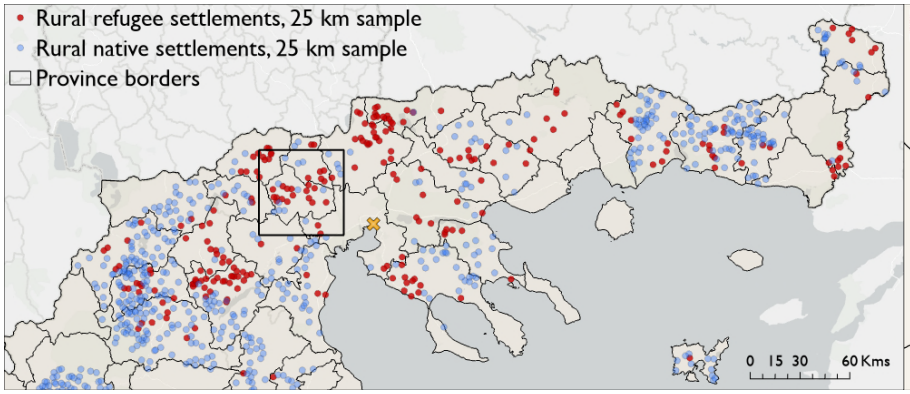
Notes. The panels map 7,769 rural settlements, distinguishing between predominantly-refugee settlements, where the share of refugees in 1928 exceeds 80% (in red dots), and native settlements, where the share of refugees is below 5% (in blue dots), and mixed settlements, where the share of refugees is between 5% and 80% (in gray squares). Panel *A* plots settlements across Greece. Panel *B* zooms into Macedonia and West Thrace. Panel *C* zooms into the area surrounding Giannitsa province in Pella, Central Macedonia.

Figure A13: Local Analysis Sample

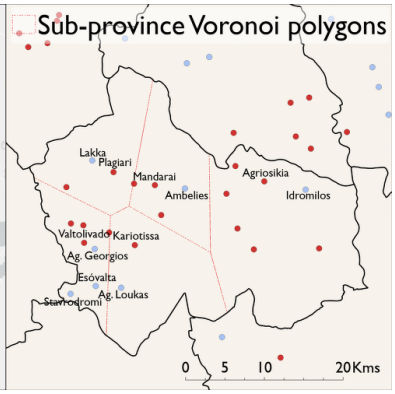
(a) Greece



(b) Macedonia and West Trace

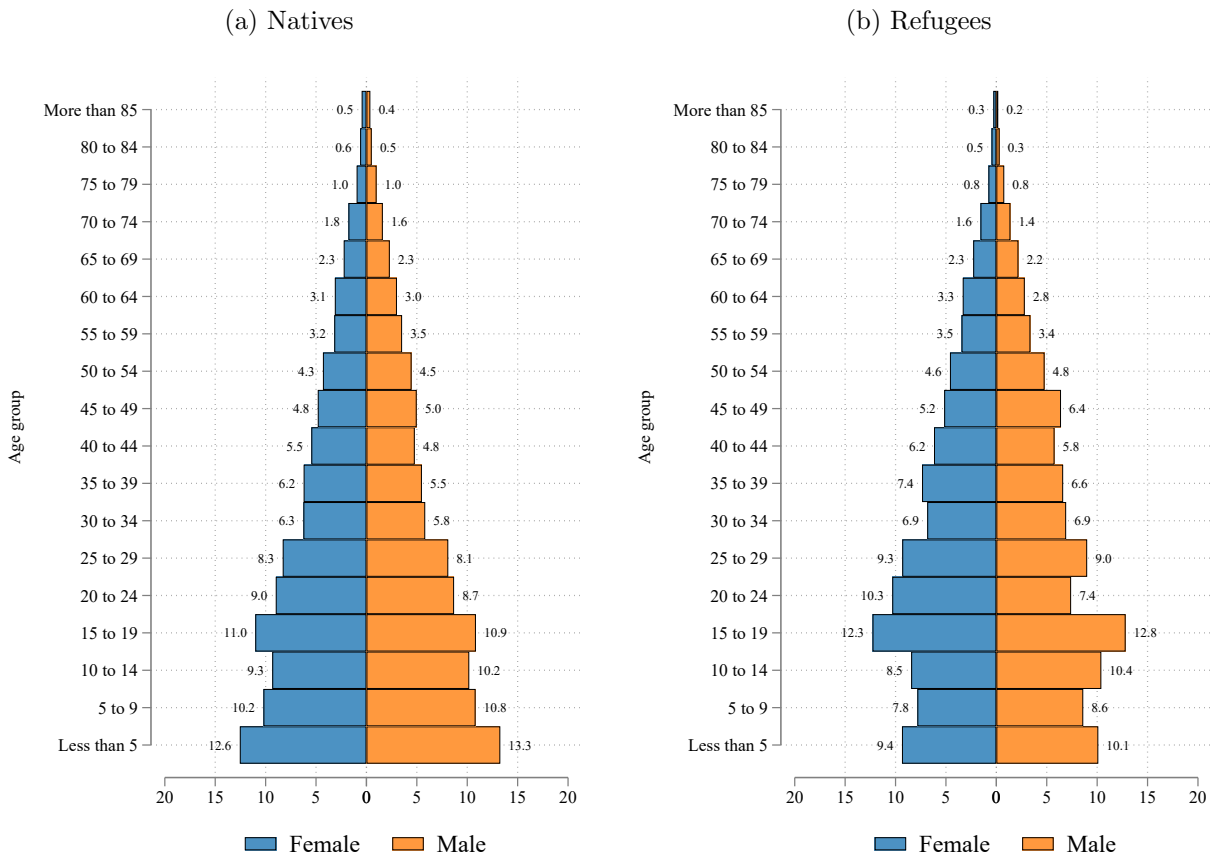


(c) Giannitsa province



Notes. The panels map the native and refugee settlements in the local sample. Each native settlement is matched to at least one nearby refugee settlement within a 25 km radius. In the refugee settlements, the share of refugees in 1928 exceeds 80% (in red dots); in native settlements, the share of refugees is below 5% (in blue dots). Panel *A* plots nearby (within 25 km) refugee and native settlements across Greece. There are 262 refugee and 1,042 native settlements. Panel *B* zooms into Macedonia and West Thrace. Panel *C* zooms into the area surrounding Giannitsa province in Pella, Central Macedonia.

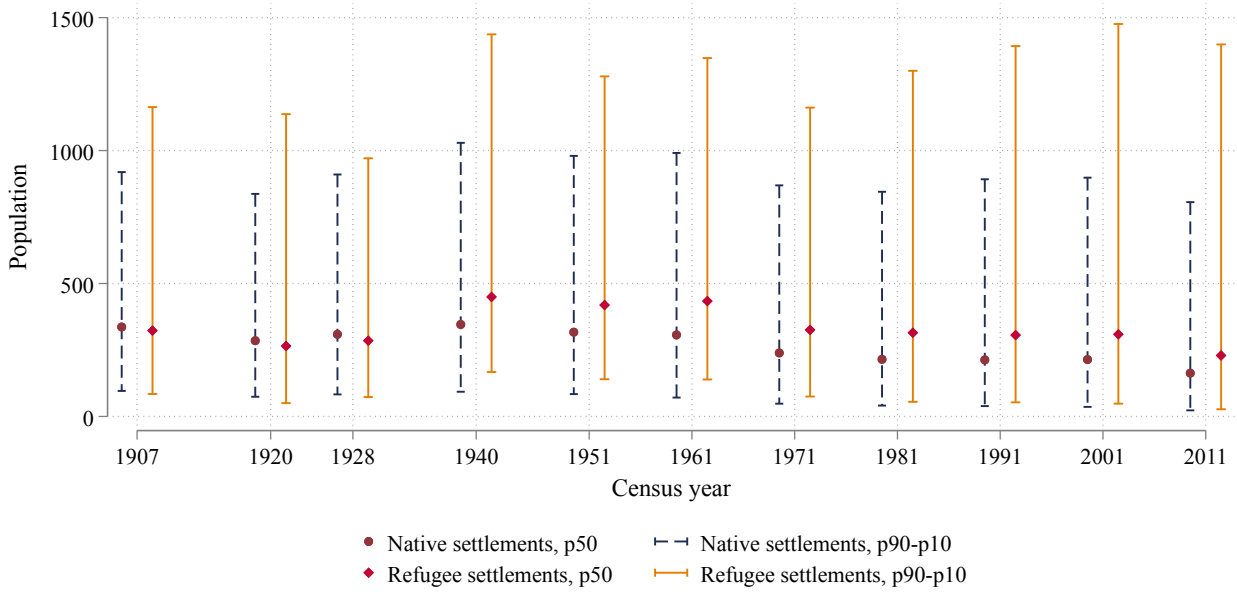
Figure A14: Population Distribution in 1928 Across Refugees and Natives



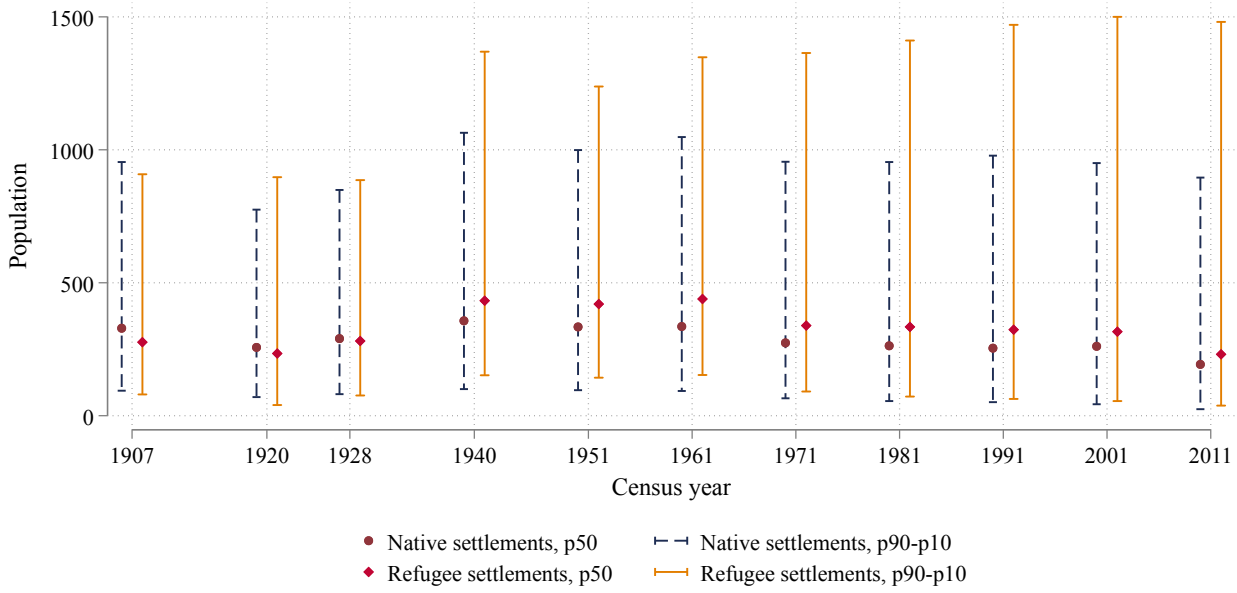
Notes. The figures plot the age distribution by gender among refugees and natives using the 1928 General Population Census.

Figure A15: Population Dynamics. Rural Native and Refugee Settlements

(a) All rural native and refugee settlements



(b) Local sample



Notes. The figures plot the evolution of the median and the 90th- and 10th-percentiles of the population in predominantly refugee and rural native settlements between 1907 and 2011, as recorded in the respective Censuses. Panel *A* plots the population statistics for all rural native and refugee settlements. Panel *B* focuses on the native and refugee settlements in the local sample.

A.2 Correlates of Refugee Share and Balancedness

Here, we report on the correlates of refugee settlement across the Greek countryside, providing details to the discussion in Section 3.3 of the main paper.

Refugee Presence Appendix Table A2 reports the OLS estimates associating the share of refugees in the total population in 1928 with various geographic, location, and ethnographic features across 7,769 rural settlements. First, by far, the most significant correlate of refugee settlement is the share of Muslims forced to leave Macedonia—and, to a lesser extent, Crete and Thessaly, with the Convention for the Population Exchange of 1923. A ten percentage point increase in the share of Muslims in 1920, just before the population exchange, is associated with an eight percentage point higher share of refugees in 1928. The share of Muslims explains about 60% of the variation in refugee share. When conditioning on geography and location features (columns (2)–(5)) or exploit within prefecture (column (4)) or province (column (5)) variation, the coefficient on the share of Muslims in 1920 drops somewhat from 0.83 to 0.67, implying, however, still considerable effects. Second, among the geographic features, the log distance to reclaimed marshlands enters with a highly significant coefficient, supporting the historical narrative that refugees quite often settled in swamps. Besides, refugees settled mainly in the lowlands, but if anything, in areas with lower-quality land. Appendix Figure A16 illustrates the strong correlation between the share of refugees in 1928 and the share of Muslim populations in the Greek countryside in 1920. Panel *A* provides the bin scatter plot showing the unconditional correlation (column (1), Appendix Table A2). Panel *B* plots the correlation conditional on province constants and other covariates (column (4)).

Balance Tests Appendix Figure A17 and A18 report "balance tests" that examine mean differences in geographic, location, and demographic features between predominantly refugee settlements, where the share of refugees in 1928 exceeded 80%, and native settlements, where the refugees' share in 1928 was below 5%. Appendix Figure A17 gives results in the baseline sample, consisting of 405 refugee and 6,091 native settlements. Appendix Figure gives "balancedness tests" in the local sample (Appendix Figure A13), which comprises 262 refugee and 1,042 settlements. In the local sample, refugee settlements are less than 25 km away than native settlements. All panels give standardized coefficients and 90% confidence intervals from OLS regressions that associate the geographic, ecological, demographic, and locational features to an indicator variable that takes the value of one for predominantly refugee settlements and zero for native settlements. Panel *A* gives unconditional specification estimates. Panel *B* gives estimates conditional on log altitude, a significant correlate. Specification 1 (left) does not include any (additional) controls. Specification 2 includes prefecture (admin-2 unit, *nomos*) fixed-effects. Specification 3 includes province (admin-3 units, *eparchies*) fixed-effects. Specification 4 (in Appendix Figure A17) includes sub-province fixed effects, based on Voronoi polygons restricting the comparisons to native and refugee settlements within a 25 km radius. In addition to the significantly higher share of Muslims in refugee settlements, there are significant differences in proximity to the

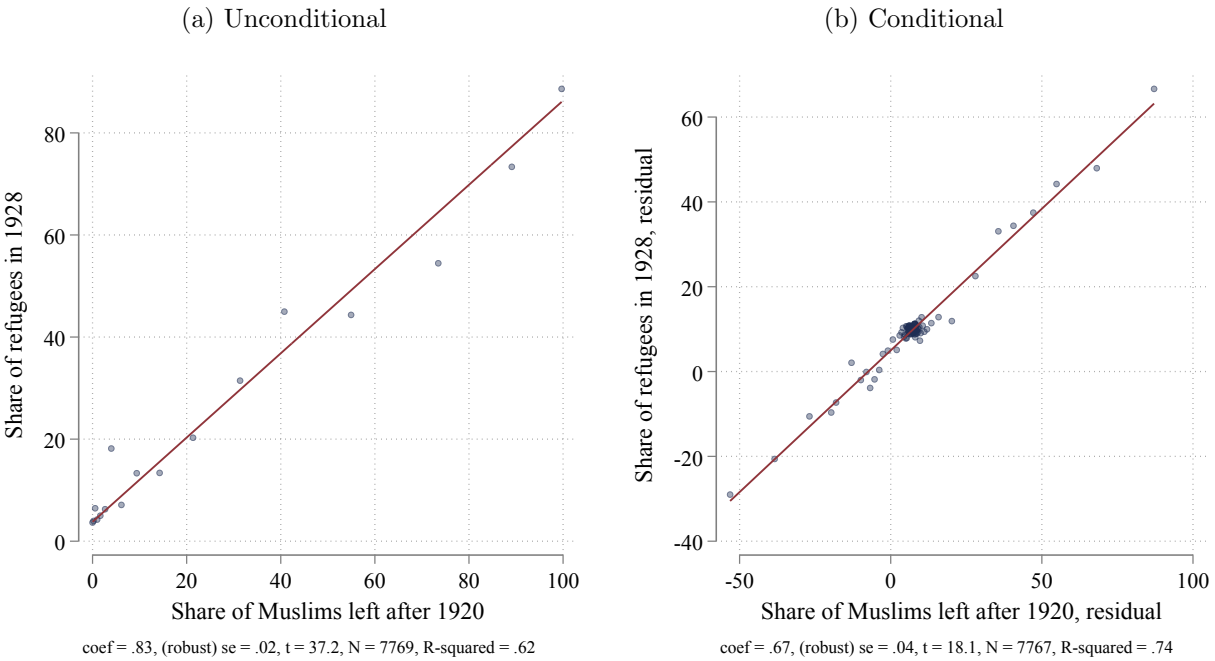
coast, marshlands, and crop suitability. Refugee settlements had a smaller population before the exchange, as many refugees settled in new villages, built by the RSC (see the discussion in Section 2). However, these differences shrink and turn statistically indistinguishable from zero when comparing settlements in the same prefecture or province, and when we condition on log altitude. When we restrict comparisons to proximate settlements in the local sample (Appendix Figure A17), geographical, location, and ecology differences are tiny.

Table A2: Correlates of Refugee Share across Rural Settlements

Dependent Variable:	Refugee share in 1928					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Muslims in 1920, imputed	0.825***	0.738***	0.682***	0.668***		
	(0.022)	(0.026)	(0.033)	(0.037)		
Log population in 1920		-2.317***	-2.271***	-2.290***	-1.776***	-1.862***
		(0.341)	(0.342)	(0.356)	(0.393)	(0.361)
Monastery in 1920		0.152	0.083	0.030	0.037	0.219
		(1.390)	(0.880)	(0.851)	(0.952)	(0.892)
Log distance to reclaimed marshlands		-2.675***	-1.058***	-1.001**	-1.099	-0.573
		(0.341)	(0.398)	(0.406)	(0.677)	(0.685)
Log distance to province capital		-0.294	-0.434	-0.318	0.358	0.135
		(0.538)	(0.444)	(0.517)	(0.800)	(0.841)
Log distance to coast		0.761	0.417	0.673	0.872	0.534
		(0.466)	(0.384)	(0.520)	(0.741)	(0.853)
Log distance to railway stations		-1.009***	-1.122***	-0.951*	-0.974*	-0.919
		(0.282)	(0.342)	(0.506)	(0.546)	(0.591)
Log crop suitability		-1.594**	-1.194*	-1.824**	-0.130	0.172
		(0.666)	(0.643)	(0.737)	(1.326)	(1.607)
Log altitude		-4.327***	-2.657***	-3.061***	-0.729	-1.776
		(0.683)	(0.656)	(0.902)	(1.056)	(1.167)
Log average precipitation		-12.087***	-1.927	-0.214	-18.187**	-4.843
		(1.874)	(2.997)	(4.257)	(8.199)	(5.115)
Log average temperature		-18.978***	-0.597	3.990	0.604	3.571
		(6.116)	(4.662)	(5.314)	(7.930)	(9.297)
R-squared	0.620	0.710	0.738	0.750	0.482	0.544
Observations	7,769	7,767	7,767	7,767	7,767	7,767
Average outcome	10.0	10.0	10.0	10.0	10.0	10.0
Prefecture FE			✓		✓	
Province FE				✓		✓

Notes. The table gives OLS estimates associating the share of refugees after the Asia Minor Catastrophe (in the fall of 1922) across 7,769 rural settlements, as recorded in the 1928 General Population Census, to various location, geographic, and demographic features. The specifications in columns (3) and (5) include prefecture fixed-effects and the specifications in column (4) and (6) province fixed-effects. Standard errors are clustered at the province level. *, **, and *** denote statistical significance at the 10%, 5%, and 1%, confidence level.

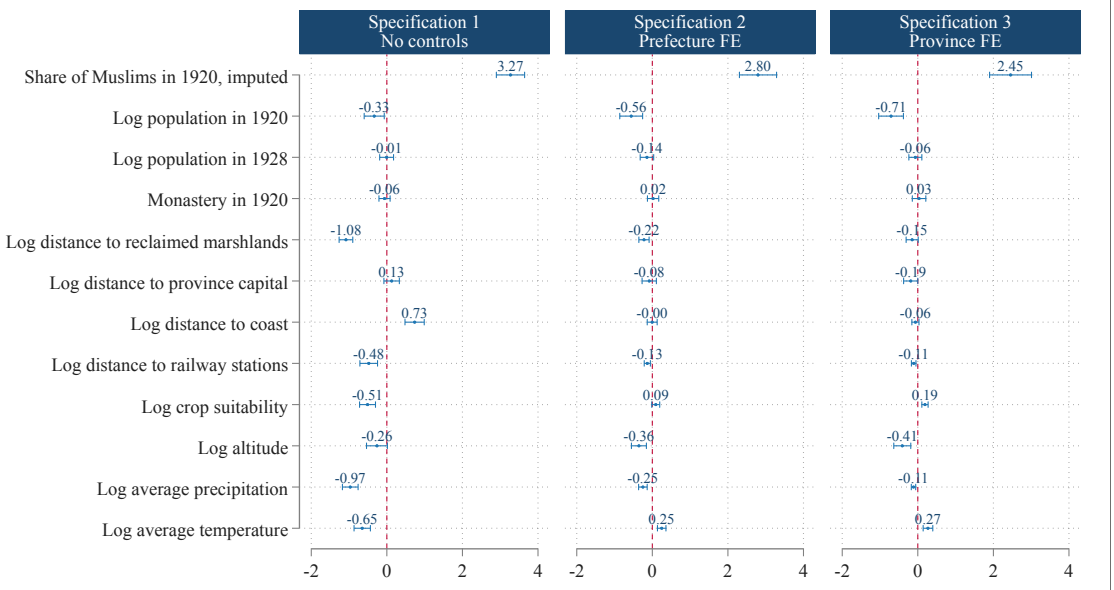
Figure A16: Binscatter Share of Refugees in 1928 and the Share of Muslims in 1920



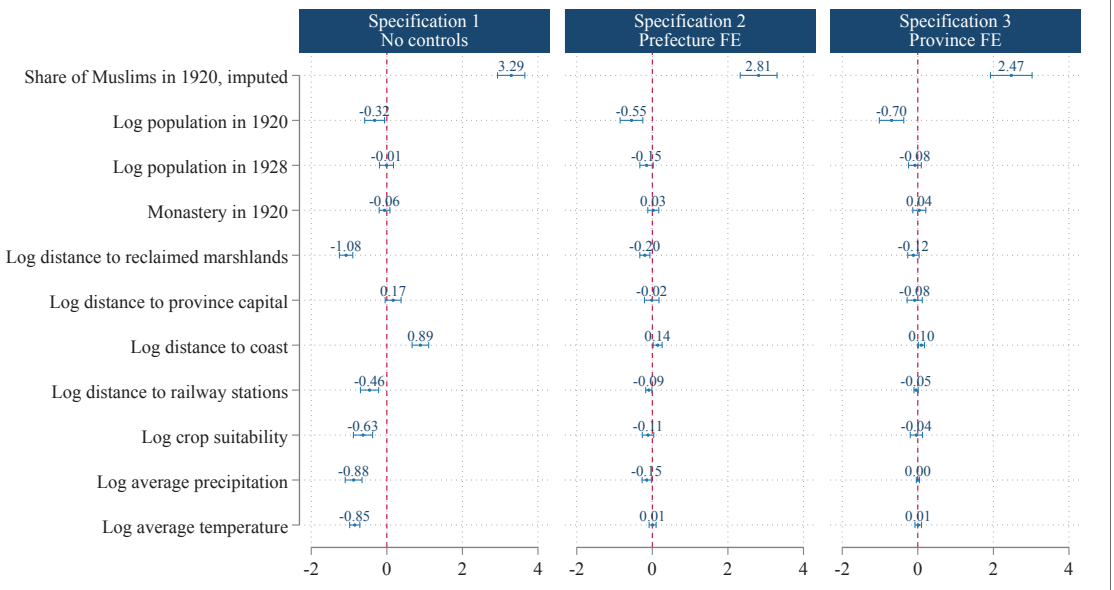
Notes. The panels provide the bin scatter plot across 100 bins visualizing the correlation between the share of Muslims departing after 1920 and the share of refugees in 1928. Panel *A* presents he unconditional correlation presented in column 1 of Appendix Table A2 while Panel *B* presents the correlation conditional on province fixed effects and covariates presented in column 4. Standard errors are clustered at the province level.

