SUPPLEMENTARY MATERIAL

GROUP TIES AMID INDUSTRIAL CHANGE
Historical Evidence from the Fossil Fuel Industry

By Noah Zucker

World Politics
doi: 10.1017/S0043887122000168

Replication data are available at:

Zucker, Noah. 2022. "Replication data for: Group Ties Amid Industrial Change: Historical Evidence from the Fossil Fuel Industry." Harvard Dataverse, V l. doi: 10.7910/DVN/PRRXMM.

Group Ties amid Industrial Change: Historical Evidence from the Fossil Fuel Industry
Supplementary Material

A. Coal Data and Production Measurement

All coal mine data are collected from editions of *Mineral Resources of the United States*, which was regularly published by the U.S. Geological Survey between 1882 and 1931 (annually between 1900 and 1931) and is now available through the HathiTrust Digital Library. Because the quality of these records is not sufficient for OCR transcription, I transcribed all data manually.

These documents contain information on the amount of coal produced, value of coal produced, mine employment, and mine working days at the county-year level. I collected production data for all available years between 1900 and 1929. In most cases, the reports list these data for individual counties. But there are some cases where counties are aggregated together for particular years owing to relatively low levels of production in some of these counties. Because I do not have information on the breakdown of production across combined counties, I exclude these observations from all analyses. As noted in the main text, I also exclude observations for counties that changed borders in a given decade; I use data from Horan and Hargis, 1995, "County Longitudinal Template, 1840–1990," ICPSR.

To measure industrial shocks, I focus on the quantity of coal produced. The per-unit value of coal produced often did not change much year to year. I do not use employment data either, as coal mines at this time featured large numbers of informal workers not included in official employment statistics. Though data on working days bypasses this issue of informal employment, this variable does not provide information on the amount of work being done on any given day.

For county-years in which no production information is listed, I assume that there was no active production (i.e., production recorded as zero in the dataset). The one exception is for anthracite mines in northeastern Pennsylvania. Between 1916 and 1921, these reports did not include county- level information on mining in these areas; data was aggregated up to the coalfield level, which covered multiple counties. In these cases, production is recorded as missing (NA); these observations are then excluded.

"Shock intensity" is the sum of all year-over-year percentage declines in a given county-decade (i