Figure A17: Balancedness across All Rural Native and Refugee Settlements

(a) Unconditional



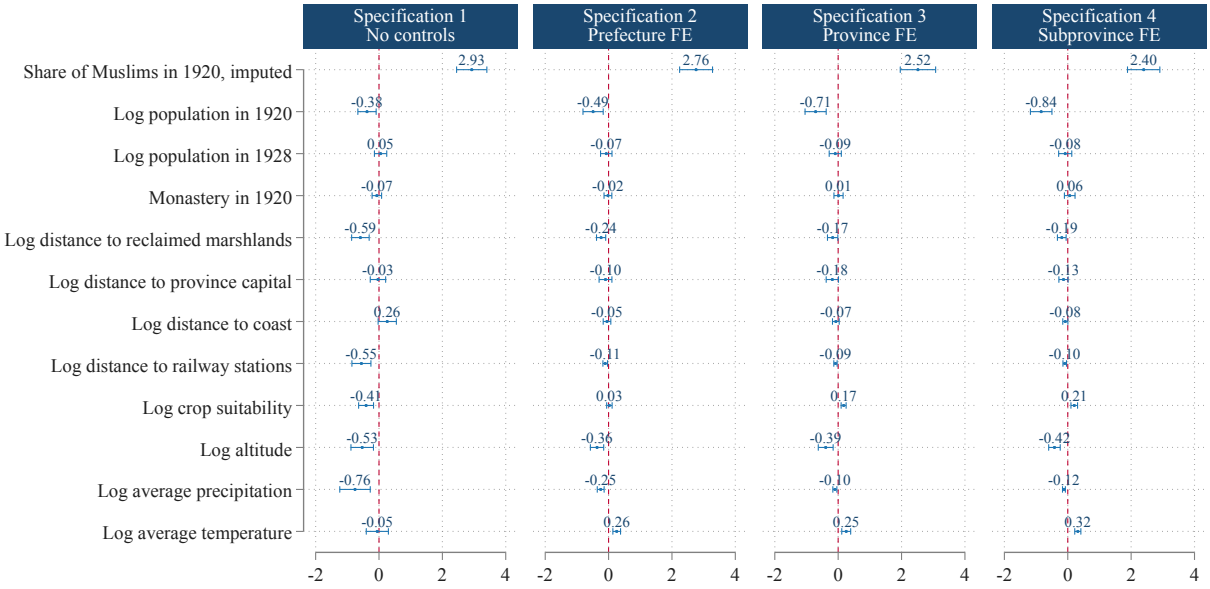
(b) Conditional on log altitude



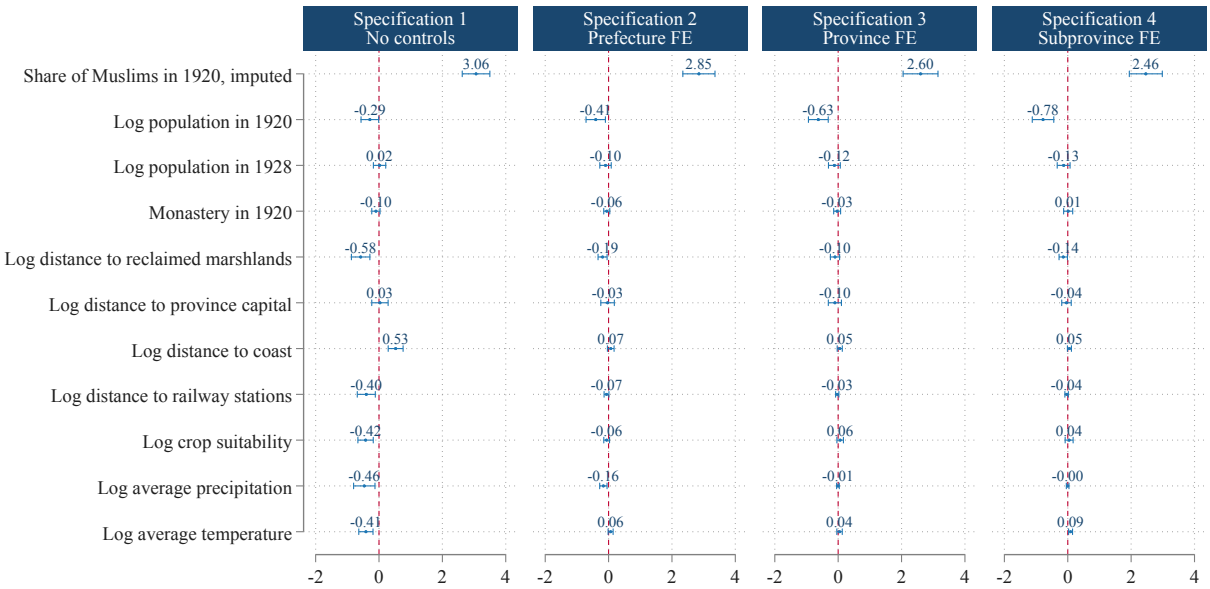
Notes. The panels explore whether various geographic, location, and population features differ between predominantly refugee and native settlements in the baseline sample of rural settlements. All specifications in both Panels give standardized coefficients and 90% confidence intervals obtained from regressing standardized settlement characteristics on a dummy variable indicating predominantly refugee settlements, where the share of refugees in 1928 exceeds 80%; the indicator equals zero for native settlements, where the share of refugees in 1928 is below 5%. Panel *A* presents unilateral unconditional specifications. Panel *B* gives the regression estimates conditioning on settlements log altitude. Specification 1 does not include any controls. Specification 2 conditions on prefecture fixed effects. Specification 3 conditions on province fixed-effects. Standard errors are clustered at the province level.

Figure A18: Balancedness across Rural Native and Refugee Settlements in the Local Sample

(a) Unconditional



(b) Conditional on log altitude



Notes. The panels explore whether various geographic, location, and population features differ between predominantly refugee and native settlements in the local sample of rural settlements. All specifications in both Panels give standardized coefficients and 90% confidence intervals obtained from regressing standardized settlement characteristics on a dummy variable indicating predominantly refugee settlements, where the share of refugees in 1928 exceeds 80%; the indicator equals zero for native settlements, where the share of refugees in 1928 is below 5%. Panel *A* presents unilateral unconditional specifications. Panel *B* gives the regression estimates conditioning on settlements log altitude. Specification 1 does not include any controls. Specification 2 conditions on prefecture fixed effects. Specification 3 conditions on province fixed-effects. Specification 4 conditions on sub-province fixed effects, based on Voronoi polygons. Standard errors are clustered at the province level.

B Displacement and Human Capital

Appendix Section B complements the descriptive and regression analyses linking human capital to forced displacement in Section 4 of the paper. First, we report descriptive evidence on (trends in) literacy rates for refugees and non-refugees before the displacement. Second, we give additional results on the association between (changes in) education and refugee lineage. Third, we report further evidence and various robustness checks, revealing the stability of the baseline estimates to alternative definitions of refugee and native settlements and different cutoffs of proximity in the local analysis. Fourth, we report results linking the share of refugees in 1928 to subsequent human capital across all rural settlements: besides predominantly refugee and autochthonous ones, we include mixed refugee-native ones. Fifth, we report gender-specific estimates.

B.1 Descriptive Patterns. Initial Differences and Dynamics

Initial Differences (1971-1981 Censuses) Appendix Figure B1 plots the mean literacy (in Panel *A*) and primary school completion rates (in panel *B*) for those born before 1902 from predominantly refugee and native settlements using the 1971 and 1981 General Population Censuses. The figure, which complements the descriptions in Figure 4 in the main paper, shows the lower levels of education for Greeks in predominantly refugee settlements compared to those from native settlements.

Pre-Displacement Differences in Education. Rural Refugees and Non-Refugees Appendix Figure B2 plots the differences in the literacy of refugees and natives by birth-cohort, as recorded in the 1928 General Population Census. Panel *A* plots the differences in literacy among the rural population across the country. Panel *B* zooms into Macedonia and West Thrace as eight out of ten rural refugees settled there. Both panels illustrate the drop in literacy for the Greek-Orthodox communities in Asia Minor following the repression of the Ottoman Empire, which accelerated after 1908 and further intensified during the Greco-Turkish hostilities after the Great War.

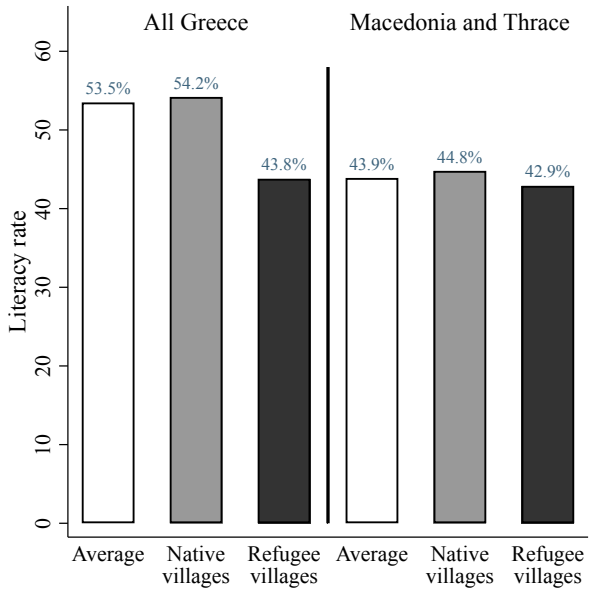
Pre-Trends in Literacy. Refugees and Non-Refugees Appendix Table B1 reports specifications exploring whether the literacy increase during the 19th century is similar for refugees and non-refugees. We retrieve from the 1928 General Population Census five-year birth-cohort level statistics on literacy rates for residents in the countryside and then regress the literacy rates on a linear time trend, an indicator variable for refugees, and an interaction between the time trend and the refugee indicator. As revealed by the trend estimate, literacy has increased steadily for natives and refugees (when the latter resided in Anatolia). The refugee indicator enters with a negative coefficient, telling of the refugees' lower literacy by about 2.5–5.7 percentage points. However, the coefficient on the refugee dummy drops in absolute value and becomes statistically indistinguishable from zero in Macedonia and Thrace when we omit the cohorts impacted by the Greco-Turkish warfare and the ethnic violence against minorities in the Ottoman Empire after the Young Turks rose in 1908. Most importantly, the coefficient on the interaction term that estimates the differential trend between natives and refugees

is small and statistically insignificant, both when we analyze the patterns across the country (columns (1)–(2)) and when we zoom into Macedonia and West Thrace (columns (3)–(4)).

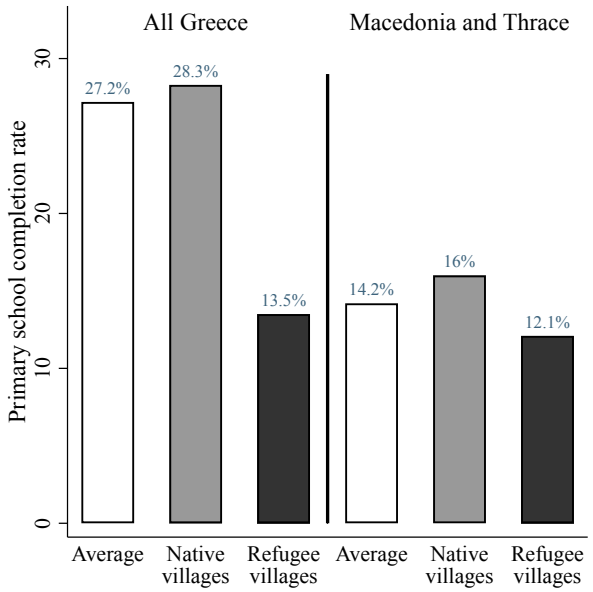
Educational Dynamics. Rural Refugee and Native Settlements Appendix Figure B3 plots the dynamics of schooling years (left) and junior high school completion rates (right) for Greeks living in predominantly refugee and native settlements, pooling the micro samples of the 1971, 1981, 1991, 2001, and 2011 General Population Censuses. Panel *A* considers all predominantly refugee and native settlements, while Panel *B* focuses on Macedonia and West Thrace. The descriptive analysis uncovers two patterns. First, in line with the tabulations of the 1928 General Population Census, human capital was lower in predominantly refugee settlements compared to native settlements for those who started schooling prior to the population exchange. Those born outside Greece before WWI who settled in the countryside after the population exchange had lower completion rates of junior high school. When we zoom into Macedonia and West Thrace, differences become muted because the human capital of non-refugees was lower than that of natives in Central and Southern Greece. Second, upon arriving in Greece, mainly in Macedonia and West Thrace, refugees in the countryside caught up and soon surpassed the natives' education. Second, the refugee-native differences in junior high completion rates for individuals in Macedonia and West Thrace in the 1950s, 1960s, and 1970s are a staggering 15 percentage points.

Figure B1: Education across Refugee and Native Settlements. Greeks born before 1902. 1971–1981 Censuses

(a) Literacy Rates



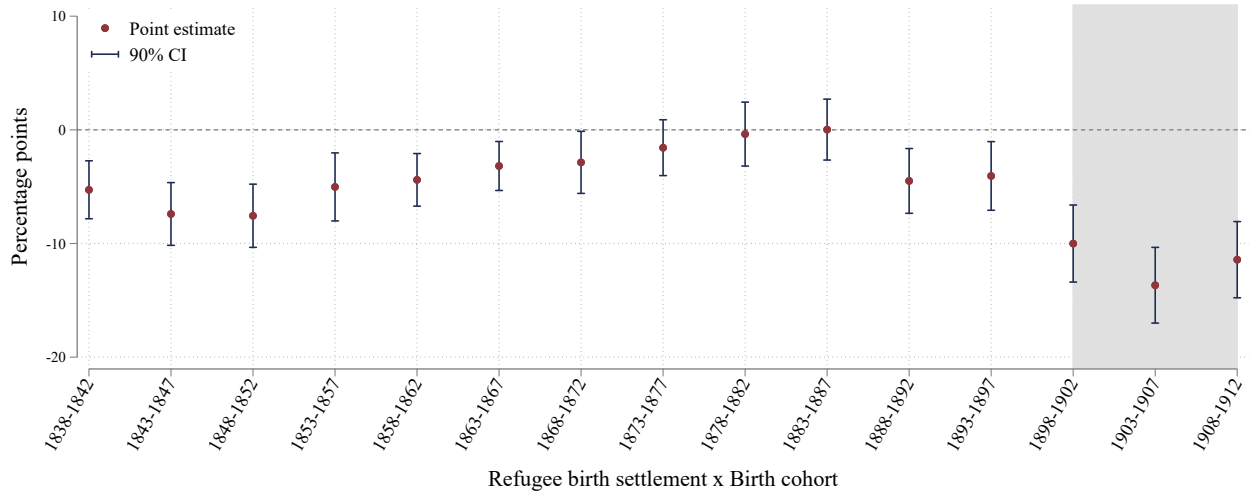
(b) Primary School Completion



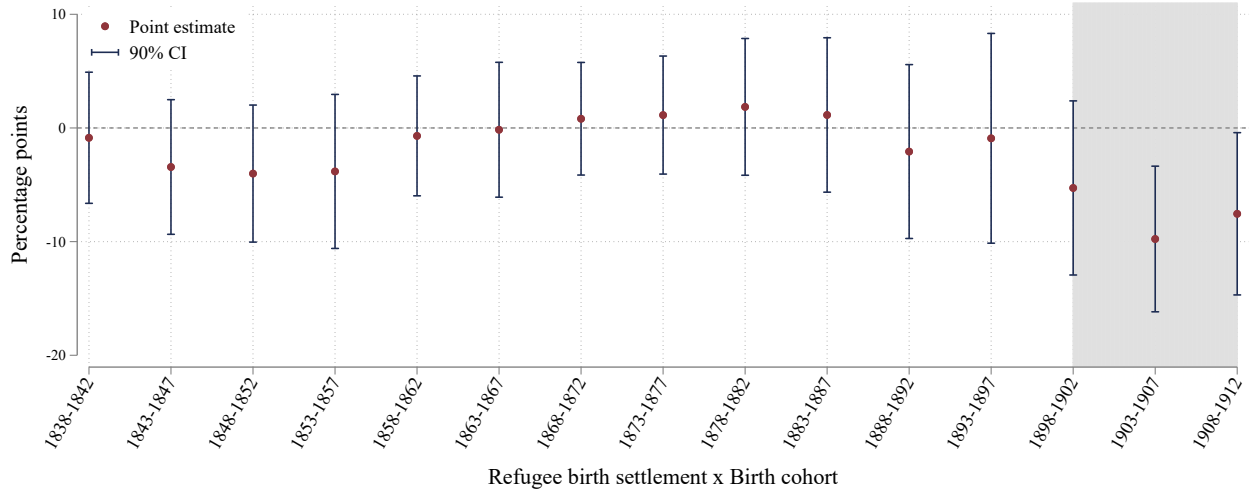
Notes. Panel *A* plots literacy rates for all Greeks (white bars) born before 1902 living in the countryside (rural), in native settlements (gray bars), and in refugee settlements (black bars). Panel *B* plots primary school completion rates for all Greeks (white bars) born before 1902 living in the countryside (rural), in refugee settlements (gray bars), and in native settlements (black bars). Data come from the 1971 and 1981 General Population Censuses. Both panels report means across Greece (left), and Macedonia and West Thrace (right).

Figure B2: Gap in Literacy between Refugees and Natives in Rural Settlements in 1928

(a) All Greece



(b) Macedonia and West Thrace



Notes. All panels plot the difference in literacy rate in 1928 between rural natives and refugees across provinces, retrieved from OLS estimates, weighted by province population. Panel *A* plots the unconditional difference across all 141 provinces. Panel *B* plots the difference across all provinces in Macedonia and West Thrace, as 79.8% of rural refugees settled there. Literacy rates are retrieved from the 1928 General Population Census. Gray indicates the cohorts affected by Ottoman repression, the Balkan Wars, the Great War, and the Greco-Turkish War. Standard errors are clustered at the prefecture level.

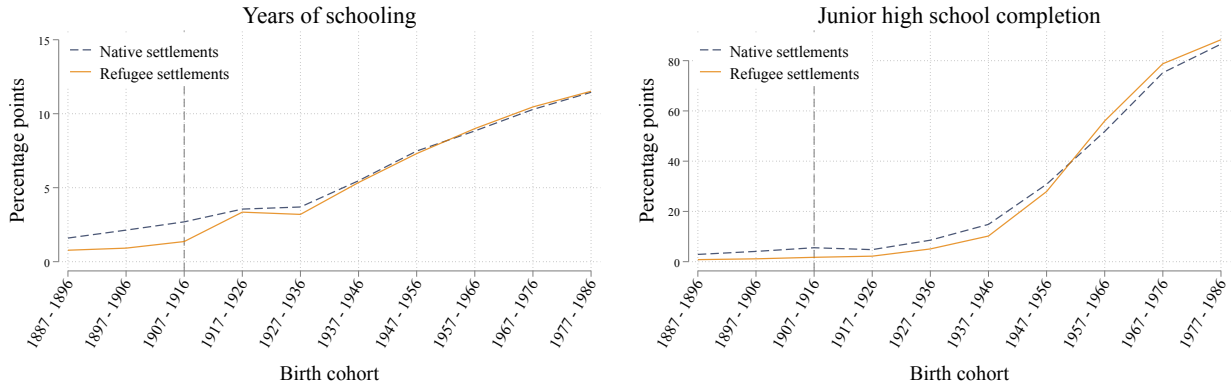
Table B1: Testing for Differential Trends across Refugees and Natives before the Population Exchange

Sample:	All sample		Macedonia & Thrace	
Cohorts:	All	Excluding 1898–1912	All	Excluding 1898–1912
	(1)	(2)	(3)	(4)
Refugee	-5.704***	-2.757**	-2.527***	-0.197
	(1.055)	(1.084)	(0.866)	(0.914)
Time trend	3.039***	2.441***	2.979***	2.542***
	(0.225)	(0.229)	(0.185)	(0.199)
Refugee × Time trend	-0.287	0.436	-0.286	0.289
	(0.263)	(0.301)	(0.238)	(0.296)
R-squared	0.957	0.958	0.966	0.960
Observations	30	24	30	24

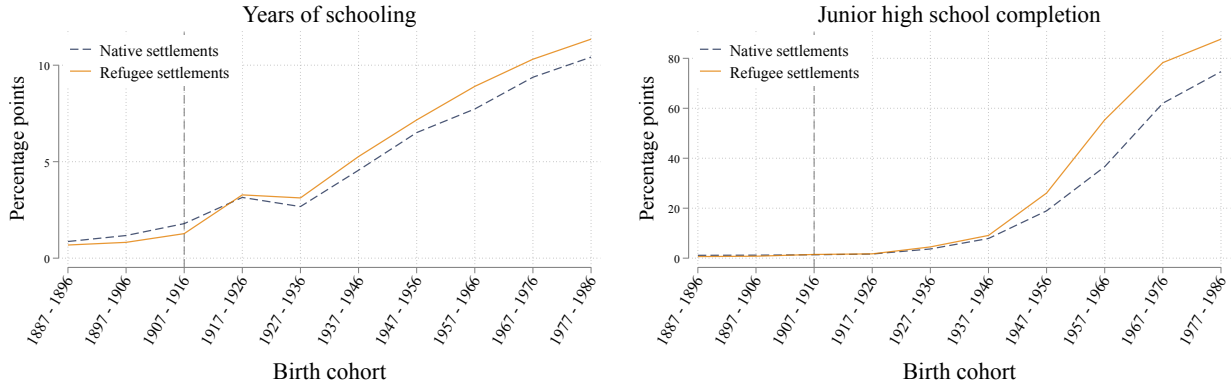
Notes. The table gives estimates testing for differential trends in literacy rates between refugees and natives, as observed in the 1928 General Population Census. All columns report OLS estimates associating average literacy rates, by 10-year birth cohorts and refugee status, to a linear time trend, an indicator variable that takes the value of one for refugees and zero for non-refugees, and an interaction between the linear time trend and the refugee indicator. Columns (1) and (2) consider all provinces; columns (3) and (4) focus on Macedonia and Western Thrace provinces. Columns (2) and (4) exclude the cohorts born between 1898 and 1912, affected by Ottoman repression, the Balkan Wars, the Great War, and the Greco-Turkish War. *, **, and *** denote statistical significance at the 10%, 5%, and 1%, confidence level.

Figure B3: Human Capital Dynamics across Rural Refugee and Native Settlements. 1971–2011 Censuses

(a) All provinces



(b) Macedonia and West Thrace



Notes. All panels plot the evolution in years of schooling and junior high school completion rates across all predominantly refugee settlements (in solid orange) and native (non-refugee) settlements (in dashed blue) across 10-year birth cohorts. Panel A considers the rural predominantly refugee and native settlements across all provinces, while Panel B restricts the sample to those in Macedonia and West Thrace. Data come from the 1971, 1981, 1991, 2001, and 2011 Population Censuses. Predominantly refugee settlements are those where the share of refugees from Anatolia in 1928 exceeds 80%; native settlements are those where the share of refugees in 1928 is lower than 5%.

B.2 Additional Results

Here, we report additional results on the link between forced displacement and subsequent human capital investments across the Greek countryside (Sections 4.2 and 4.3).

B.2.1 Difference-in-Difference Estimates. Local Sample

Appendix Figure B4 plots the event-study graphs showing the within-settlement correlation between human capital and living in refugee settlements across birth cohorts in the local sample; the comparison is between predominantly refugee villages and towns (where the refugee share exceeded 80% in 1928) and native settlements (where the share of refugees was below 5%) in the same province, no more than 25 kilometers away; regression equation (1). While standard errors are wider compared to the baseline sample estimates, as the sample is considerably smaller, the patterns are similar to the within-province estimates in Figure 6. First- and second-generation refugees have a four to five percentage points higher likelihood of completing primary education. Over time, second-generation refugees born in the 1940s–1970s have a four to six percentage points higher chance of completing junior high school; second- and even third-generation refugees born after WWII have a two to six percentage point higher likelihood of completing high school, while the latest cohorts of Greeks born in refugee settlements have a one to two percentage points higher propensity finishing tertiary education.

B.2.2 Conditioning on Current Residence

Appendix Figure B5 plots cohort-specific estimates on the refugee indicator that shed light on the differential dynamics of schooling between Greeks born in predominantly refugee and native settlements, using the 2011 General Population Census that gives details on the respondent's family birthplace. These estimates thus complement the regression estimates in Table 2. Two results emerge. First, years of schooling are considerably higher for Greeks from refugee settlements compared to those born in native settlements in the same prefecture (admin-2 level unit), province (admin-3 level unit), and sub-province level in the local sample. Second, while the cohort-specific estimates are somewhat noisy, they illustrate that the educational gain of second-, third-, and fourth-generation refugees is considerable and stable over time.

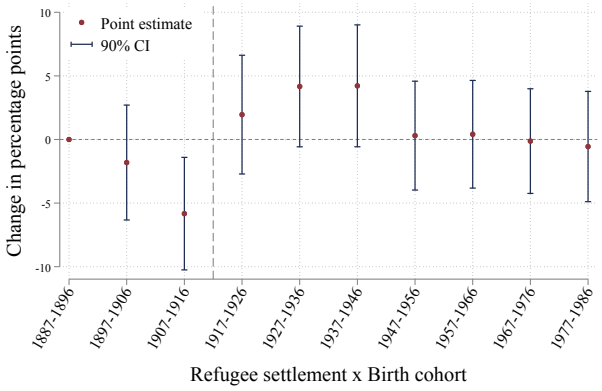
Appendix Table B2 associates primary school, junior high school, high school, and tertiary education completion rates with the refugee indicator that takes the value of one for Greeks from predominantly refugee settlements and zero for individuals from native settlements, conditioning on municipality of residence in 2011 (regression equation (2)). Panel A gives the coefficients across Greek citizens, aged 25–90, born in all rural predominantly refugee and native settlements; Panels B and C report the estimates for individuals who continue residing in their birthplace (non-migrants) and for Greeks who moved out (internal migrants), respectively. The results complement the ones in Table 2–Panel A in the paper, which looks at schooling years. The educational gain of second-, third-, and fourth-generation refugees is present both when looking at Greeks living in their families' home-

towns (Panel B) and at internal migrants (Panel C), who mainly moved to Athens, Thessaloniki, and Piraeus. Besides, the patterns are similar to the difference-in-difference specifications (using all post-1971 Censuses) in Table 1 in the main paper.

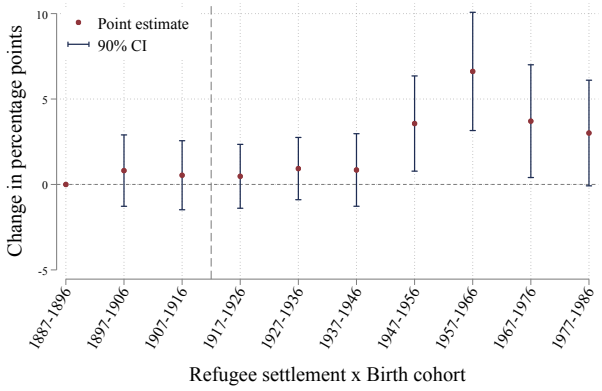
Appendix Figure B6 plots the corresponding cohort-specific estimates that allow exploring the dynamics of educational attainment rates for the two groups. We condition on the municipality of residence. The patterns illustrate the dynamics of uprootedness over three generations. First- and second-generation refugees have a much higher propensity, about ten percentage points, to complete primary schooling than Greeks born in native settlements (Panel A). As primary school completion reaches 100% for all Greeks, then the educational advantage of Greeks born in refugee settlements emerges for junior high and high school completion (Panels B and C). Gradually, the refugee-native educational differences also become evident, strengthening over time, when looking at college/university completion (Panel D). In addition, the patterns in Appendix Figure B6 also echo the difference-in-difference estimates in the main paper (Figure 6).

Figure B4: Education Dynamics across Refugee and Native Settlements. Local Sample. 1971–2011

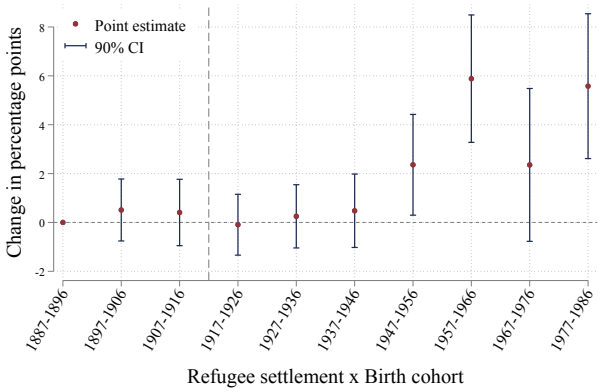
(a) Primary school completion



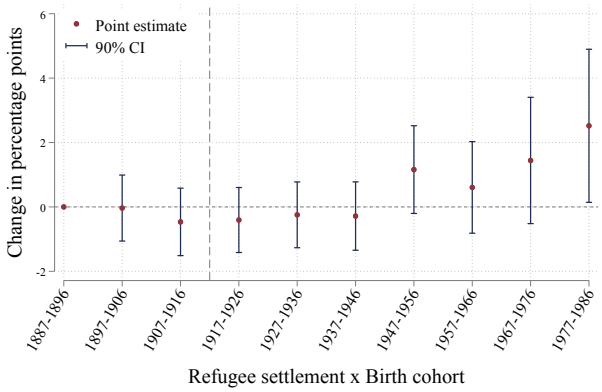
(b) Junior high school completion



(c) High school completion



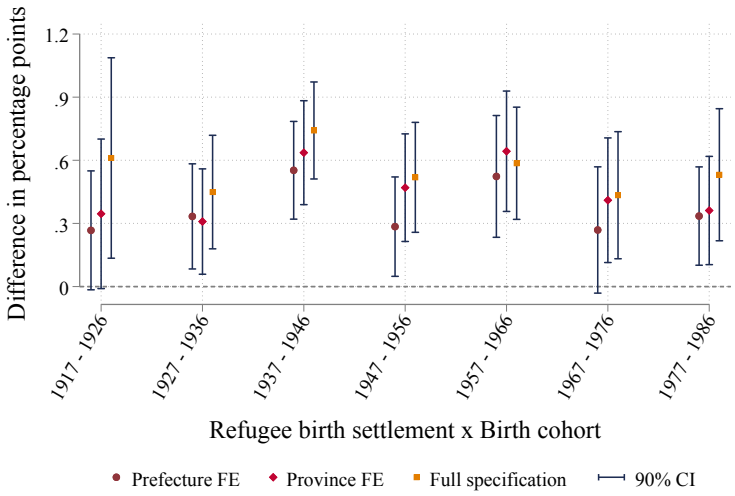
(d) Tertiary education completion



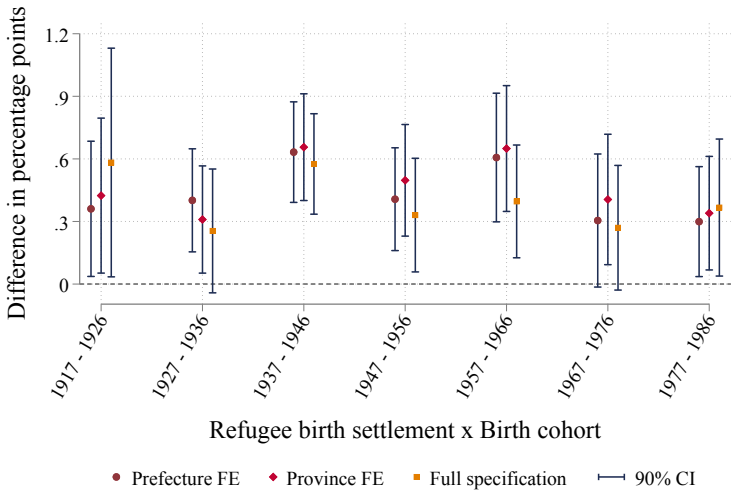
Notes. The panels plot the event-study OLS (linear probability model) estimates obtained from associating indicator variables for completing primary school (in Panel *A*), for competing junior high school (in Panel *B*), for competing high school (in Panel *C*), and for completing tertiary education (in Panel *D*) on a birth-cohort varying dummy variable indicating whether an individual resides in a predominantly refugee settlement, where the share of refugees in 1928 exceeds 80%. Estimation is in the local sample that comprises native and refugee settlements in the same sub-province unit (Voronoi polygon) within a 25 km radius. All specifications include settlement fixed effects, interactions of census-year and birth-cohort fixed effects, interactions of sub-province fixed effects and birth-cohort fixed effects, interactions of individual characteristics (age and gender) and birth-cohort fixed effects, and interactions of various settlement characteristics and birth-cohort fixed effects. The settlement characteristics are: log distance to the coast, the provincial capital, railway stops, and reclaimed marshlands; geographic characteristics: crop suitability, altitude, precipitation, and temperature; and the log population in 1920, an indicator for settlements with the monastery in 1920. The 1897–1906 birth cohort serves as the excluded cohort. Standard errors are clustered at the settlement-of-residence level.

Figure B5: Years of Schooling across Rural Refugee and Native Birth Settlements, cond. on Residence (2011)

(a) All Settlements



(b) Local Sample



Notes. The panels present the cohort-level estimates obtained from regressing years of schooling on a dummy variable indicating whether an individual was born in a predominantly refugee settlement, where the share of refugees in 1928 exceeds 80%, for each birth-cohort separately; regression equation (2). Panel *A* gives estimates across all rural refugee and native birth settlements. Panel *B* focuses on the local analysis sample, where refugee birth settlements are matched to native ones no more than 25 km away in the same sub-province unit. Maroon circles and red diamonds plot the estimates obtained from parsimonious specifications exploiting within-birth-prefecture (*nomos*) and within-birth-province (*eparchia*) variation; these specifications only condition on interactions of the birth prefecture or the birth province fixed effects. Orange squares plot the estimates obtained from a rich (full) specification that also controls for individual characteristics (age and gender) and birth settlement characteristics, and also for sub-province fixed effects in Panel *B*. Standard errors are clustered at the birth-settlement and the municipality-of-residence levels.

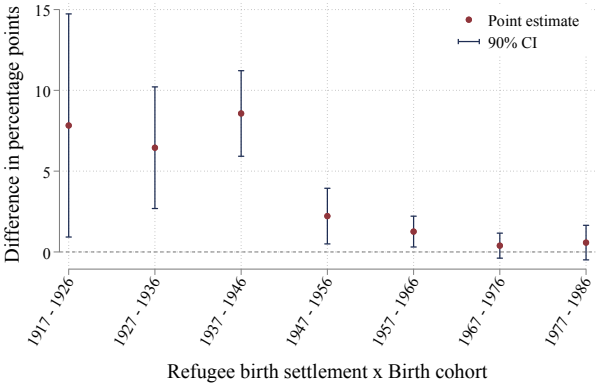
Table B2: Educational Attainment across Native and Refugee Birth Settlements, cond. on Residence (2011)

Dependent Variable:	Completed education level:							
	Primary school		Jun. high school		High school		Tertiary education	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A:	**All sample**							
Born in a refugee settlement	4.365***	3.961***	6.788***	5.401***	5.327***	4.076***	2.191***	1.423**
	(0.804)	(0.795)	(1.311)	(1.240)	(1.176)	(1.134)	(0.601)	(0.568)
R-squared	0.297	0.311	0.323	0.360	0.227	0.267	0.059	0.088
Observations	549,480	549,474	549,480	549,474	549,480	549,474	549,480	549,474
Average outcome in native settlements	86.2	86.2	45.6	45.6	33.4	33.4	11.4	11.4
# Refugee observations	44,115	44,114	44,115	44,114	44,115	44,114	44,115	44,114
# Refugee birth settlements	357	357	357	357	357	357	357	357
# Native observations	505,365	505,360	505,365	505,360	505,365	505,360	505,365	505,360
# Native birth settlements	5,788	5,788	5,788	5,788	5,788	5,788	5,788	5,788
# Birth provinces for FE	141	141	141	141	141	141	141	141
# Residence-municipality FE		1,021		1,021		1,021		1,021
Panel B:	**Non-migrants**							
Born in a refugee settlement	3.780***	4.688***	5.238***	4.613***	3.901***	3.632**	1.858***	0.988
	(0.981)	(1.247)	(1.296)	(1.745)	(1.199)	(1.631)	(0.637)	(0.606)
R-squared	0.321	0.331	0.402	0.415	0.296	0.309	0.087	0.095
Observations	243,038	243,005	243,038	243,005	243,038	243,005	243,038	243,005
Average outcome in native settlements	83.2	83.2	40.0	40.0	26.8	26.8	6.7	6.7
# Refugee observations	18,259	18,255	18,259	18,255	18,259	18,255	18,259	18,255
# Refugee birth settlements	351	351	351	351	351	351	351	351
# Native observations	224,779	224,750	224,779	224,750	224,779	224,750	224,779	224,750
# Native birth settlements	5,688	5,688	5,688	5,688	5,688	5,688	5,688	5,688
# Birth provinces for FE	140	140	140	140	140	140	140	140
# Residence-municipality FE		943		943		943		943
Panel C:	**Internal migrants**							
Born in a refugee settlement	3.977***	3.043***	6.762***	4.996***	5.572***	3.961***	1.732**	1.154
	(0.905)	(0.884)	(1.542)	(1.424)	(1.421)	(1.317)	(0.805)	(0.740)
R-squared	0.286	0.298	0.298	0.328	0.215	0.249	0.066	0.088
Observations	289,213	289,208	289,213	289,208	289,213	289,208	289,213	289,208
Average outcome in native settlements	88.8	88.8	50.9	50.9	39.4	39.4	15.4	15.4
# Refugee observations	22,674	22,674	22,674	22,674	22,674	22,674	22,674	22,674
# Refugee birth settlements	355	355	355	355	355	355	355	355
# Native observations	266,539	266,534	266,539	266,534	266,539	266,534	266,539	266,534
# Native birth settlements	5,704	5,704	5,704	5,704	5,704	5,704	5,704	5,704
# Birth provinces for FE	140	140	140	140	140	140	140	140
# Residence-municipality FE		1,020		1,020		1,020		1,020
Cohort × Individual controls	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Birth settlement controls	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Birth province FE	✓	✓	✓	✓	✓	✓	✓	✓
Residence municipality FE		✓		✓		✓		✓

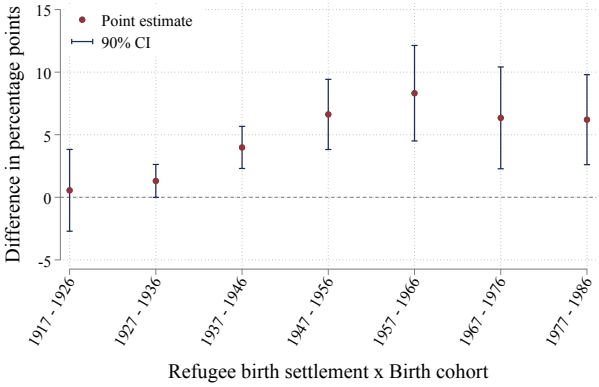
Notes. The Panels report OLS estimates associating being born in a predominantly refugee settlement and educational attainment across all rural predominantly refugee and native birth settlements. The dependent variable is primary school completion in columns (1)–(2), junior high school completion in columns (3)–(4), high school completion in columns (5)–(6), tertiary education completion in columns (7)–(8). Panel *A* considers all residents of rural refugee and native birth settlements, while Panel *B* restricts estimation to individuals who reside in their birth settlement and who did not migrate abroad, and Panel *C* restrict estimation to those who reside in a municipality other than their birth settlement. All reported regressions control for the interactions of the birth-province fixed effects and birth-cohort fixed effects, the interactions of birth-settlement characteristics and birth-cohort fixed effects and the interactions of individual characteristics (age and gender) and birth-cohort fixed effects. Columns (2), (4), (6), and (8) control for municipality-of-residence fixed effects. Birth settlement characteristics are: log distance to the coast, the provincial capital, railway stops, and reclaimed marshlands; geographic characteristics: crop suitability, altitude, precipitation, and temperature; and the log population in 1920, an indicator for settlements with the monastery in 1920. Standard errors are clustered at the birth-settlement and the municipality-of-residence levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1%, confidence level.

Figure B6: Educational Attainment across Rural Native and Refugee Birth Settlements, cond. on Residence (2011)

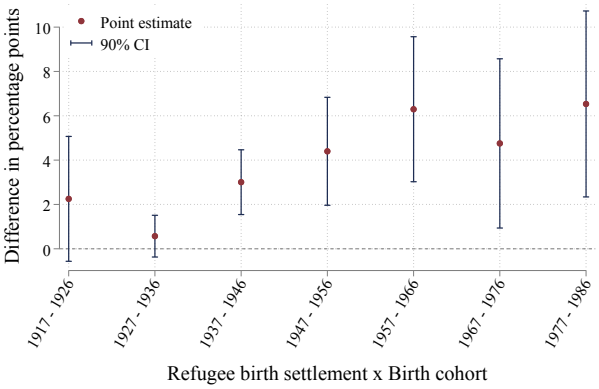
(a) Primary school completion



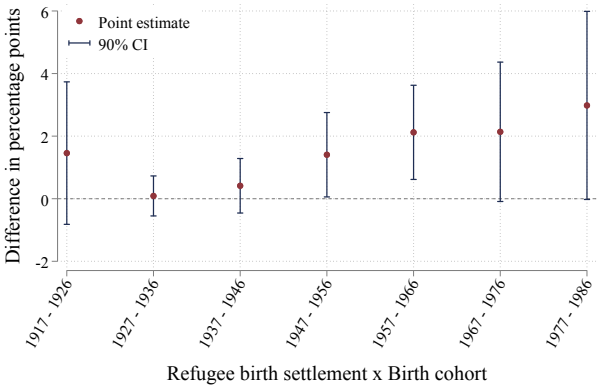
(b) Junior high school completion



(c) High school completion



(d) Tertiary education completion



Notes. The panels present the cohort-level estimates obtained from regressing indicator variables for completing primary school (in Panel *A*), for competing junior high school (in Panel *B*), for competing high school (in Panel *C*), and for completing tertiary education (in Panel *D*) on a dummy variable indicating whether an individual resides in a predominantly refugee settlement, where the share of refugees in 1928 exceeds 80%, for each birth-cohort separately. The analysis considers all rural refugee and native birth settlements. All panels control for birth-province fixed effects, birth-settlement characteristics, and individual characteristics (age and gender). Birth settlement characteristics are: log distance to the coast, the provincial capital, railway stops, and reclaimed marshlands; geographic characteristics: crop suitability, altitude, precipitation, and temperature; and the log population in 1920, an indicator for settlements with the monastery in 1920. Standard errors are clustered at the birth-settlement and the municipality-of-residence levels.

B.3 Further Evidence

B.3.1 Sectoral Employment

Appendix Table B3 reports conditional on residence estimates (regression equation 1), which also account for individuals' sector of employment. The sample consists of respondents between the ages of 25 and 55 in the 2011 General Population Census (25% micro sample). We focus on this age interval as employment drops considerably among Greeks after the mid-fifties. Column (1) gives the estimates on the refugee indicator that takes the value of one for respondents from predominantly refugee settlements (where the share of refugees in 1928 exceeded 80%) and zero for native settlements (where the refugee share was below 5% in 1928) among 25-55-year-olds. When we condition on the respondents' current residence, in (2), the coefficient on the refugee indicator is somewhat smaller than the estimate in the main paper, estimated across 25-90-year-olds (0.53 vs 0.58). In columns (3)-(4), we augment the specification with broad sector of employment fixed effects (15 industries), which, conservatively, we allow to differ across birth cohorts. In columns (5)-(6), we use the most granular sectoral disaggregation of the 2011 Cesnus that distinguishes across 260 sectors. The estimate in (6) suggests that third and fourth-generation refugees have about 0.3 years more schooling than Greeks from native settlements in the same *eparchia*, even when looking among Greeks working in the same industry and living in the same municipality. The coefficient is approximately a third lower than the full sample estimate (0.53) in column (2). This suggests that sectoral employment is a significant underlying factor behind refugee families' higher propensity to invest in their children's education.

B.3.2 Spatial Spillovers

Appendix Table B4 reports difference-in-difference specifications testing for spatial spillovers from refugee to native settlements. For each native settlement, we calculate the share of refugees (as recorded in the 1928 General Population Census) within 5 km, considering all predominantly refugee and native villages in the baseline sample. Columns (1) and (3) report OLS estimates across all rural native settlements, while columns (2) and (4) focus on rural native settlements in the local analysis sample. The specifications in (1) and (2) show that the presence of refugees within 5 km does not affect natives' years of schooling after the resettlement of refugees (controlling for the log of total population among refugee and native settlements within 5 km). Likewise, the estimates in (3) and (4), where we control for the share of refugees in 1928 and log the total population within a 5 to 25 km radius, also yield little evidence of spillovers.

B.3.3 Gender

Below, we report descriptive evidence, difference-in-difference specifications (regression equation (1)), and conditional-on-residence (regression equation (2)) results distinguishing by respondent's gender.

Before-After Mean Differences Appendix Figure B7–Panel *A* reports mean literacy rates for men and women refugees and non-refugees living in the Greek countryside using the breakdown provided by the 1928 General Population Census. These results thus complement the patterns in Figure 4, which does not distinguish by gender. Literacy rates are lower for both men and women refugees as compared to non-refugee men and women across the country (left) and in Macedonia and West Thrace (right); differences in Macedonia and West Thrace were smaller as literacy rates were overall higher in Sterea Ellada, the Peloponnese, the Ionian Sea Islands, and Crete. Panel *B* reports the mean completion rates of junior high school for men and women born in predominantly refugee and native settlements using the 2011 General Population Census. The considerable advantage of natives in literacy before the Population Exchange has disappeared, with junior high completion rates being similar in 2011 for those born in refugee and native settlements (left). Zooming into Macedonia and West Thrace yields even more striking patterns (right). Women and men from predominantly refugee settlements have significantly higher human capital than Greeks from native settlements; the difference in junior high school completion rates is considerable, about ten percentage points for both men and women.

Difference-in-Difference Estimates Appendix Table B5 reports the difference-in-difference specification separately for men (odd-numbered columns) and women (even-numbered columns), complementing the analysis in Table 1 in the main paper. The coefficients for women are higher across all educational attainment categories (and years of schooling). However, statistical tests fail to reject the null hypothesis of equality of coefficients between gender. The somewhat stronger uprootedness effects for women may reflect their higher propensity to work in manufacturing and services (as compared to agriculture) and more substantial returns to schooling in the "modern sector."

Dynamics Appendix Figure B8 plots the cohort-specific estimates of the refugee indicator when regression equations (specification (1) in Panel *A*) and (specification (2) in Panel *B*) are estimated separately for men and women to explore dynamics. While the cohort-specific coefficients are noisy, the dynamic analysis shows that refugee men fare better than men born or living in native settlements in the initial post-displacement decade(s), and, over time, the refugee-native difference in schooling for women becomes more prominent and stable.

B.3.4 Sensitivity Analysis. Refugee and Native Settlements

Cutoffs Appendix Figure B9 reports OLS estimates associating years of schooling with the refugee indicator using alternative classifications for predominantly refugee and native settlements. Panel *A* reports difference-in-difference specifications, which compare the difference in years of schooling in refugee and native rural settlements before and after the forced population exchange (regression equation 1). Panel *B* reports OLS specifications that compare schooling years between those from predominantly refugee and non-refugee settlements residing in the same municipality in 2011 (regression equation (2)). We do two permutations. First, we define refugee settlements as those where, in 1928,

the share of refugees from Anatolia exceeded 50%, 60%, 70%, 80% (baseline), and 90%. Second, we define non-refugee (native) settlements as those where the share of refugees in 1928 is less than 10%, 5% (baseline), and even 1%. The significant educational gain for Greeks born in refugee compared to native settlements is present across all permutations.

Very Local Analysis Appendix Figure B10 gives difference-in-difference specifications (in Panel *A*) and specifications conditioning on residence in 2011 (in Panel *B*) restricting estimation to Greeks living in proximate refugee and native settlements in the same province, using increasingly narrower radii of 20 km, 15 km, 10 km, and 5 km to define the local sample. Even when comparing Greeks living in refugee to native settlements no more than 10 km—or even 5 km, we find a gain for second-, third-, and fourth-generation refugees with about 0.4 to 0.5 extra years of schooling.

Accounting for Spatial Autocorrelation Appendix Figure B11 presents inference robustness tests with alternative means to account the spatial autocorrelation. We compute Conley standard errors correcting for spatial autocorrelation using a uniform kernel, which is shown to approach nominal null rejection rates compared to a bartlett kernel (Conley (1999); Bester, Conley and Hansen (2011)). Following the suggestion of Colella et al. (2019), which provides Monte Carlo simulations showing that the null rejection rates approaches nominal rates when the error correction threshold approaches the underlying data generating process, we present Conley standard errors for several distance thresholds, ranging from 25 km to 200 km, using the Stata command `acreg` (Colella et al., 2023). Panels *A* and *B* plot the 90% confidence intervals obtained by correcting the standard errors of the difference-in-difference estimates for spatial autocorrelation in the baseline and the local samples, respectively. Panels *C* and *D* give the 90% confidence intervals adjusted for spatial autocorrelation across predominantly refugee and native birth settlements, conditioning on the current municipality of residence. The coefficients' statistical significance is not sensitive to alternative ways of accounting for spatial autocorrelation.

B.3.5 All Rural Settlements

Difference-in-Difference Estimates Appendix Table B6 reports difference-in-difference specifications, estimated across 7,507 rural settlements (stable geographic units), which associate the various educational outcomes to the share of refugees in 1928 We estimate the regression equation (1) using the share of refugees in 1928 rather than the binary refugee indicator on the RHS. Settlements with a higher share of refugees experienced a more substantial increase in schooling after the Population Exchange. Column (1) suggests that a rise in refugee share of 25 percentage points is associated with an extra 0.1 schooling years and a 1.0–1.5 percentage points higher likelihood of completing primary or junior high school. The estimates with the refugee share imply similar effects to the ones in the paper, where we limit comparisons between those born and living in predominantly refugee and native settlements. Moving from a settlement without any refugees to an exclusive refugee one in the same

province is associated with an increase of about 0.4 years of schooling.

Cond. on Residence Estimates Appendix Table B7 presents results from a variant of regression equation (2), replacing the refugee indicator with the share of refugees in 1928 across all rural birth settlements. So the sample now includes all types of rural birth settlements, predominantly refugee, native and mixed. We do so using the 2011 General Population Census, which, besides residence, also reports an individual's birthplace. Exploring information from Greeks who may have moved from their birth settlement—and conditioning on residence in 2011 yields similarly highly significant estimates of the refugee share on educational attainment.

Dynamics Appendix Figure B12 plots the event-study graphs showing the within-settlement correlation correlation between schooling and refugees' share in 1928 across birth cohorts. Panels *A* and *B* consider years of schooling as the outcome variable, while Panels *C*–*E* focus on the highest completed level of education. Panel *A* plots the cohort-specific estimates on the share of refugees of a parsimonious specification that conditions on settlement fixed effects, interactions between prefecture constants and birth-cohort fixed effects, and interactions between census and birth-cohort fixed effects. The coefficients in Panels *B*–*E* stem from a specification with a rich set of controls (e.g., the interactions of province constants and birth-cohort fixed effects, the interactions of settlement characteristics and birth-cohort fixed effects, and the individual characteristics and birth-cohort fixed effects, etc). The pre-displacement patterns show a drop in primary school attainment for refugee communities in Anatolia that faced Ottoman repression were impacted by WWI and the subsequent Greco-Turkish War. The correlation between human capital and the share of refugees turned positive for refugees arriving in Greece when they were about to start school (cohorts born in 1917–26). The correlation strengthens for Greeks born after that the resettlement was completed (cohorts born post-1927). Panels *C*–*F* illustrate the underlying dynamics. The likelihood of completing primary schooling is positively associated with the share of refugees for Greeks born in the 1920s and 1930s. As the share of Greeks completing primary education approached 100%, the completion rate of junior high school and then high school are higher in settlements with more substantial refugee presence. The correlation between the share of refugees and the completion of university education becomes significant and positive for those born after WWII.

Summary The analysis across all rural settlements reveals a strong association between human capital and the share of refugees in 1928 in the settlement of residence or birthplace. While these estimates may not have a causal interpretation due to mixing, and perhaps other factors related to initial settlement features (selection), they are in line, both qualitatively and quantitatively, with the baseline estimates (in Section 4) based on comparisons of nearby (in the same province and even sub-province level) almost exclusively refugee and native settlements, where the role of geography, location, and local features are muted.

Table B3: Displacement and Years of Schooling in 2011, cond. on Residence and Industry

	(1)	(2)	(3)	(4)	(5)	(6)
Born in a refugee settlement	0.604***	0.531***	0.369***	0.351***	0.352***	0.324***
	(0.162)	(0.167)	(0.105)	(0.114)	(0.097)	(0.105)
R-squared	0.108	0.162	0.363	0.379	0.422	0.434
Observations	154,461	154,453	153,617	153,609	154,395	154,387
Average outcome in native settlements	10.8	10.8	10.8	10.8	10.8	10.8
Cohort × Individual controls	✓	✓	✓	✓	✓	✓
Cohort × Birth settlement controls	✓	✓	✓	✓	✓	✓
Cohort × Birth province FE	✓	✓	✓	✓	✓	✓
Cohort × Industry (1-digit) FE			✓	✓		
Cohort × Industry (3-digit) FE					✓	✓
Residence municipality FE		✓		✓		✓
# Refugee observations	12,148	12,145	12,080	12,077	12,137	12,134
# Refugee birth settlements	353	353	353	353	353	353
# Native observations	142,313	142,308	141,537	141,532	142,258	142,253
# Native birth settlements	5,556	5,556	5,555	5,555	5,556	5,556
# Birth provinces for FE	139	139	139	139	139	139
# Residence-municipality FE		1,008		1,008		1,008

Notes. Both panels report OLS estimates associating years of schooling and a binary variable that takes the value of one for those born in a refugee settlement (where the share of refugees exceeded 80% in 1928) and zero for those born in a native settlement (where the share of refugees in 1928 was below 5%). The sample is restricted to Greek citizens, aged 25–55, who are employed. Columns (1)–(2) consider replicates the baseline estimation presented in Table 2–Columns (1) and (2), Columns (3)–(4) control for the interactions of 1-digit industry codes (N=15) with birth-cohort fixed effects. Columns (5)–(6) control for the interactions of 3-digit industry codes (N=260) with birth-cohort fixed effects. Specifications include interactions of individual characteristics (age and gender) and birth-cohort fixed effects and interactions of birth-settlement characteristics and birth-cohort fixed effects. Specifications in Panel *A* include interactions of the birth-province fixed effects and birth-cohort fixed effects. All specifications in Panel *B* include interactions of the birth-subprovince (Voronoi polygon) fixed effects and birth-cohort fixed effects. Columns (2), (4), and (6) control for municipality-of-residence fixed effects. Birth settlement characteristics are: log distances to the coast, to the provincial capital, to railway stops, and to reclaimed marshlands; geographic characteristics: crop suitability, altitude, precipitation, and temperature; and the log population in 1920, an indicator for settlements with the monastery in 1920. Standard errors are (double) clustered at the birth-settlement and the municipality-of-residence levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1%, confidence level.

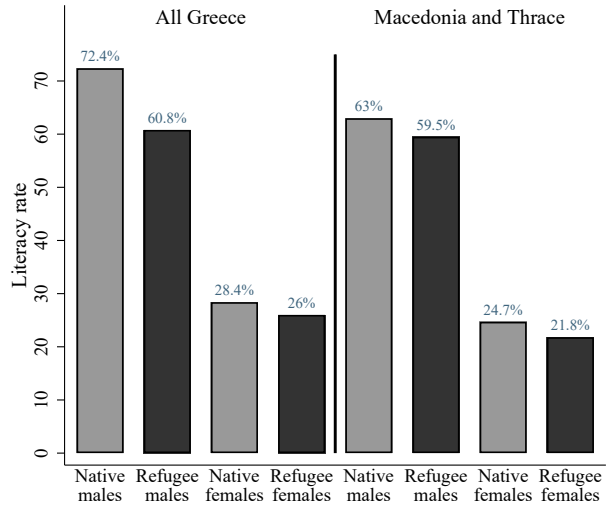
Table B4: Educational Attainment Spillovers to Native Settlements (1971–2011)

Dependent Variable:	Years of schooling			
	All sample	Local sample	All sample	Local sample
	(1)	(2)	(3)	(4)
Refugee share in 1928 within 5 kms × Post	0.002	-0.000	0.001	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Refugee share in 1928 within 5 kms × War cohort	0.000	0.001	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Log total population within 5 kms × Post	0.017*	0.017	0.017	0.030
	(0.010)	(0.020)	(0.010)	(0.021)
Log total population within 5 kms × War cohort	0.019*	-0.005	0.018*	-0.001
	(0.011)	(0.019)	(0.011)	(0.020)
Refugee share in 1928 within 5 to 25 kms × Post			0.002	-0.001
			(0.006)	(0.005)
Refugee share in 1928 within 5 to 25 kms × War cohort			-0.002	-0.003
			(0.005)	(0.005)
Log total population within 5 to 25 kms × Post			0.062	0.402***
			(0.040)	(0.146)
Log total population within 5 to 25 kms × War cohort			0.022	0.112
			(0.037)	(0.123)
R-squared	0.458	0.485	0.458	0.485
Observations	1,348,231	266,541	1,348,231	266,541
Average outcome in native settlements: 1887-1906	2.0	1.5	2.0	1.5
Average outcome in native settlements: 1907-1916	2.7	2.1	2.7	2.1
Average outcome in native settlements: 1917-1986	6.4	6.1	6.4	6.1
Settlement FE	✓	✓	✓	✓
Cohort × Census year FE	✓	✓	✓	✓
Cohort × Individual controls	✓	✓	✓	✓
Cohort × Settlement controls	✓	✓	✓	✓
Cohort × Province FE	✓		✓	
Cohort × Sub-province FE		✓		✓
# Native settlements	5,922	1,001	5,922	1,001
# Admin units for FE	137	120	137	120

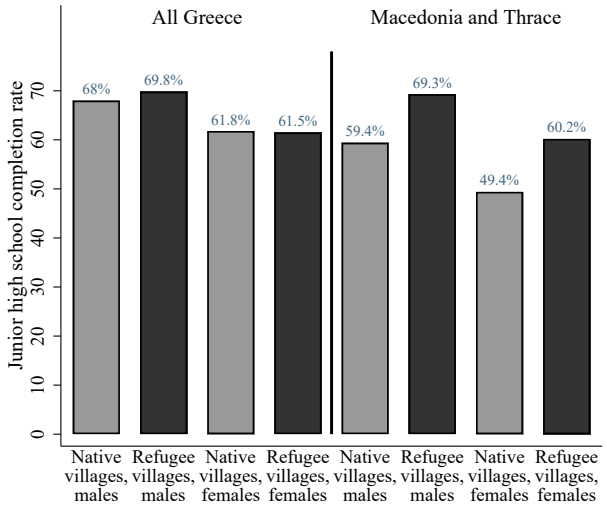
Notes. This table presents the OLS estimates associating the share of refugees in 1928 in the nearby residence settlement to years of schooling across all rural native settlements. *Post* is an indicator for cohorts born after 1917 and that started school in Greece after the Population Exchange. *War cohort* is an indicator for cohorts born between 1907 and 1916 who were of school age during the Balkan Wars, the Great War, and the Greco-Turkish War. All columns control for settlement fixed effects, the interactions of province (*eparchia*) fixed effects and birth-cohort fixed, the interactions of census-year and birth-cohort fixed effects, the interactions of individual characteristics (age and gender) and birth-cohort fixed effects, and the interactions of settlement characteristics and birth-cohort fixed effects. Settlement characteristics are: log distance to the coast, the provincial capital, railway stops, and reclaimed marshlands; geographic characteristics: crop suitability, altitude, precipitation, and temperature; and the log population in 1920, an indicator for settlements with the monastery in 1920. The 1897–1906 birth-cohort serves as the excluded cohort. Standard errors are clustered at the settlement-of-residence level. *, **, and *** denote statistical significance at the 10%, 5%, and 1%, confidence level.

Figure B7: Education in 1928 and 2011 across Refugees and Natives by Gender

(a) Literacy in 1928



(b) Junior high school completion in 2011



Notes. Panel *A* plots literacy rates for Greeks, aged 15–64, in 1928 living in the countryside (rural) by gender: refugees (gray bars) and natives (black bars). Literacy rates are retrieved from the 1928 General Population Census. Panel *B* plots the share of Greeks, aged 15–64, who had completed junior high school (9 years of schooling) in 2011 using data from the 2011 Population Census: gray bars across native rural settlements and black bars across predominantly refugee settlements (villages and towns). Both panels report means across Greece (left), and Macedonia and West Thrace (right).

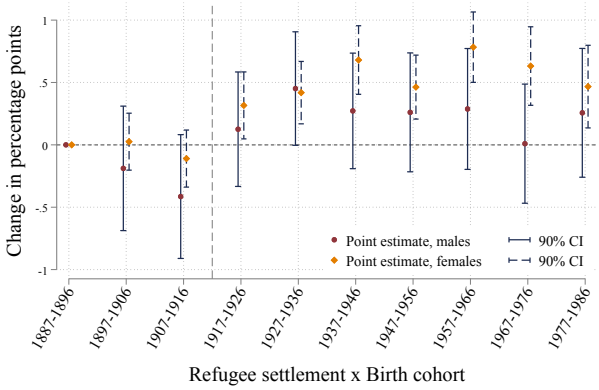
Table B5: Forced Displacement and Gender-Specific Human Capital (1971–2011)

Dependent Variable:	Years of schooling		Completed education level: Primary school		Jun. high school		High school		Tertiary education	
Sample:	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Refugee settlement × Post	0.421***	0.521***	4.210*	3.600***	2.601**	4.880***	2.314***	3.412***	0.515	1.413***
	(0.157)	(0.103)	(2.196)	(1.355)	(1.017)	(0.695)	(0.814)	(0.671)	(0.449)	(0.352)
Refugee settlement × War cohort	-0.265	-0.130	-3.911*	-2.317	-0.281	0.599	-0.302	0.013	-0.311	-0.237
	(0.163)	(0.100)	(2.302)	(1.496)	(0.968)	(0.483)	(0.667)	(0.459)	(0.450)	(0.187)
R-squared	0.359	0.528	0.236	0.411	0.339	0.464	0.239	0.372	0.065	0.121
Observations	737,006	739,464	737,006	739,464	737,006	739,464	737,006	739,464	737,006	739,464
Average outcome in native settlements: 1887-1906	3.3	0.9	49.6	14.0	7.1	0.9	2.0	0.5	0.7	0.2
Average outcome in native settlements: 1907-1916	3.9	1.5	58.0	24.4	9.7	1.6	3.5	0.7	0.9	0.2
Average outcome in native settlements: 1917-1986	7.0	5.7	84.8	70.3	34.0	24.0	22.3	17.9	6.5	5.1
Settlement FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Census year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Individual controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Settlement controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Province FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Refugee observations	62968	65334	62968	65334	62968	65334	62968	65334	62968	65334
# Refugee settlements	359	358	359	358	359	358	359	358	359	358
# Native observations	674,038	674,130	674,038	674,130	674,038	674,130	674,038	674,130	674,038	674,130
# Native settlements	5,897	5,866	5,897	5,866	5,897	5,866	5,897	5,866	5,897	5,866
# Admin units for FE	142	140	142	140	142	140	142	140	142	140

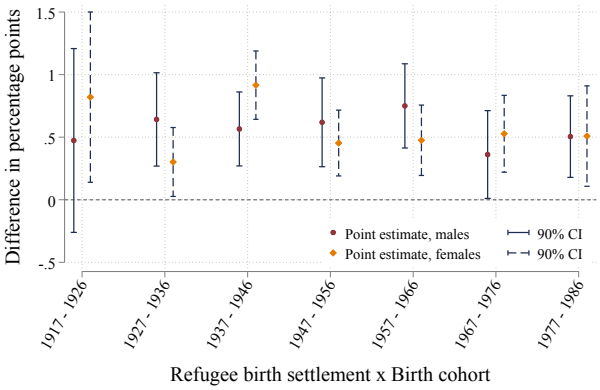
Notes. All specifications report OLS estimates associating various educational outcomes to a binary refugee settlement indicator, *Refugee settlement*, that identifies those residing in predominantly refugee villages where the share of refugees in 1928 exceeded 80%. The binary indicator equals zero for Greeks residing in native settlements, where the refugees' share in 1928 was below 5%. *Post* is an indicator for cohorts born after 1917 and that started school in Greece after the Population Exchange. *War cohort* is an indicator for cohorts born between 1907 and 1916 who were of school age during the Balkan Wars, the Great War, and the Greco-Turkish War. Odd-numbered columns restrict the sample to males, while even-numbered columns restrict it to females. The dependent variable in columns (1)–(2) is years of schooling; in columns (3)–(4) the outcome is an indicator for completing primary school; in columns (5)–(6) an indicator for completing junior high school; in columns (7)–(8) an indicator for completing high school; and in columns (9)–(10) an indicator for completing tertiary education. Odd-numbered columns focus on the sample of all rural refugee and native settlements, conditioning on birth-cohort specific province (*eparchia*) fixed effects. Even-numbered columns focus on the local sample, conditioning on birth-cohort specific sub-province (Voronoi polygon) fixed effects. In each Voronoi polygon, refugee villages and native settlements are less than 25 km away. All specifications include settlement fixed effects, interactions of census-year and birth-cohort fixed effects, interactions of individual characteristics (age and gender) and birth-cohort fixed effects, and interactions of various settlement characteristics with birth-cohort fixed effects. Settlement characteristics are: log distance to the coast, the provincial capital, railway stops, and reclaimed marshlands; geographic characteristics: crop suitability, altitude, precipitation, and temperature; and the log population in 1920, an indicator for settlements with the monastery in 1920. The 1887–1906 birth cohort serves as the excluded generation. Standard errors are clustered at the settlement-of-residence level. *, **, and *** denote statistical significance at the 10%, 5%, and 1%, confidence level.

Figure B8: Forced Displacement and Schooling. Gender-Specific Estimates

(a) Diff-in-Diff (1971-2011)



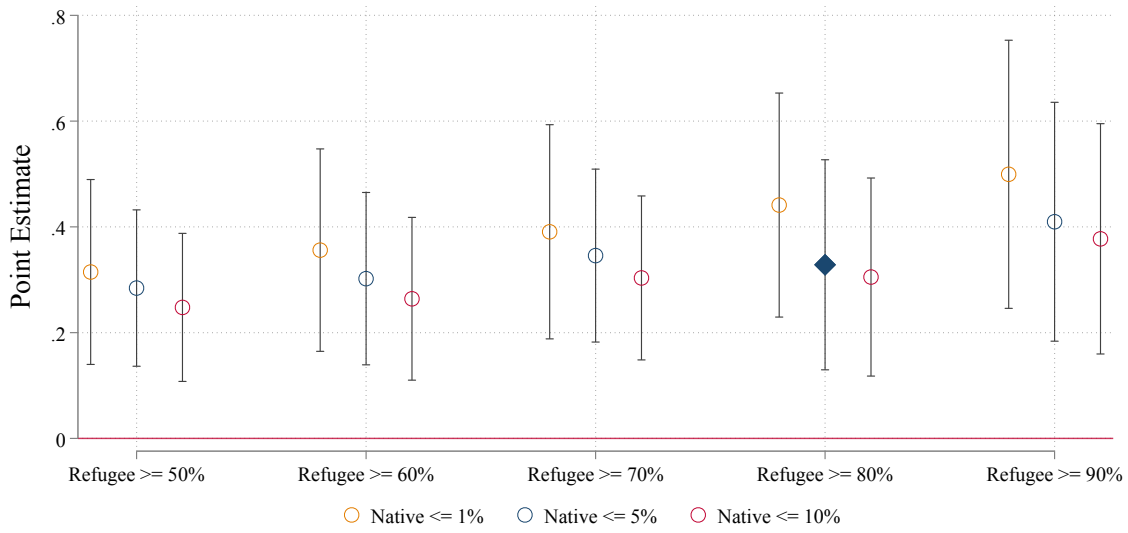
(b) Cond. on Residence (2011)



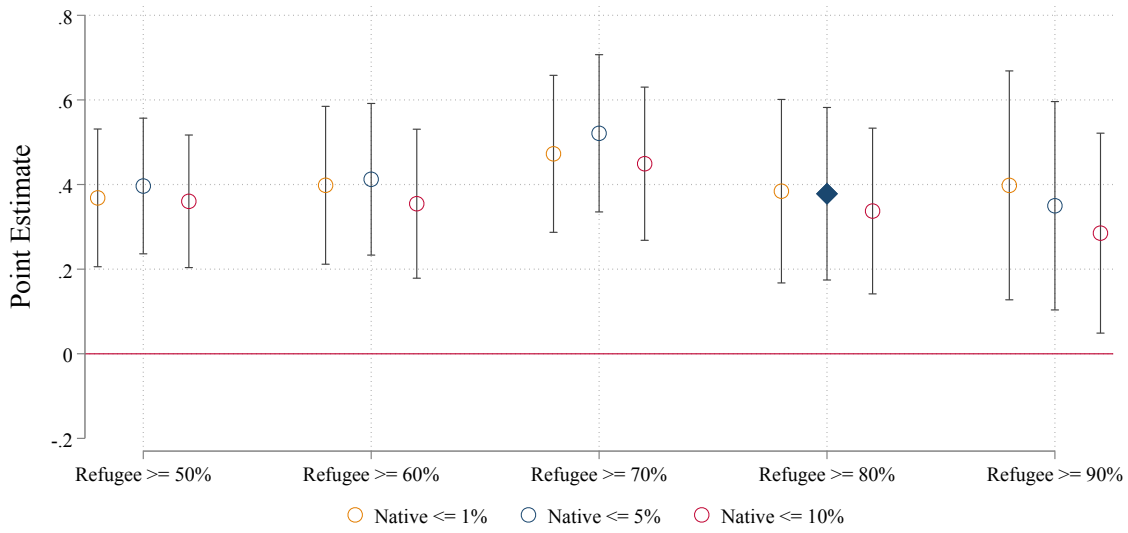
Notes. The panels explore the difference-in-difference estimates (Panel *A*) and the estimates conditioning on residence (Panel *B*) by gender. Refugee settlements are those where the share of refugees in 1928 exceeds 80%, and native settlements are those where the share of refugees from Anatolia in 1928 is below 5%. Maroon markers and solid lines present the estimated coefficients and 90% confidence intervals obtained from restricting the analysis to males, while orange markers and dashed lines present those from restricting the analysis to females. The confidence intervals in Panel *A* are based on standard errors clustered at the municipality-of-residence level, while in Panel *B* are based on standard errors double clustered at the birth-settlement and the municipality-of-residence levels. The difference-in-difference specifications use micro samples from the 1971, 1981, 1991, 2001, and 2011 General Population Censuses. The conditional-on-residence specifications use the micro-sample from the 2011 General Population Census.

Figure B9: Sensitivity Analysis I. Classification Predominantly Refugee and Native Rural Settlements

(a) Diff-in-Diff (1971–2011)



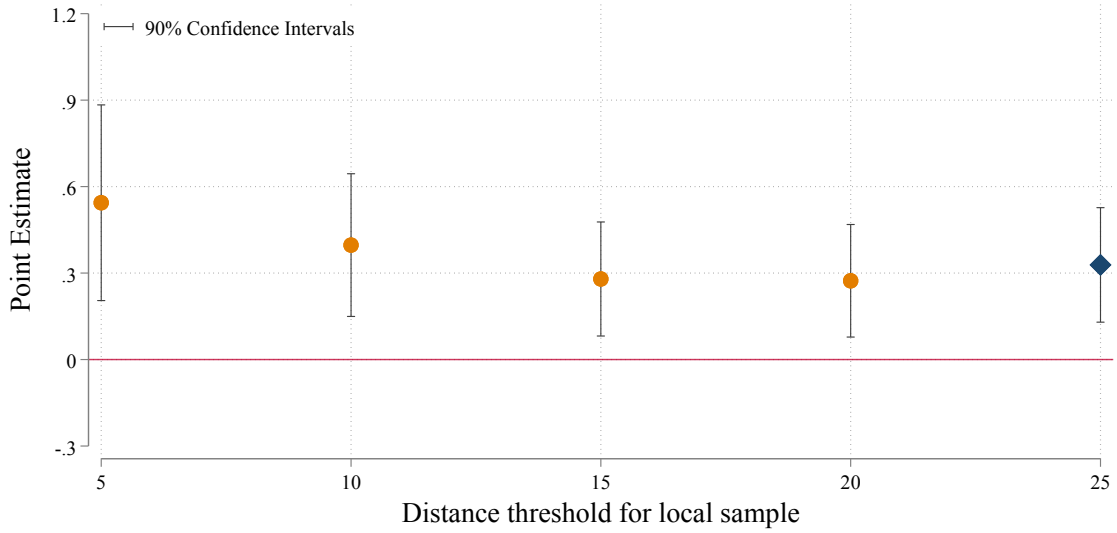
(b) Cond. on Residence (2011)



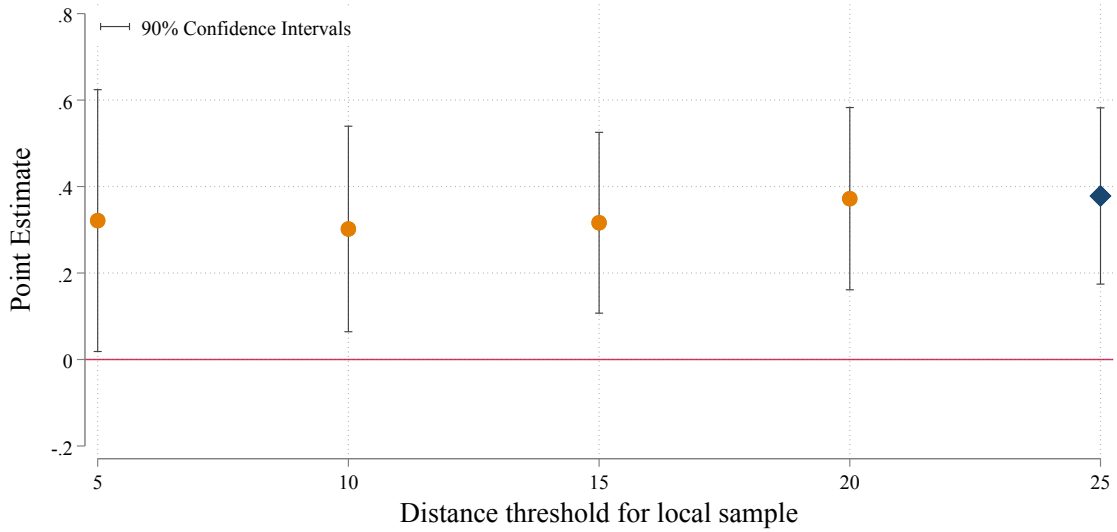
Notes. The panels explore the sensitivity of the difference-in-difference estimates (Panel *A*) and the estimates conditioning on residence (Panel *B*) to changing the definition of (predominantly) refugee and native settlements. Each panel gives 15 coefficients and 90% confidence intervals: refugee settlements are those where the share of refugees in 1928 exceeds 50%, 60%, 70%, 80%, and 90% and native settlements are those where the share of refugees in 1928 is below 1%, 5%, and 10%. The diamond shape in solid blue gives the estimates obtained using the baseline definition of refugee settlements (the share of refugees in 1928 exceeds 80%) and native settlements (the share of refugees in 1928 is below 5%). The confidence intervals in Panel *A* are based on standard errors clustered at the municipality-of-residence level, while in Panel *B* are based on standard errors double clustered at the birth-settlement and the municipality-of-residence levels. The difference-in-difference specifications use micro samples from the 1971, 1981, 1991, 2001, and 2011 General Population Censuses. The conditional-on-residence specifications use the micro sample from the 2011 General Population Census.

Figure B10: Sensitivity Analysis II. Proximity Refugee and Native Rural Settlements

(a) Diff-in-Diff (1971–2011)



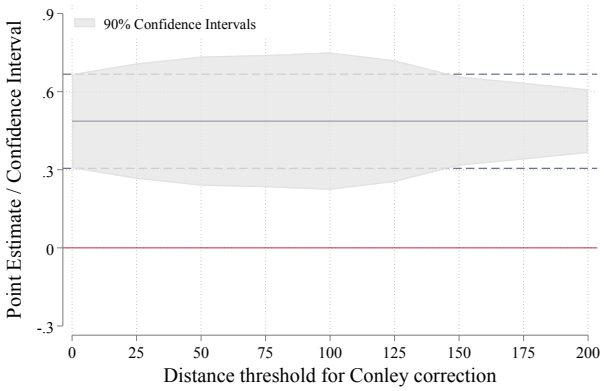
(b) Cond. on Residence (2011)



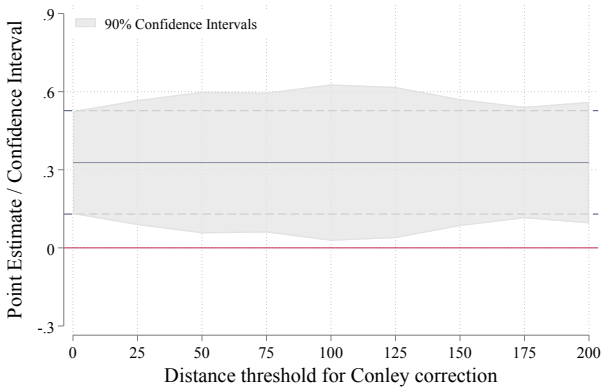
Notes. The panels explore the sensitivity of the difference-in-difference estimates (Panel *A*) and the estimates conditioning on residence (Panel *B*) to changing the local sample: the maximum distance between predominantly refugee and native settlements in the same sub-province unit (Voronoi polygons) in the local sample. Each panel gives 5 coefficients and 90% confidence intervals from alternative distance cutoffs: 25 kilometers (baseline), 20 km, 15 km, 10 km, and 5 km. Refugee settlements are those where the share of refugees in 1928 exceeds 80%, and native settlements are those where the share of refugees in 1928 is below 5%. The confidence intervals in Panel *A* are based on standard errors clustered at the residence-settlement level, while in Panel *B* are based on standard errors double clustered at the birth-settlement and the municipality-of-residence levels. The difference-in-difference specifications use micro samples from the 1971, 1981, 1991, 2001, and 2011 General Population Censuses. The conditional-on-residence specifications use the micro-sample from the 2011 General Population Census.

Figure B11: Sensitivity Analysis III. Conley Standard Errors with Varying Distance Thresholds

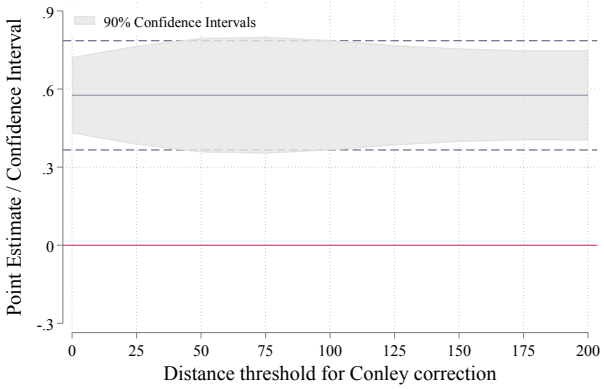
(a) Diff-in-Diff, All Rural Sample (1971–2011)



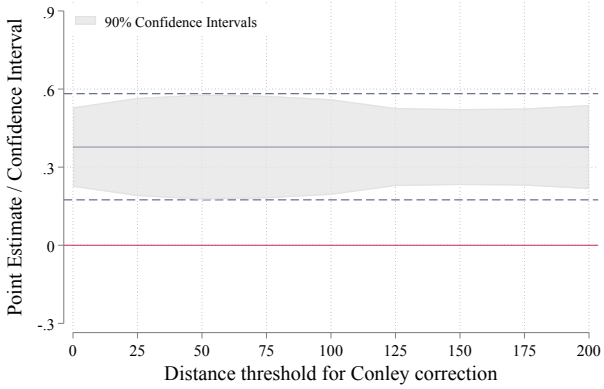
(b) Diff-in-Diff, Local Sample (1971–2011)



(c) Cond. on Residence, All Rural Sample (2011)



(d) Cond. on Residence, Local Sample (2011)



Notes. All panels explore the sensitivity of the difference-in-difference estimates (Panels A and B) and the estimates conditioning on residence (Panels C and D) to correcting the standard errors for spatial autocorrelation for several distance thresholds that vary between 0 and 200 kilometers, increasing incrementally by 25 kilometers. In each reported estimate, the settlements within the indicated distance threshold are assumed to be correlated to each other, while settlements that are farther away are not. The gray horizontal line represents the point estimates. The dashed navy horizontal lines represent the baseline 90% confidence intervals, while the shaded are represents confidence intervals correcting for spatial autocorrelation. Panels A and C consider all rural predominantly refugee and native (birth) settlements in the same province, while Panels B and D restrict the sample to the rural predominantly refugee and native (birth) settlements in the same sub-province unit (Voronoi polygons) within a 25 km radius. The difference-in-difference specifications use micro samples from the 1971, 1981, 1991, 2001, and 2011 General Population Censuses. The conditional-on-residence specifications use the micro-sample from the 2011 General Population Census.

Table B6: Educational Attainment across All Rural Settlements (1971–2011)

Dependent Variable:	Years of schooling	Completed education level: Primary school	Jun. high school	High school	Tertiary education
	(1)	(2)	(3)	(4)	(5)
Share of refugees in 1928 × Post	0.004***	0.042***	0.032***	0.024***	0.006*
	(0.001)	(0.010)	(0.007)	(0.006)	(0.003)
Share of refugees in 1928 × War cohort	-0.002**	-0.035***	0.004	0.002	-0.001
	(0.001)	(0.011)	(0.004)	(0.003)	(0.002)
R-squared	0.464	0.356	0.406	0.305	0.092
Observations	2,123,467	2,123,467	2,123,467	2,123,467	2,123,467
Average outcome in native settlements: 1887-1906	2.0	31.0	3.9	1.2	0.4
Average outcome in native settlements: 1907-1916	2.7	40.7	5.6	2.0	0.5
Average outcome in native settlements: 1917-1986	6.4	77.6	29.0	20.1	5.8
Settlement FE	✓	✓	✓	✓	✓
Cohort × Census Year FE	✓	✓	✓	✓	✓
Cohort × Individual controls	✓	✓	✓	✓	✓
Cohort × Settlement controls	✓	✓	✓	✓	✓
Cohort × Province FE	✓	✓	✓	✓	✓
# Settlements	7,507	7,507	7,507	7,507	7,507
# Admin units for FE	142	142	142	142	142

Notes. This table presents the OLS estimates associating the share of refugees in 1928 in the residence settlement to various educational outcomes across all rural settlements. *Post* is an indicator for cohorts born after 1917 and that started school in Greece after the Population Exchange. *War cohort* is an indicator for cohorts born between 1907 and 1916 who were of school age during the Balkan Wars, the Great War, and the Greco-Turkish War. The dependent variable is: years of schooling in column (1); an indicator for completing primary school in column (2); an indicator for completing junior high school in column (3); an indicator for completing high school in column (4); an indicator for completing tertiary education in column (5). All columns control for settlement fixed effects, the interactions of province (*eparchia*) fixed effects and birth-cohort fixed, the interactions of census-year and birth-cohort fixed effects, the interactions of individual characteristics (age and gender) and birth-cohort fixed effects, and the interactions of settlement characteristics and birth-cohort fixed effects. Settlement characteristics are: log distance to the coast, the provincial capital, railway stops, and reclaimed marshlands; geographic characteristics: crop suitability, altitude, precipitation, and temperature; and the log population in 1920, an indicator for settlements with the monastery in 1920. The 1897–1906 birth-cohort serves as the excluded cohort. Standard errors are clustered at the settlement-of-residence level. *, **, and *** denote statistical significance at the 10%, 5%, and 1%, confidence level.

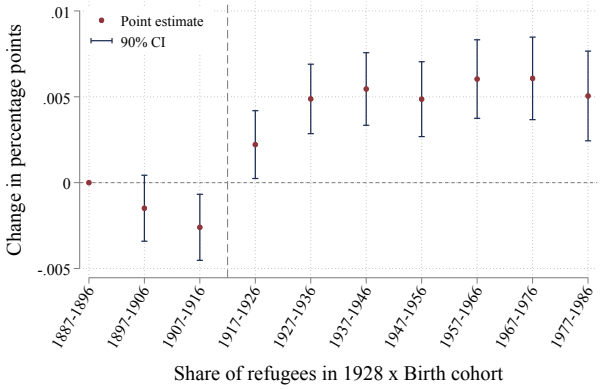
Table B7: Educational Attainment across Rural Birth Settlements, conditional on Residence (2011)

Dependent Variable:	Years of schooling		Completed education level: Primary school		Jun. high school		High school		Tertiary education	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share of refugees in 1928, birth settlement	0.005***	0.003***	0.036***	0.026***	0.047***	0.028***	0.031***	0.015*	0.008	0.003
	(0.001)	(0.001)	(0.006)	(0.005)	(0.011)	(0.009)	(0.010)	(0.008)	(0.005)	(0.005)
R-squared	0.358	0.403	0.288	0.302	0.331	0.366	0.233	0.272	0.062	0.090
Observations	761,507	761,501	761,507	761,501	761,507	761,501	761,507	761,501	761,507	761,501
Average outcome in native settlements	8.0	8.0	86.2	86.2	45.6	45.6	33.4	33.4	11.4	11.4
Cohort × Individual controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Birth settlement controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort × Birth province FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Residence municipality FE		✓		✓		✓		✓		✓
# Refugee observations	44,115	44,114	44,115	44,114	44,115	44,114	44,115	44,114	44,115	44,114
# Refugee birth settlements	357	357	357	357	357	357	357	357	357	357
# Native observations	505,365	505,361	505,365	505,361	505,365	505,361	505,365	505,361	505,365	505,361
# Native birth settlements	5,788	5,788	5,788	5,788	5,788	5,788	5,788	5,788	5,788	5,788
# Birth provinces for FE	141	141	141	141	141	141	141	141	141	141
# Residence-municipality FE		1,022		1,022		1,022		1,022		1,022

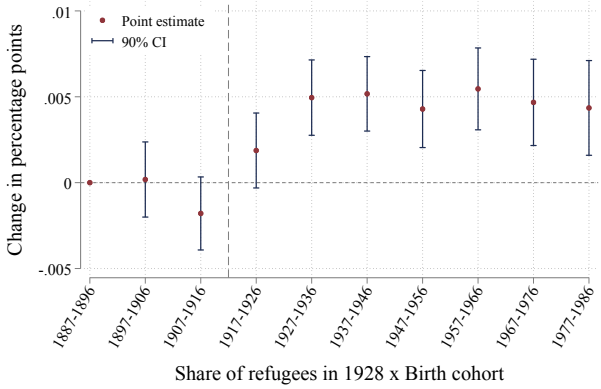
Notes. This table presents the OLS estimates associating the 1928 share refugees in the birth settlement to educational attainment across rural settlements. The dependent variable is years of schooling in columns (1)–(2), primary school completion in columns (3)–(4), junior high school completion in columns (5)–(6), high school completion in columns (7)–(8), tertiary education completion in columns (9)–(10). All reported regressions control for the interactions of the birth-province fixed effects and birth-cohort fixed effects, the interactions of birth-settlement characteristics and birth-cohort fixed effects and the interactions of individual characteristics (age and gender) and birth-cohort fixed effects. Columns (2), (4), (6), (8), and (10) control for municipality-of-residence fixed effects. Birth settlement characteristics are: log distance to the coast, the provincial capital, railway stops, and reclaimed marshlands; geographic characteristics: crop suitability, altitude, precipitation, and temperature; and the log population in 1920, an indicator for settlements with the monastery in 1920. Standard errors are clustered at the birth-settlement and the municipality-of-residence levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1%, confidence level.

Figure B12: Human Capital and Refugee Share across All Rural Settlements (1971–2011)

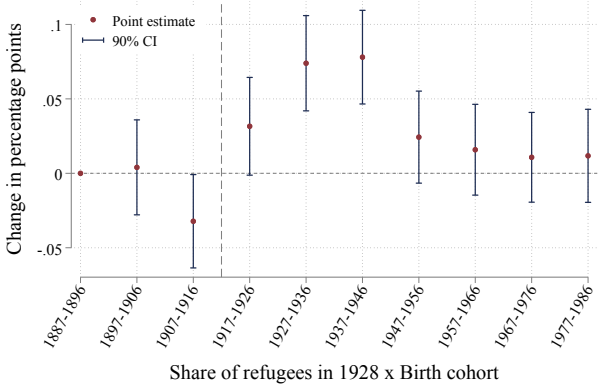
(a) Years of schooling, Prefecture FEs



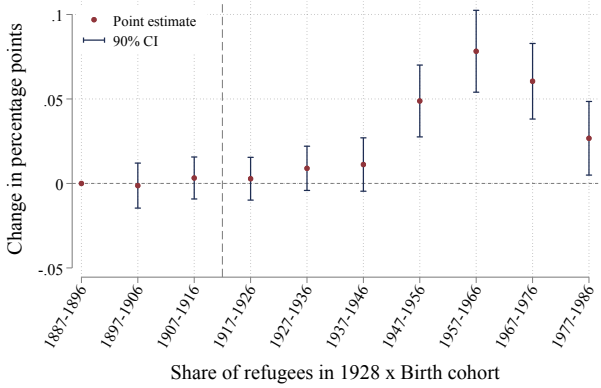
(b) Years of schooling, Full specification



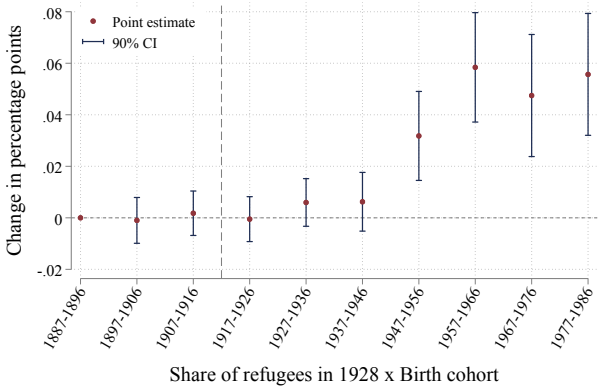
(c) Primary school completion



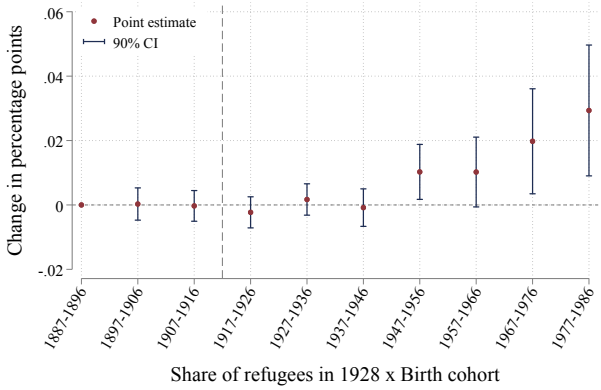
(d) Junior high school completion



(e) High school completion



(f) Tertiary education



Notes. The panels present the event-study estimates obtained from regressing educational outcomes on the share of refugees in 1928 across all rural settlements. Outcome variable in consideration is years of schooling in Panels *A* and *B*; completing primary school in Panel *C*, junior high school in Panel *D*, high school in Panel *E*, and tertiary education in Panel *F*. Panel *A* controls for settlement fixed effects, the interactions between prefecture constants and birth-cohort fixed effects, and the interactions between census-year and birth-cohort fixed effects. Panels *B*-*F* controls for the interactions of province constants and birth-cohort fixed effects, the interactions of settlement characteristics and birth-cohort fixed effects, and the individual characteristics and birth-cohort fixed effects in addition. Settlement characteristics are: log distance to the coast, the provincial capital, railway stops, and reclaimed marshlands; geographic characteristics: crop suitability, altitude, precipitation, and temperature; and the log population in 1920, an indicator for settlements with the monastery in 1920. The 1887–1896 birth-cohort serves as the excluded cohort. Standard errors are clustered at the settlement-of-residence level.

C Human Capital Portability

This Appendix Section reports descriptive statistics and sensitivity checks on the link between forced displacement and subsequent human capital investments in portable skills among college graduates, complementing the analysis in Section 5 of the main paper.

Degree Portability Appendix Table C1 lists the 50 most popular degrees in the Greek tertiary education system, tabulating the 2011 Population Census (sample 25%) alongside our classification of the skills' transferability based on Chat-GPT 4. The table gives the tabulations for the country and the rural areas in our baseline sample of predominantly refugee and native settlements, respectively.

Descriptive Patterns Appendix Figure C1 presents the average skill portability of tertiary education degrees among individuals born in native and refugee settlements. The first two sets of bars show that those from refugee settlements are more likely to complete a transferable degree by 3.5 percentage points in the baseline sample and by 5 percentage points in the local sample, where refugee and native villages are within 25 km radius in the same province. The last three sets of bars present the relative distribution of lawyers, doctors, and engineers among these professions. Lawyers are overrepresented among individuals from native settlements, while the share of doctors and engineers is significantly higher among those from predominantly refugee settlements.

Sensitivity Appendix Figure C2 reports sensitivity checks of the association between the portability of tertiary degree skills and being born in a predominantly refugee settlement, reported in Table 3. Panel *A* reports 15 specification coefficients altering the cutoff of refugees' share in 1928 to classify predominantly refugee and native settlements. Across all permutations, Greeks born in majority or predominantly refugee settlements are considerably more likely to complete tertiary education with portable skills. Panel *B* gives five permutations of the baseline specification linking an indicator of skill portability of university education to the refugee indicator that takes the value of one for an individual whose birthplace is a predominantly refugee settlement (refugee share in 1928 exceeds 80%) and zero for Greeks born in native settlements, where the share of refugees in 1928 is below 5% in the local sample. The panel gives estimates restricting comparisons between native and refugee settlements in the same province (*eparchia*) no more than 25 km, 20 km, 15 km, 10 km, and 5 km away. The estimates are stable even though the sample drops.

Spatial Autocorrelation Appendix Figure C3 explores inference accounting for spatial correlation, using Conley standard errors with a uniform kernel (Conley (1999) and Bester, Conley and Hansen (2011)). Following the suggestion of Colella et al. (2019), we present Conley standard errors for several distance thresholds: ranging from 25 km to 200 km, increasing by 25 km in each iteration, using the Stata command `acreg` (Colella et al., 2023). Panels *A* and *B* plot the 90% confidence intervals adjusted for spatial autocorrelation, conditioning on the municipality of residence across all rural

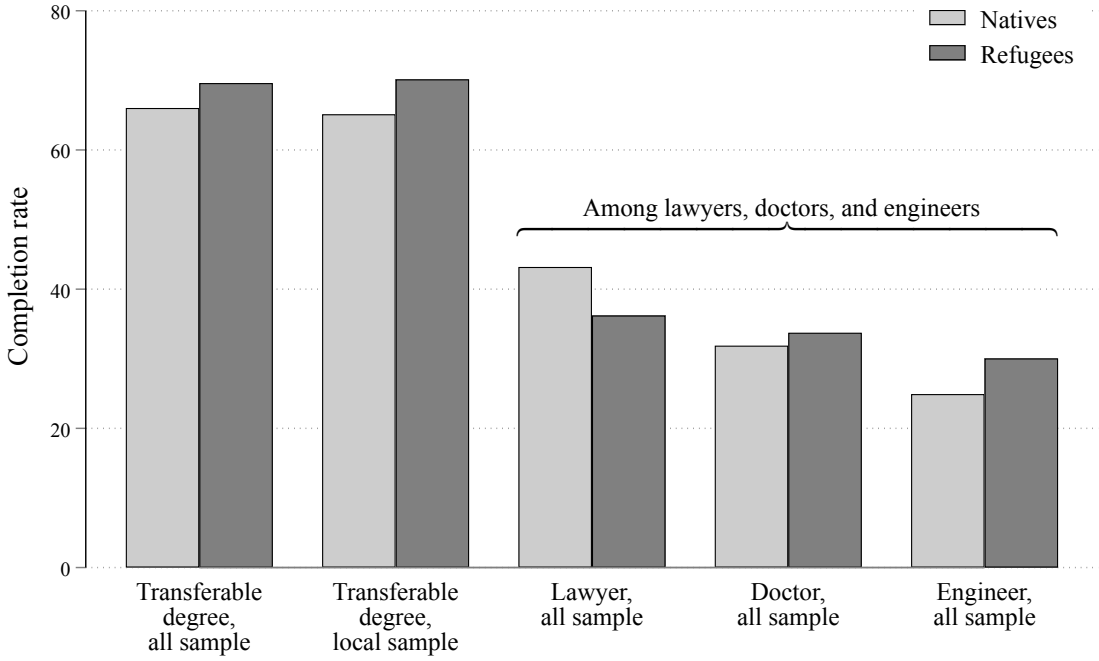
predominantly refugee and native birth settlements and those in the local sample where refugee and native birth settlements are within 25 km in the same province, respectively. The estimates' statistical significance is not sensitive to the particular cutoff used in the spatial autocorrelation adjustment.

Table C1: Top 50 Most Popular College Degrees and Skill Portability (2011)

		All Greece			Main rural sample		
Study field	Transferable Degree	Number	Share	Cumulative share	Number	Share	Cumulative share
Economic Sciences	1	23,469	6.9	6.9	4,070	6.6	6.6
Law	0	19,209	5.7	12.6	3,069	4.9	11.5
Primary Education Teachers	0	18,400	5.4	18.1	6,167	9.9	21.5
Business Administration	1	16,736	5.0	23.0	2,815	4.5	26.0
Medicine	1	15,450	4.6	27.6	2,304	3.7	29.7
Accounting	1	10,632	3.1	30.7	2,081	3.4	33.1
Civil Engineering	1	10,530	3.1	33.9	1,663	2.7	35.7
Mathematics	1	8,360	2.5	36.3	1,754	2.8	38.6
Greek Literature	0	8,303	2.5	38.8	1,713	2.8	41.3
Nursing	1	8,126	2.4	41.2	2,013	3.2	44.6
Physical Education and Sport Science	1	7,906	2.3	43.5	1,473	2.4	47.0
Military Officers	0	6,007	1.8	45.3	1,582	2.6	49.5
Physics	1	5,915	1.8	47.1	1,077	1.7	51.2
English Language and Literature	1	5,647	1.7	52.1	656	1.1	52.3
Electrical and Computer Engineering	1	5,754	1.7	48.8	746	1.2	53.5
Kindergarten Teachers	0	5,662	1.7	50.4	1,353	2.2	55.7
Mechanical Engineering	1	5,253	1.6	55.3	689	1.1	56.8
Engineering	1	5,505	1.6	53.7	1,107	1.8	58.6
Information Technology	1	4,941	1.5	56.8	572	0.9	59.5
Architectural Engineers	1	4,770	1.4	58.2	406	0.7	60.2
Political Sciences	1	4,657	1.4	59.5	1,091	1.8	61.9
Philosophy	1	4,465	1.3	62.2	923	1.5	63.4
Electrical Engineering	1	4,522	1.3	60.9	954	1.5	64.9
Agronomist	0	4,194	1.2	63.4	1,025	1.7	66.6
Dentistry	1	3,719	1.1	66.8	561	0.9	67.5
Theology	0	3,884	1.1	64.6	1,024	1.7	69.2
Pharmacists	1	3,781	1.1	65.7	626	1.0	70.2
French Language and Literature	1	3,487	1.0	67.8	249	0.4	70.6
Tourism Enterprises	1	3,229	1.0	68.8	456	0.7	71.3
Chemical Engineers	1	2,890	0.9	72.4	387	0.6	71.9
Chemistry	1	3,107	0.9	70.6	491	0.8	72.7
Policemen	0	2,953	0.9	71.5	743	1.2	73.9
History and Archaeology	1	3,135	0.9	69.7	493	0.8	74.7
Psychology	1	2,758	0.8	73.2	253	0.4	75.1
Business Organisation and Management	1	2,211	0.7	74.5	297	0.5	75.6
Sociology	1	2,272	0.7	73.9	393	0.6	76.2
Biology	1	1,966	0.6	76.4	253	0.4	76.6
Public Administration	0	2,163	0.6	75.8	525	0.8	77.5
Electronics	1	2,171	0.6	75.2	379	0.6	78.1
Physiotherapy	1	1,618	0.5	78.9	240	0.4	78.5
Infant and Child Care	0	1,696	0.5	77.4	343	0.6	79.0
Naval Officers	0	1,665	0.5	77.9	273	0.4	79.5
Plant Production	0	1,737	0.5	76.9	480	0.8	80.3
Fine Arts	1	1,644	0.5	78.4	140	0.2	80.5
Midwives	1	1,599	0.5	79.3	375	0.6	81.1
Accounting and Finance	1	1,506	0.4	79.8	286	0.5	81.5
Social Work	1	1,467	0.4	80.2	339	0.5	82.1
Electronic Computing Systems	1	1,426	0.4	80.6	255	0.4	82.5
Geology	1	1,405	0.4	81.5	197	0.3	82.8
Medical Laboratories	1	1,421	0.4	81.1	289	0.5	83.3

Notes. This table lists the 50 most popular degrees studied in tertiary education along with their skill portability status and prevalence. The degrees are sorted in descending order by their popularity.

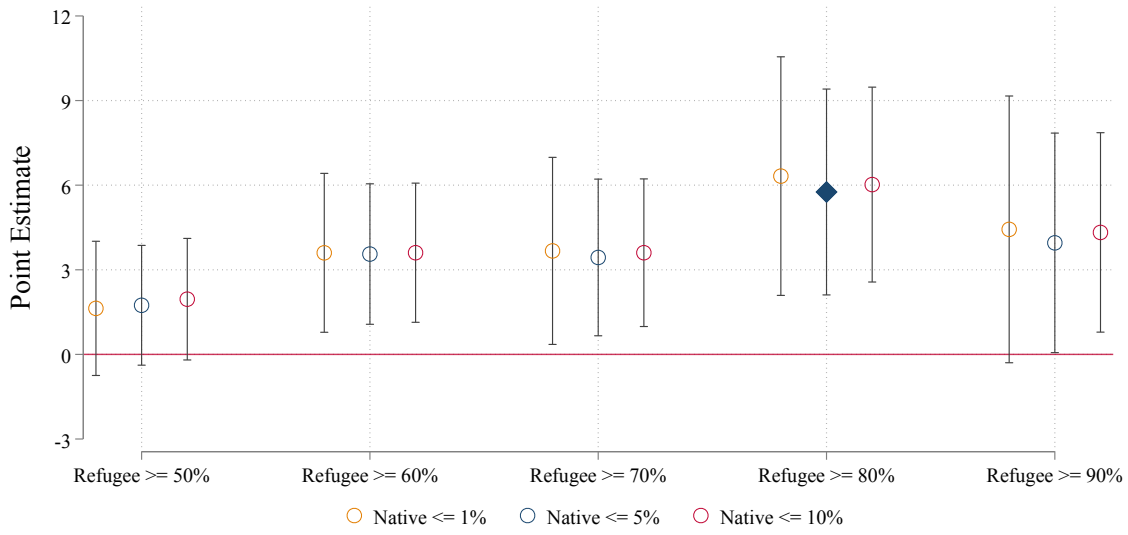
Figure C1: Skill Transferability across Refugee and Native Settlements (2011 Census)



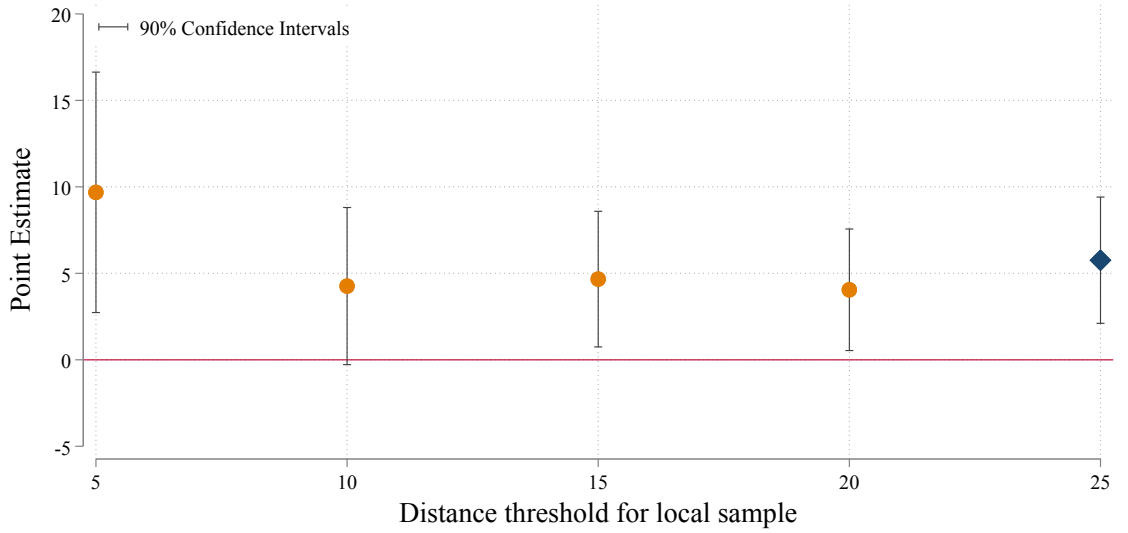
Notes. This figure presents the average skill transferability of tertiary degrees across refugee and native birth settlements among those with a tertiary education. It also shows the relative distribution of lawyers, doctors, and engineers among those with a degree in law, medicine, and engineering (excluding civil engineering) from universities and polytechnics across all rural refugee and native birth settlement. All sample stands for all rural predominantly refugee and native birth settlements; local sample focuses on a sample where native and refugee birth settlements are within a 25 km radius in the same province.

Figure C2: Sensitivity Checks. Forced Displacement and Skill Portability (2011)

(a) Classification of Native and Refugee Settlements



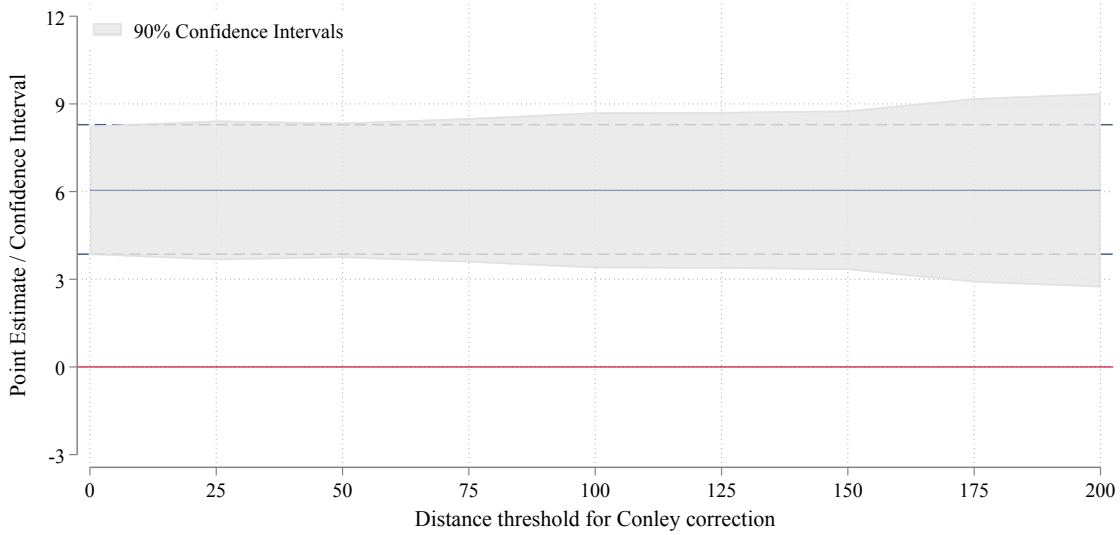
(b) Distance Threshold. Local Sample Analysis



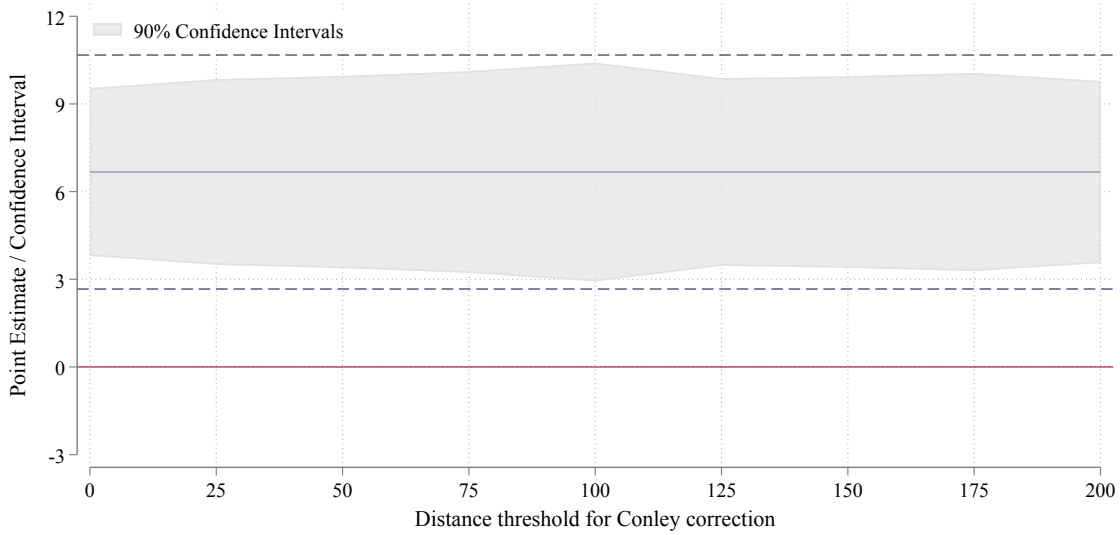
Notes. Panel *A* presents 15 estimates changing the definition of (predominantly) refugee and native settlements. Panel *B* plots 5 estimates changing the maximum distance between refugee and native birth settlements to build the sub-province clusters in the local analysis sample. In all OLS specifications, the dependent variable is an indicator for tertiary education degrees with portable skills. The diamond shape in solid blue presents the coefficient estimates, from Table 3–column (4), based on the baseline definition of refugee birth settlements (share of refugees in 1928 exceeds 80%) and native birth settlements (share of refugees in 1928 is below 5%), and the baseline definition of local sample, where refugee birth settlements are matched to native ones no more than 25 km away in the same sub-province unit. The round hollow shapes in Panel *A* represent the estimates obtained from using alternative definitions of refugee settlements (share of refugees in 1928 exceeds 50%, 60%, 70%, 80%, and 90%) and native settlements (share of refugees in 1928 is below 1% and 10%). The round orange shapes in Panel *B* represent the estimates obtained using alternative distance thresholds: 5 km, 10 km, 15 km, 20 km, and 25 km. The two panels also give 90% confidence intervals based on standard errors double clustered at the birth-settlement and the municipality-of-residence levels. The conditional-on-residence specifications use the micro-sample from the 2011 General Population Census.

Figure C3: Sensitivity Analysis III. Conley Standard Errors with Varying Distance Thresholds

(a) All sample



(b) Local analysis sample



Notes. The figure explores the sensitivity of the estimates conditioning on residence to correcting the standard errors for spatial autocorrelation for several distance thresholds that vary between 0 and 200 kilometers, increasing incrementally by 25 kilometers. In each reported estimate, the settlements within the indicated distance threshold are assumed to be correlated to each other, while settlements that are farther away are not. The gray horizontal line represents the point estimates. The dashed navy horizontal lines represent the baseline 90% confidence intervals, while the shaded are represents confidence intervals correcting for spatial autocorrelation. Panels *A* considers all rural predominantly refugee and native birth settlements in the same province (Table 3–column (2)), while Panel *B* restricts the sample to the predominantly refugee and native birth settlements in the same sub-province unit (Voronoi polygons) within a 25 km radius (Table 3–column (4)).

D Mechanisms: Additional Results

This Appendix Section complements the heterogeneity analysis that delves into additional mechanisms linking the initial forced displacement experience to subsequent human capital investments (in Section 6 of the paper). The appendix gives additional results, summary statistics, and descriptive evidence across all aspects we explore heterogeneity.

Refugees' Settlement Characteristics Appendix Table D1 gives summary statistics (mean, median, max, min, standard deviation, the 10th and the 90th percentile) of all demographic, geographic, locational, and ethnographic variables we use in the heterogeneity analysis. Panel *A* reports summary statistics across 384 or 405 predominately refugee settlements, defined as those where the share of refugees in 1928 was above 80% of the total population. Panel *B* gives the statistics weighted by settlements' population in 1928.

Displacement Trajectory by Origin Region The Greek-Orthodox people from Smyrna (Izmir), Ionia, and the coastal cities had to flee in chaotic conditions, given the swift advance of Turkish troops in September 1922. The Eastern Thrace Greeks abandoned their homes in the fall of 1923, crossing the Evros River to Western Thrace, but in a relatively organized manner, carrying movable property and animals. The Cappadocia-Cilicia communities left their homes after the Convention for the Population Exchange but without the ability to carry assets. As part of the voluntary population exchange, the Greek-Orthodox communities from Bulgaria, unaffected by the Greco-Turkish wars, moved relatively peacefully south with their belongings. Pontic Greeks, living on the southern shores of the Black Sea and the Caucasus, faced labor camp internments and large-scale massacres. Those who survived arrived in 1923 and several even later. Appendix Figure D5 plots the share of refugees by origin as recorded in the 1928 General Population Census and the Rural Refugee Catalog.

Heterogeneity by Educational Level Appendix Figures D1–D4 present the heterogeneity analysis for different levels of educational attainment: primary school completion, junior high and high school completion, and tertiary education. It complements the baseline results in the main paper using schooling years.

Origin Shares Appendix Figure D5–Panel *A* plots the share of refugees by origin as recorded in the 1928 General Population Census for the entire population of the country and for rural areas separately. Panel *B* gives the distribution based on the Rural Refugee Catalog for all rural settlements that received at least one land grant and for rural settlements in the local sample.

Origin Distribution at Settlement Level. Mappings Appendix Figure D6 maps the share of refugees from Asia Minor and Cappadocia (Panel *A*), Pontus and Thrace (Panel *B*), Eastern Thrace (Panel *C*), and Constantinople (Panel *D*) across 141 provinces based on the breakdown provided by

the 1928 General Population Census. The four maps also overlay rural settlements where majority of land grant recipients originated from each region, based on the Rural Refugee Catalog compiled by the RSC in the mid-1920s.

Balance Test. Cappadocian Settlements vs. Other Origin Settlements Appendix Figure D7 explores differences in geographic, locational, and ecological features between refugee settlements where the majority of the population is from Cappadocia and other refugee settlements (with a non-Cappadocian majority). The two types of refugee settlements are rather balanced in terms of their characteristics. Once we condition on prefecture (level-2 admin unit) and province (level-3 admin unit) fixed effects, there are no statistically significant differences.

Event-Study by Origin Region Appendix Figure D8 plots the cohort-specific estimates of the refugee indicators distinguishing the main origin region of refugees settled with a land grant (Asia Minor, Cappadocia, Pontus and Caucasus, East Thrace, and Bulgaria) to quantify the evolution of refugee-native differences in schooling . Panel A gives the difference-in-difference estimates that compare the evolution of schooling years of Greeks in refugee and native settlements after the forced displacement to the pre-population exchange, pooling together the micro samples of all post-1971 General Population Censuses. Panel B gives conditional on residence estimates that compare Greeks born in predominantly refugee and native settlements using the 2011 General Population Census. While the estimates are somewhat noisy, the analysis clearly shows that Greeks born in refugee settlements where the majority in 1928 were from Cappadocia, who did not speak Greek, do not experience an educational gain. In contrast, the estimates for second-, third-, and fourth-generation refugees from Asia Minor, Cappadocia, Pontus and Caucasus, Pontus, East Thrace, and mixed settlements are positive despite some fluctuations.

Balance Test. New Refugee Settlements vs. Former-Muslim Refugee Settlements Appendix Figures D9 explore differences in geographic, locational, and ecological features between the former-Muslim and predominantly new refugee settlements in the baseline sample. New refugee settlements were somewhat more populated in 1928, closer to the coast and marshlands and at a lower elevation. However, once conditioning on (log) altitude, differences become muted.

New and Former-Muslim Refugee Settlements. Population Over Time Appendix Figure D10 plots the evolution of the population in the two types of rural refugee settlements, former Muslim and new ones, and native villages and towns between 1907 and 2011, as recorded in the respective Censuses. Panel A gives the evolution of the median population, along with 90th- and 10th-percentiles for all rural and native refugee settlements, while Panel B focuses on the native and refugee settlements in the local analysis sample.

Table D1: Summary Statistics across Settlements

	Mean	SD	Min	10p	Median	90p	Max	Obs.
Panel A:	**Settlement characteristics**							
Majority of refugees from Constantinople (%)	0.3	5.1	0.0	0.0	0.0	0.0	100.0	384
Majority of refugees from Asia Minor, coastal (%)	19.5	39.7	0.0	0.0	0.0	100.0	100.0	384
Majority of refugees from Cappadocia (%)	7.0	25.6	0.0	0.0	0.0	0.0	100.0	384
Majority of refugees from Pontus and Caucasus (%)	38.3	48.7	0.0	0.0	0.0	100.0	100.0	384
Majority of refugees from Thrace (%)	24.5	43.1	0.0	0.0	0.0	100.0	100.0	384
Majority of refugees from Bulgaria (%)	1.8	13.4	0.0	0.0	0.0	0.0	100.0	384
No majority region of origin (%)	8.6	28.1	0.0	0.0	0.0	0.0	100.0	384
Share of refugees from province centres in Turkey (%)	5.0	10.1	0.0	0.0	1.7	11.4	95.3	384
Origin diversity at settlement	0.7	0.2	0.0	0.3	0.7	0.9	1.0	384
Origin (province) diversity at settlement	0.5	0.2	0.0	0.1	0.5	0.8	0.9	384
Former-Muslim settlement (%)	71.9	45.0	0.0	0.0	100.0	100.0	100.0	405
Population in 1928	449.1	488.5	9.0	73.0	285.0	971.0	4,041.0	405
Agricultural suitability	27.5	10.6	11.5	16.1	25.8	40.2	71.0	405
Agricultural similarity to origin	85.2	6.2	61.3	78.1	86.0	92.2	99.3	384
Agricultural-suitability gain	-6.2	15.5	-37.4	-23.0	-8.8	14.6	60.3	384
Distance to province capital	15.6	7.8	1.9	6.5	14.1	27.7	39.0	405
Population density in 1920, province level	31.4	23.3	13.3	14.1	27.0	37.4	235.5	405
Share of manufacturing in 1920, province level	12.6	7.3	3.7	5.6	12.1	21.4	35.6	405
Share of commerce in 1920, province level	4.6	2.8	1.2	1.7	4.2	8.0	17.4	405
Share of public administration in 1920, province level	1.6	1.1	0.4	0.5	1.4	2.8	7.6	405
Number of teachers per 100 school-age children, province level	1.8	0.3	0.9	1.4	1.8	2.1	3.5	405
Panel B:	**Settlement characteristics, population weighted**							
Majority of refugees from Constantinople (%)	0.2	4.1	0.0	0.0	0.0	0.0	100.0	177,974
Majority of refugees from Asia Minor, coastal (%)	22.4	41.7	0.0	0.0	0.0	100.0	100.0	177,974
Majority of refugees from Cappadocia (%)	6.1	23.9	0.0	0.0	0.0	0.0	100.0	177,974
Majority of refugees from Pontus and Caucasus (%)	26.5	44.1	0.0	0.0	0.0	100.0	100.0	177,974
Majority of refugees from Thrace (%)	26.9	44.4	0.0	0.0	0.0	100.0	100.0	177,974
Majority of refugees from Bulgaria (%)	4.6	20.9	0.0	0.0	0.0	0.0	100.0	177,974
No majority region of origin (%)	13.3	34.0	0.0	0.0	0.0	100.0	100.0	177,974
Share of refugees from province centres in Turkey (%)	4.4	7.5	0.0	0.0	2.1	9.6	95.3	177,974
Origin diversity at settlement	0.7	0.3	0.0	0.2	0.8	0.9	1.0	177,974
Origin (province) diversity at settlement	0.6	0.3	0.0	0.1	0.6	0.8	0.9	177,974
Former-Muslim settlement (%)	66.1	47.3	0.0	0.0	100.0	100.0	100.0	181,882
Population in 1928	979.1	826.6	9.0	213.0	734.0	2,088.0	4,041.0	181,882
Agricultural suitability	27.0	9.7	11.5	16.1	26.4	36.4	71.0	181,882
Agricultural similarity to origin	85.9	5.3	61.3	80.0	86.0	91.8	99.3	177,974
Agricultural-suitability gain	-7.4	13.6	-37.4	-25.0	-8.8	10.2	60.3	177,974
Distance to province capital	14.1	7.5	1.9	6.2	12.3	25.6	39.0	181,882
Population density in 1920, province level	34.2	27.6	13.3	16.5	29.0	37.4	235.5	181,882
Share of manufacturing in 1920, province level	12.9	7.7	3.7	4.6	12.1	21.4	35.6	181,882
Share of commerce in 1920, province level	4.7	3.0	1.2	1.7	4.2	9.4	17.4	181,882
Share of public administration in 1920, province level	1.6	1.2	0.4	0.5	1.4	2.8	7.6	181,882
Number of teachers per 100 school-age children, province level	1.7	0.3	0.9	1.3	1.7	2.1	3.5	181,882

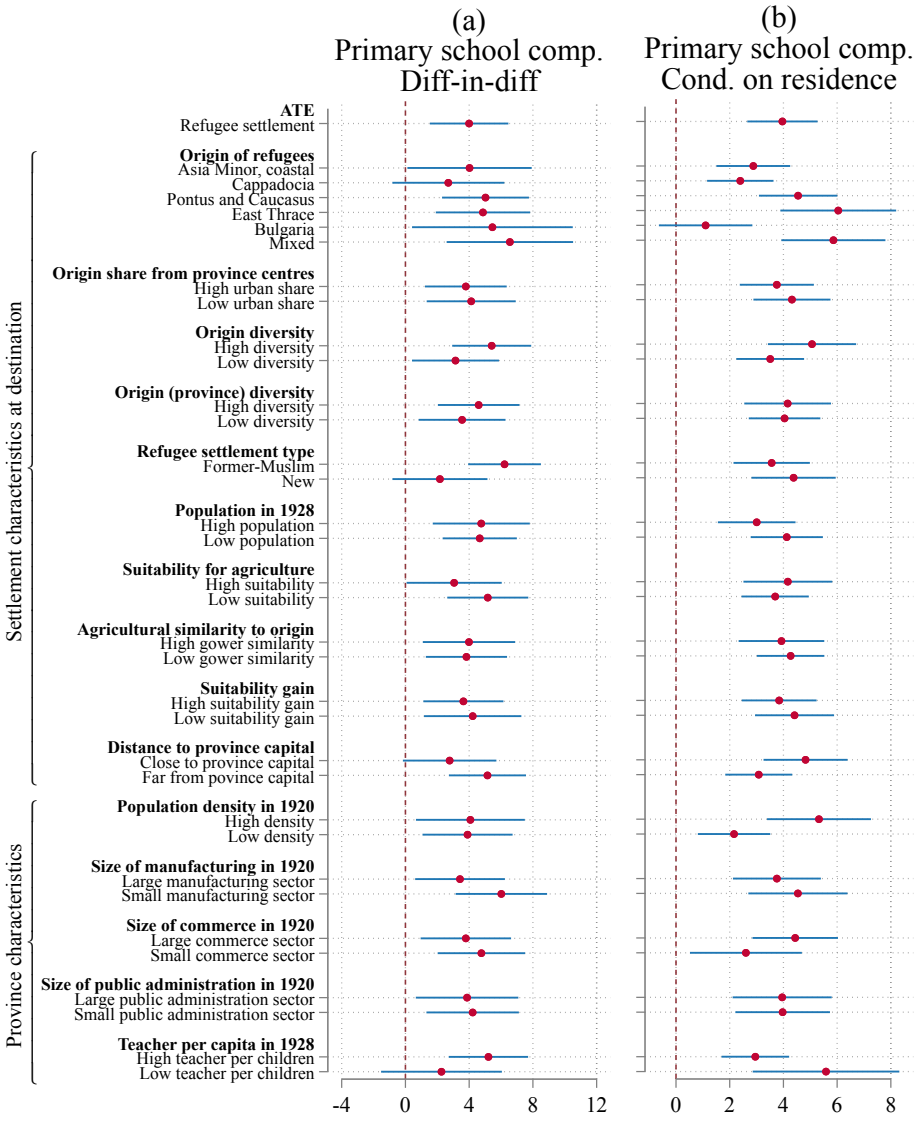
Notes. The table gives summary statistics of various ethnographic, geographic, and locational features across predominantly refugee settlements, defined as those where the share of refugees in 1928 exceeded 80%. Panel *A* gives unweighted summary statistics across all rural predominantly refugee settlements. Panel *B* gives statistics weighting the settlements by their population in 1928.

Table D2: Refugee Origin Settlement Characteristics and Displacement Trajectories

	Constantinople	**Eastern Thrace**	**Asia Minor**	**Pontus**	**Cappadocia**
Settlement	Mainly in Athens connections with Greek cities (trading)	West Thrace	Aegean Sea Islands Athens Sterea Elada	Rural Macedonia, Thessaloniki	Central Northern Greece
Trajectory	Peacefully, transferring assets Trade, Services	Cross Evros Fall 1922 carry movables and animals	Repression since 1908 Exodus, Fire of Smyrna Attacks by Turkish Army	Post Population Exchange Convention (1923); Some via Caucasus, Russia	Post Population Exchange Convention mid/late 1923 & 1924
Main Activity					
Agriculture	0.43	0.71	0.63	0.71	0.71
Herding	0.00	0.00	0.02	0.09	0.03
Fishing	0.29	0.00	0.08	0.00	0.00
Merchants	0.29	0.14	0.20	0.10	0.21
Mining	0.00	0.14	0.08	0.04	0.05
Main Agriculture					
Tobacco	0.00	0.00	0.12	0.21	0.00
Cereal	0.40	0.33	0.29	0.59	0.68
Silk Cocoons	0.00	0.33	0.39	0.01	0.00
Vineyards	0.00	0.33	0.21	0.02	0.32
Language					
Turkish Spoken	0.26	0.25	0.37	0.49	0.75
Only Turkish	0.00	0.00	0.14	0.31	0.57
Only Pontic	0.00	0.12	0.08	0.26	0.04
Only Greek	0.74	0.62	0.53	0.21	0.00
Only Cappadocian	0.00	0.00	0.00	0.00	0.20
# Settlements	19	8	333	250	75

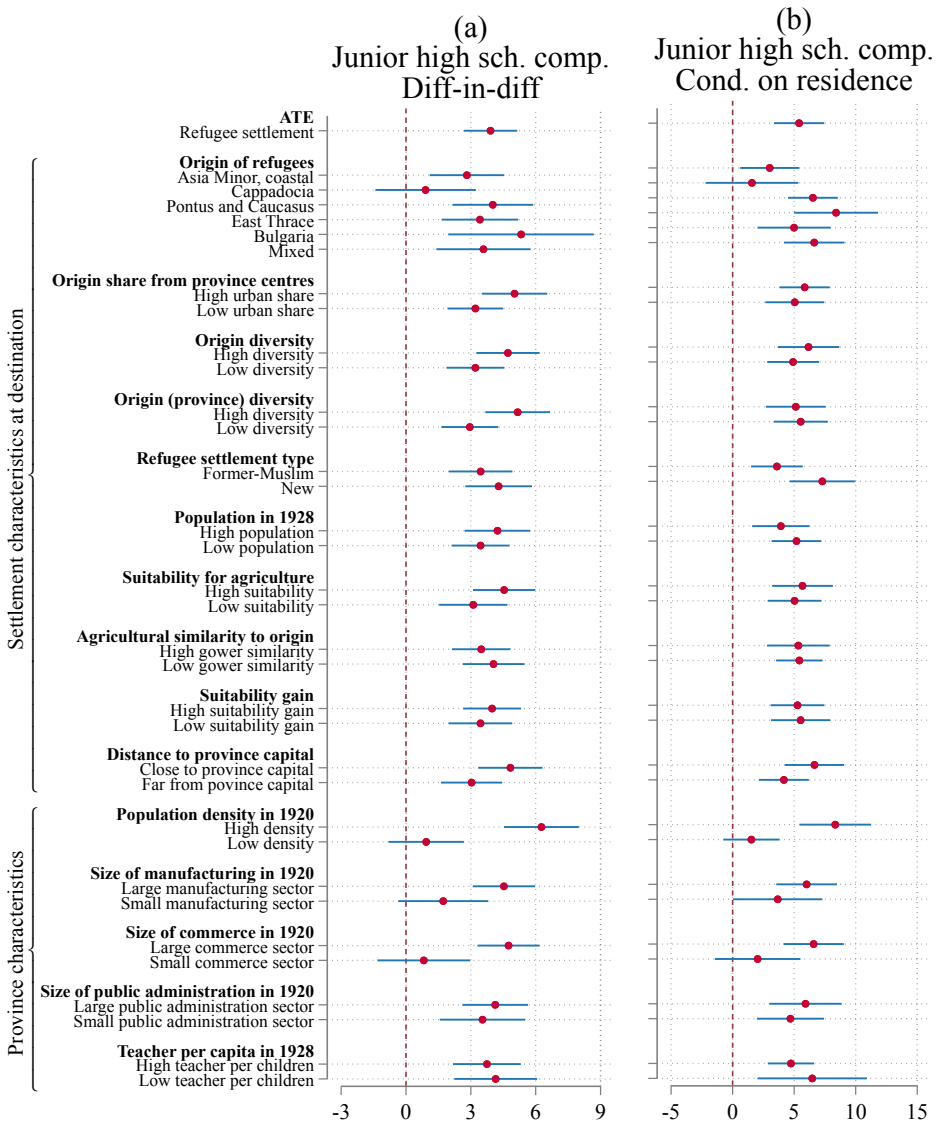
Notes. The table summarizes the main characteristics and displacement trajectories for Greek-Orthodox refugees from five broad regions in Anatolia: (i) Constantinople, (ii) Eastern Thrace, (iii) Coastal Asia Minor, (iv) Pontus, and (v) Cappadocia. The underlying data on main occupation, agricultural production, and language spoken are retrieved from the seconding of 1,500 interviews (oral histories) from the archive of *the Center for Asian Minor Studies*, *Xenophanes: Periodical Publication of the Association of Asia Minor Greeks "Anatoli"*, and Sia Anagnostopoulou (1997) monograph, "The Greek Orthodox Communities of Asia Minor."

Figure D1: Displacement and Primary School Completion by Settlement Characteristics (1971–2011)



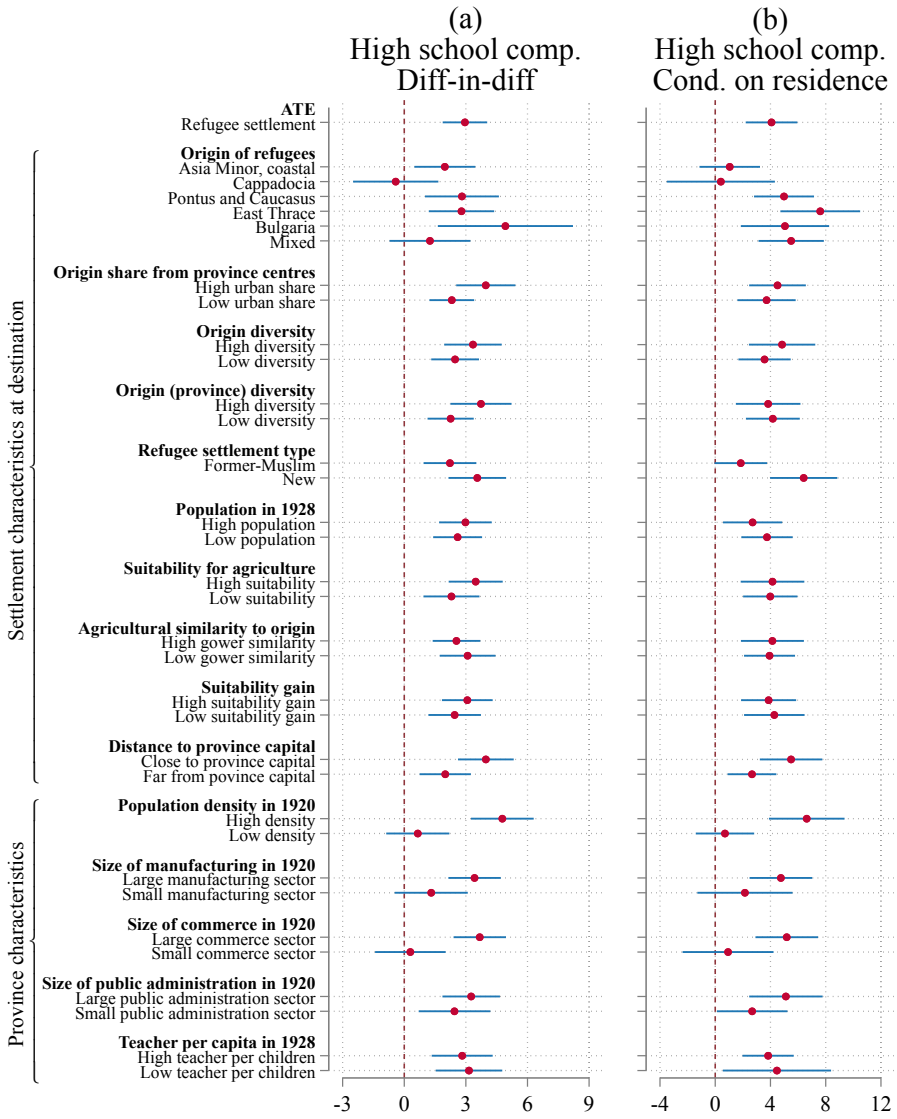
Notes. The panels present the heterogeneous treatment effects of difference-in-difference estimates (Panel *A*) and the estimates conditional on residence (Panel *B*) by settlement characteristics at destination and settlement province characteristics. For each characteristic we consider, we split the sample into two by the population-weighted median value of the characteristic at the settlement level in 1928; see Panel *B* of Appendix Table D1 to see the cutoff values. The confidence intervals in Panel *A* are based on standard errors clustered at the settlement-of-residence level, while in Panel *B* are based on standard errors double clustered at the birthplace-settlement and the municipality-of-residence levels. The difference-in-difference specifications use micro samples from the 1971, 1981, 1991, 2001, and 2011 General Population Censuses. The conditional-on-residence specifications use the micro-sample from the 2011 General Population Census.

Figure D2: Displacement and Junior High School Completion by Settlement Characteristics (1971–2011)



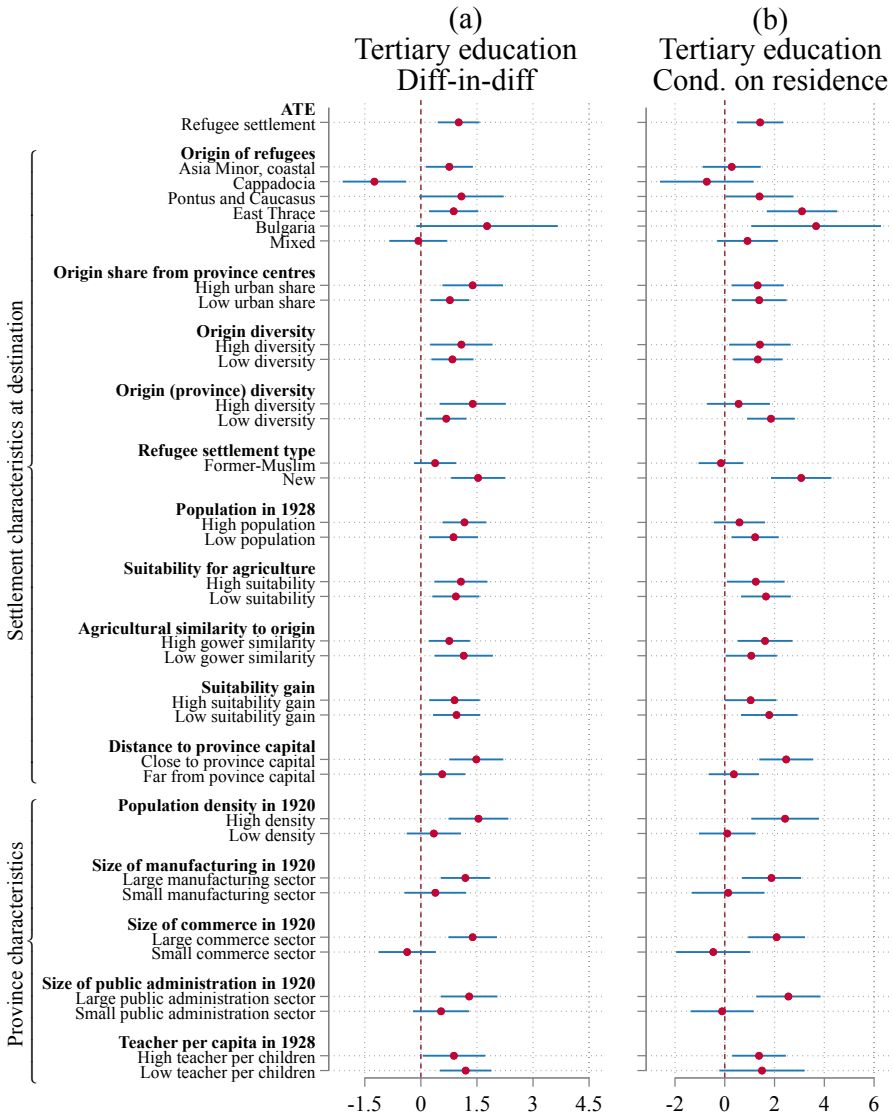
Notes. The panels present the heterogeneous treatment effects of difference-in-difference estimates (Panel *A*) and the estimates conditional on residence (Panel *B*) by settlement characteristics at destination and settlement province characteristics. For each characteristic we consider, we split the sample into two by the population-weighted median value of the characteristic at the settlement level in 1928; see Panel *B* of Appendix Table D1 to see the cutoff values. The confidence intervals in Panel *A* are based on standard errors clustered at the settlement-of-residence level, while in Panel *B* are based on standard errors double clustered at the birthplace-settlement and the municipality-of-residence levels. The difference-in-difference specifications use micro samples from the 1971, 1981, 1991, 2001, and 2011 General Population Censuses. The conditional-on-residence specifications use the micro-sample from the 2011 General Population Census.

Figure D3: Displacement and High School Completion by Settlement Characteristics (1971–2011)



Notes. The panels present the heterogeneous treatment effects of difference-in-difference estimates (Panel *A*) and the estimates conditional on residence (Panel *B*) by settlement characteristics at destination and settlement province characteristics. For each characteristic we consider, we split the sample into two by the population-weighted median value of the characteristic at the settlement level in 1928; see Panel *B* of Appendix Table D1 to see the cutoff values. The confidence intervals in Panel *A* are based on standard errors clustered at the settlement-of-residence level, while in Panel *B* are based on standard errors double clustered at the birthplace-settlement and the municipality-of-residence levels. The difference-in-difference specifications use micro samples from the 1971, 1981, 1991, 2001, and 2011 General Population Censuses. The conditional-on-residence specifications use the micro-sample from the 2011 General Population Census.

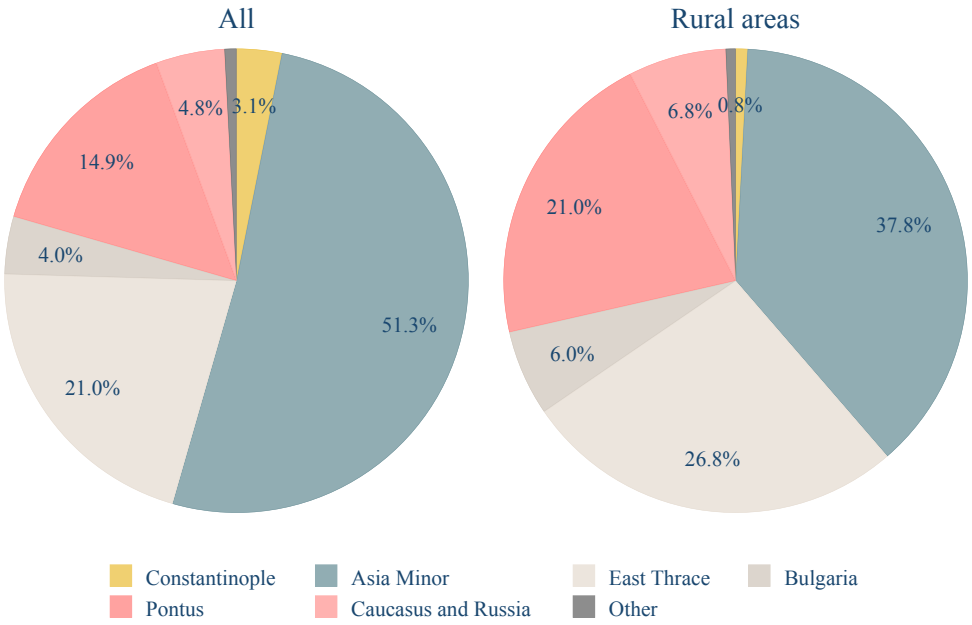
Figure D4: Displacement and Tertiary Education by Settlement Characteristics (1971–2011)



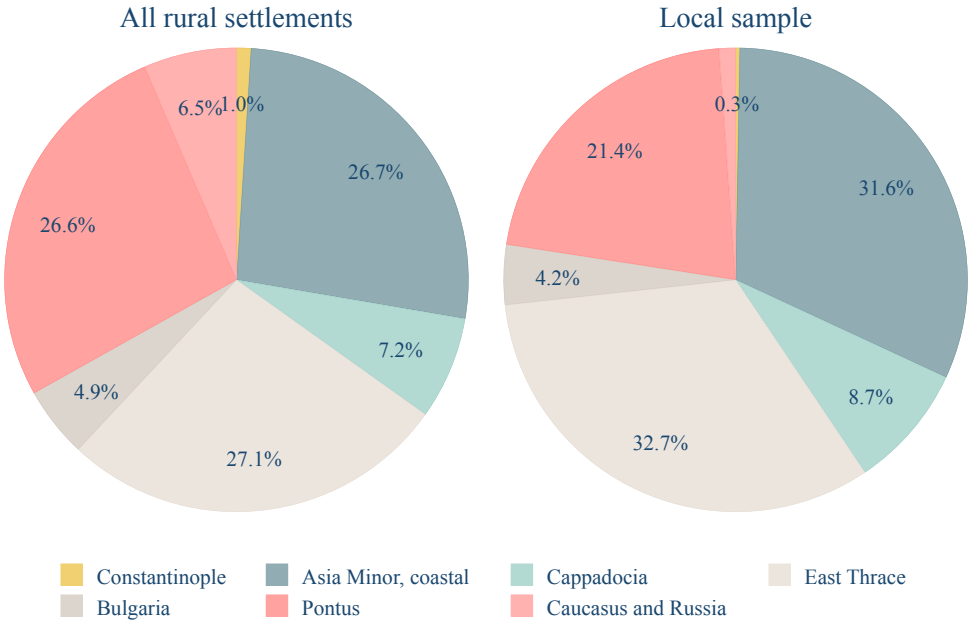
Notes. The panels present the heterogeneous treatment effects of difference-in-difference estimates (Panel *A*) and the estimates conditional on residence (Panel *B*) by settlement characteristics at destination and settlement province characteristics. For each characteristic we consider, we split the sample into two by the population-weighted median value of the characteristic at the settlement level in 1928; see Panel *B* of Appendix Table D1 to see the cutoff values. The confidence intervals in Panel *A* are based on standard errors clustered at the settlement-of-residence level, while in Panel *B* are based on standard errors double clustered at the birthplace-settlement and the municipality-of-residence levels. The difference-in-difference specifications use micro samples from the 1971, 1981, 1991, 2001, and 2011 General Population Censuses. The conditional-on-residence specifications use the micro-sample from the 2011 General Population Census.

Figure D5: Refugees' Origin

(a) 1928 General Population Census



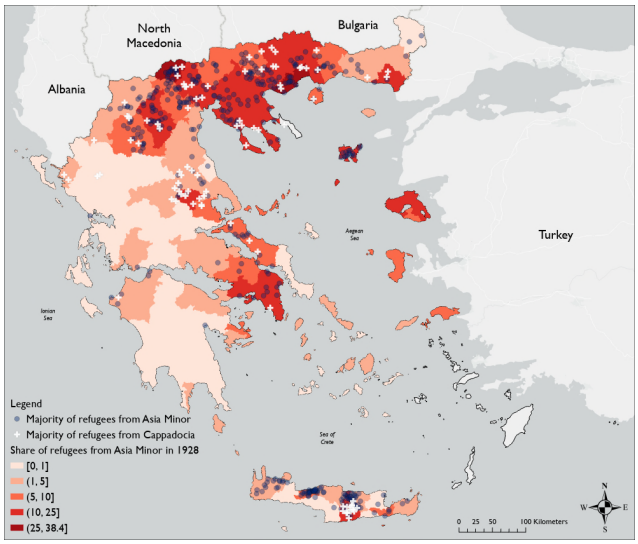
(b) Rural Catalog



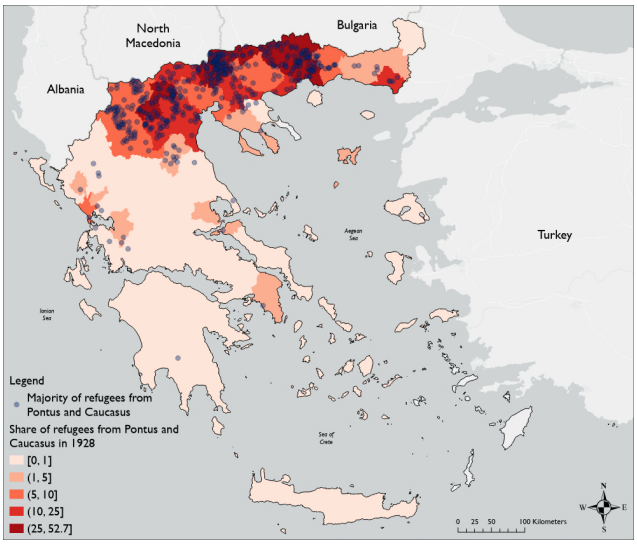
Notes. The Figures present the distribution of refugees based on region of origin. Panel *A* uses information on region of origin coming from the 1928 General Population Census and presents the share of refugees from different regions across all population (left) and rural population (right). Panel *B* uses information coming from the place of origin of refugees with land grant coming from the Rural Catalog and presents the share of refugees from different regions in the all rural settlements with a land grant (left) and among rural settlements with a land grant in the local sample (right), which focuses on a sample of rural predominantly-refugee and native settlements within a 25 km radius.

Figure D6: Refugee's Origin and Rural Settlement in Greece

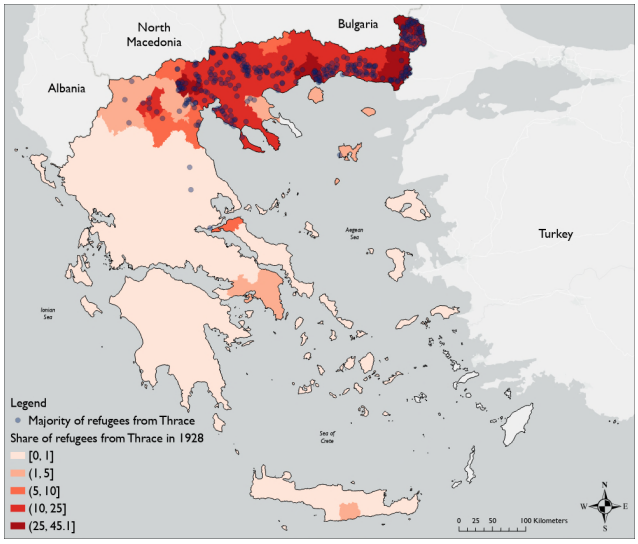
(a) Asia Minor and Cappadocia



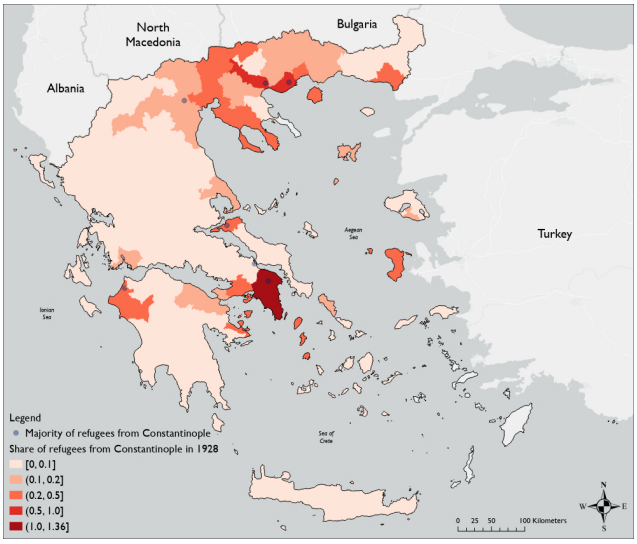
(b) Pontus and Caucasus



(c) Eastern Thrace



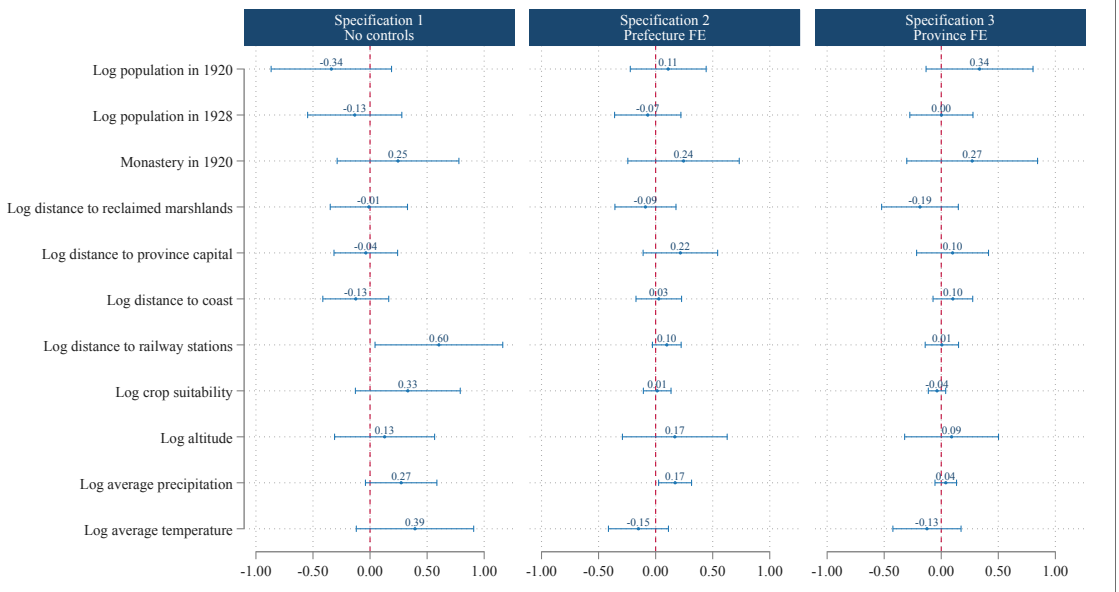
(d) Constantinople



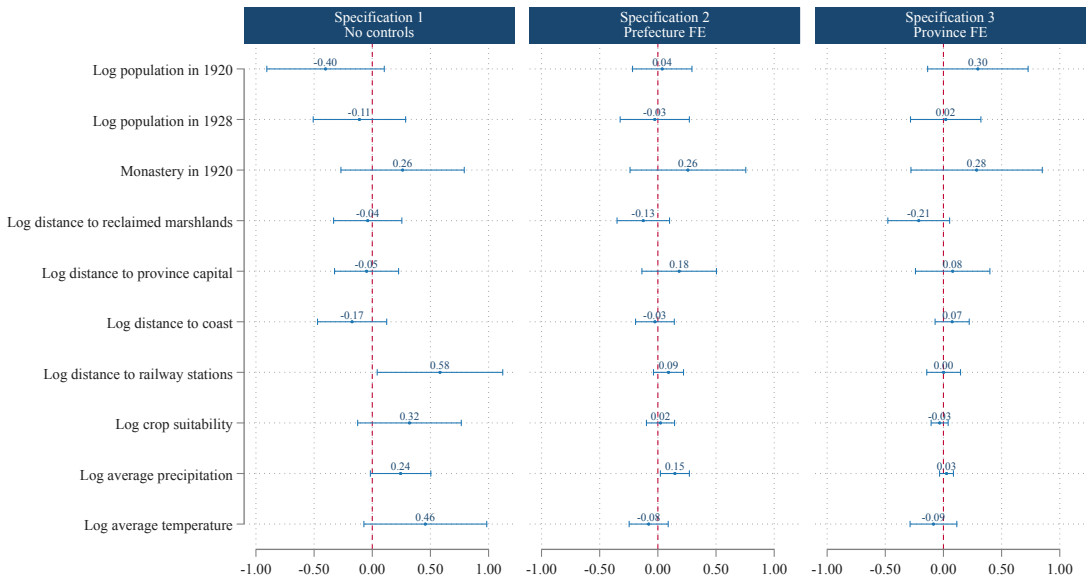
Notes. The Figure plots the share of refugees from Asia Minor and Cappadocia (Panel *A*), Pontus and Caucasus (Panel *B*), Eastern Thrace (Panel *C*), and Constantinople (Panel *D*) over the total population across 141 provinces (admin-3 level units) based on the breakdown provided by the 1928 Population Census. The 1928 Population Census tabulates refugees from Cappadocia as part of the refugees from Asia Minor. Dots (and plus signs in Panel *A*) portray settlements where refugees from each respective region in Anatolia received the majority of land grants, using data from the Rural Refugee Catalog compiled by the Refugee Settlement Committee in the mid-1920s.

Figure D7: Balance of observable characteristics across majority-Cappadocian and other refugee settlements

(a) Unconditional



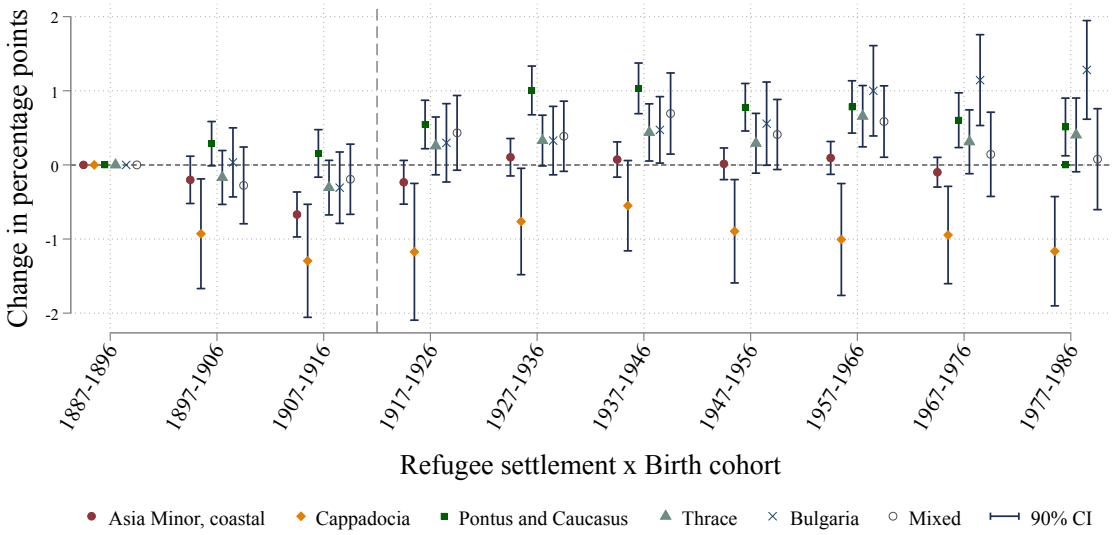
(b) Conditional on log altitude



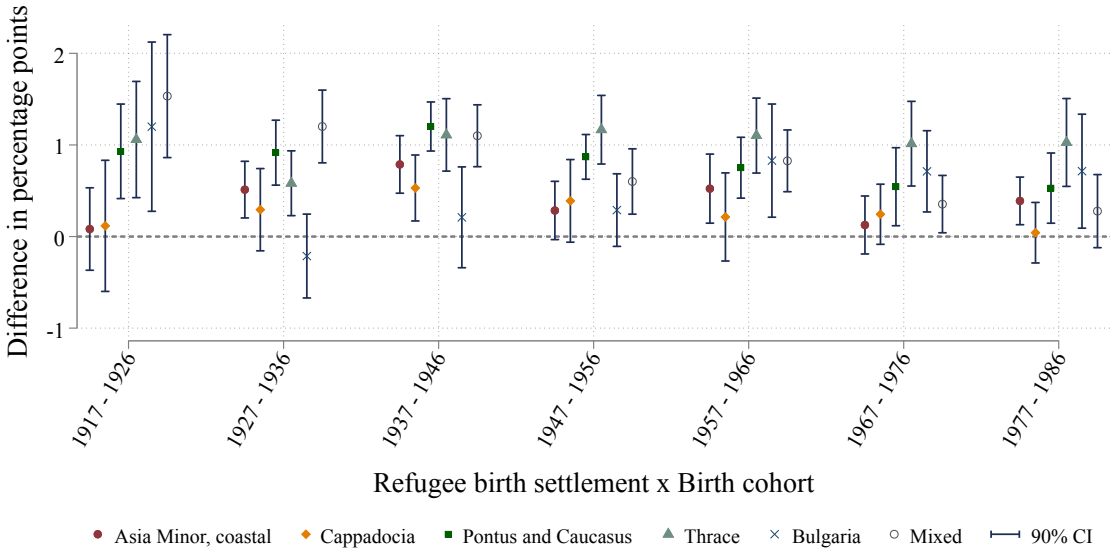
Notes. The panels explore whether various geographic, location, and population features differ between majority-Cappadocian refugee and other refugee settlements across all predominantly refugee settlements in the baseline sample. All specifications in both Panels give standardized coefficients and 90% confidence intervals obtained from regressing standardized settlement characteristics on a dummy variable indicating predominantly refugee settlements, where the share of refugees in 1928 exceeds 80%; the indicator equals zero for native settlements, where the share of refugees in 1928 is below 5%. Panel *A* presents unilateral unconditional specifications. Panel *B* gives the regression estimates conditioning on settlements log altitude. Specification 1 does not include any controls. Specification 2 conditions on prefecture fixed effects. Specification 3 conditions on province fixed-effects. Standard errors are clustered at the province level.

Figure D8: Forced Displacement and Schooling by Origin. Cohort-Level Estimates

(a) Difference-in-differences, 1971–2011



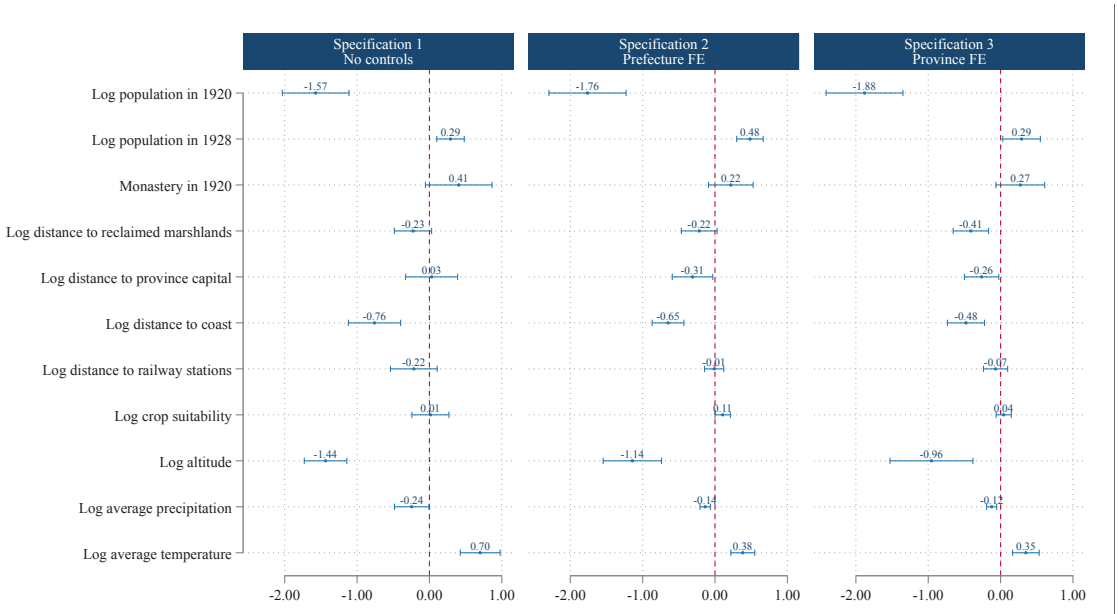
(b) Conditional on residence, 2011



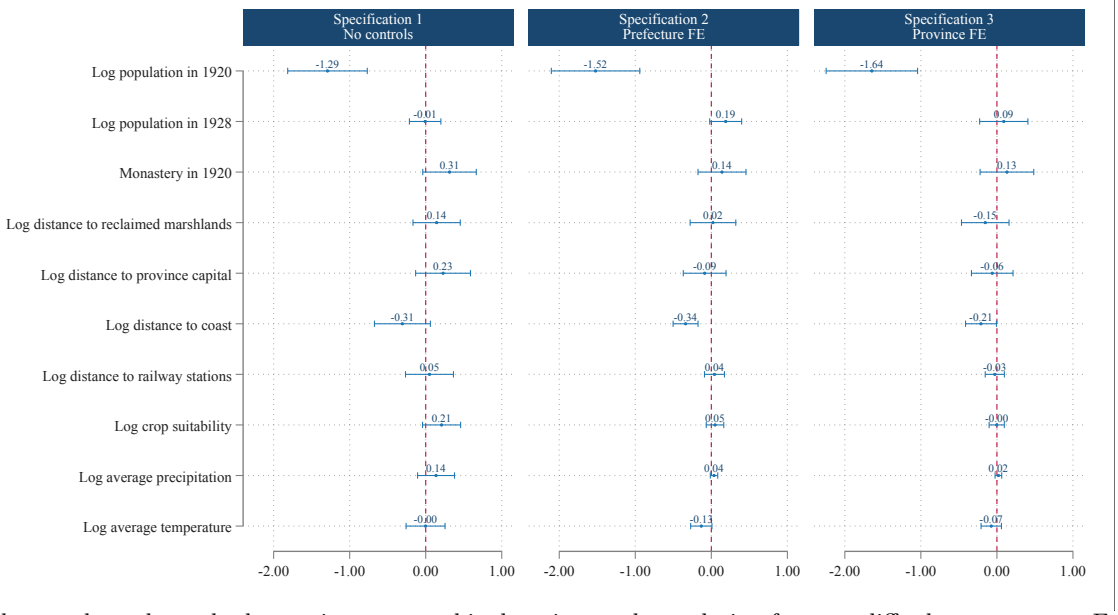
Notes. Panel *A* plots the event-study OLS estimates identifying the time-varying effect of residing in a predominantly refugee settlement interacted with the origin region of the majority of refugees with a land grant; see Figure 5 for the list of covariates. The 1887–96 birth cohort serves as the excluded cohort. Panel *B* plots the cohort-level estimates obtained from regressing years of schooling on a dummy variable indicating whether an individual was born in a predominantly refugee settlement interacted with the origin region of the majority of refugees with a land grant; see Appendix Figure B5. The confidence intervals in Panel *A* are based on standard errors clustered at the settlement-of-residence level, while in Panel *B* is based on standard errors double clustered at the birth-settlement and the municipality-of-residence levels. The difference-in-difference specifications use micro samples from the 1971, 1981, 1991, 2001, and 2011 General Population Censuses. The conditional-on-residence specifications use the micro-sample from the 2011 General Population Census.

Figure D9: Balance of observable characteristics across predominantly-Muslim and predominantly-non-Muslim refugee settlements

(a) Unconditional



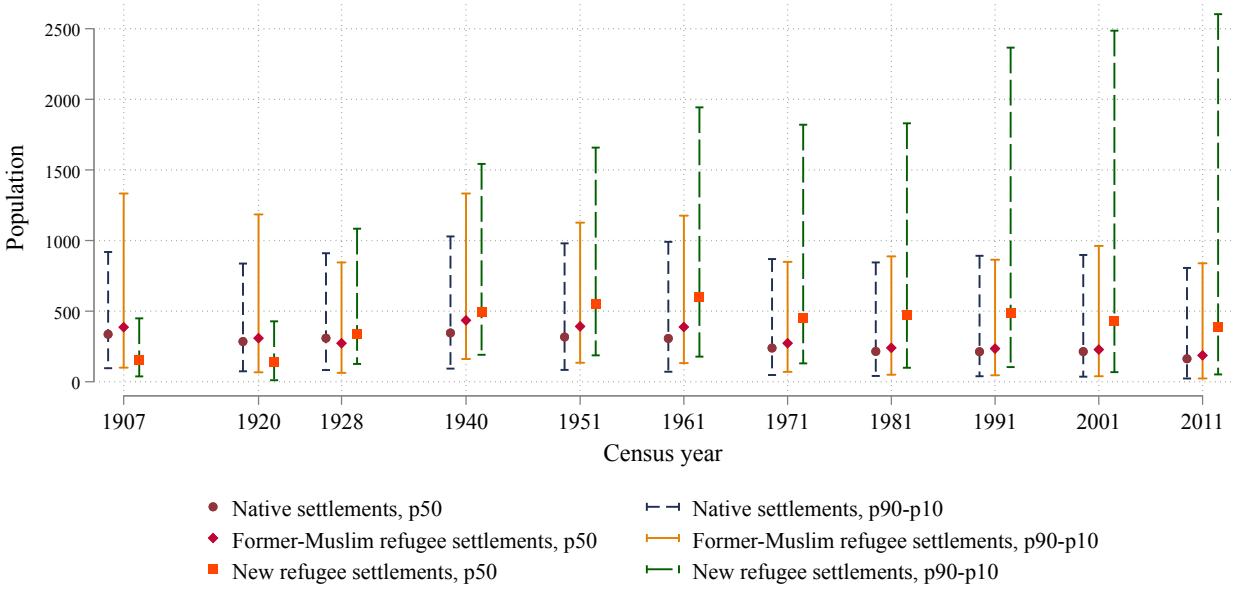
(b) Conditional on log altitude



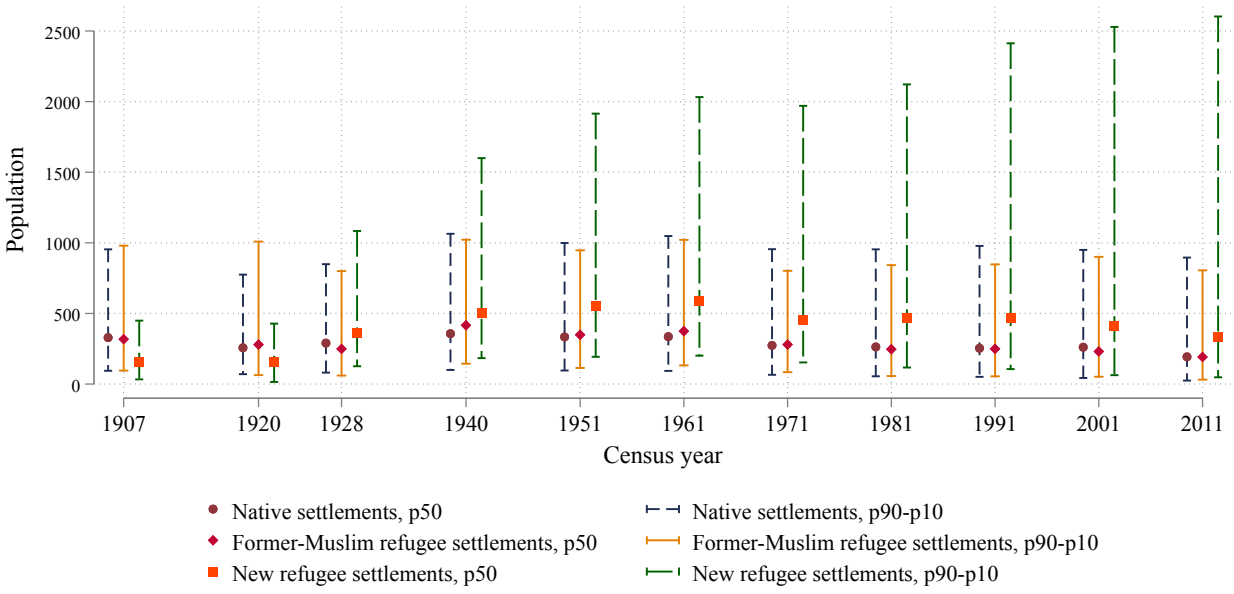
Notes. The panels explore whether various geographic, location, and population features differ between across Former-Muslim refugee settlements and new refugee settlements across all predominantly refugee settlements in the baseline sample. All specifications in both Panels give standardized coefficients and 90% confidence intervals obtained from regressing standardized settlement characteristics on a dummy variable indicating predominantly refugee settlements, where the share of refugees in 1928 exceeds 80%; the indicator equals zero for native settlements, where the share of refugees in 1928 is below 5%. Panel *A* presents unilateral unconditional specifications. Panel *B* gives the regression estimates conditioning on settlements log altitude. Specification 1 does not include any controls. Specification 2 conditions on prefecture fixed effects. Specification 3 conditions on province fixed-effects. Standard errors are clustered at the province level.

Figure D10: Population Dynamics. Rural Native and (Former-Muslim and New) Refugee Settlements

(a) All rural native and refugee settlements



(b) Local sample



Notes. The figures plot the evolution of the median and the 90th- and 10th-percentiles of the population in rural refugee and native settlements between 1907 and 2011, as recorded in the respective Censuses. Refugee villages are split into two categories: (i) Former-Muslim refugee settlements where the (imputed) share of Muslims in 1920 exceeds 80%; (ii) New refugee settlements with a corresponding share of the Muslim population in 1920 below 80%. Panel *A* plots the population statistics for all rural and native refugee settlements. Panel *B* focuses on the native and refugee settlements in the local analysis sample.

Appendix References

Anagnostopoulou, S. (1997): *Greek Orthodox Communities in Asia Minor - 19th century - 1919*. Ellenika Grammata, Athens.

Bester, C. A., T. G. Conley, and C. B. Hansen (2011): "Inference with Dependent Data Using Cluster Covariance Estimators," *Journal of econometrics*, 165(2), 137–151.

Colella, F., R. Lalive, S. O. Sakalli, and M. Thoenig (2019): "Inference with Arbitrary Clustering," *IZA DP No. 12584*.

——— (2023): "acreg: Arbitrary Correlation Regression," *The Stata Journal*, 23(1), 119–147.

Conley, T. G. (1999): "GMM estimation with cross-sectional dependence," *Journal of Econometrics*, 92(1), 1–45.

Güvenç, S. (2010): *Atlas of Old and New Toponyms of Northern Greece: Before and After the Population Exchange [Mübadele Öncesi ve Sonrası Eski ve Yeni Adları ile Kuzey Yunanistan Yer Adlari Atlası]*. Golden Medya Matbaacılık, Istanbul.

Minasidis, C. (2024): ""This Expulsion is Explained in Many Ways": Ottoman Greek Orthodox Internal Exiles during the Great War (1914–1918)," *New Perspectives on Turkey*, 70, 89–106.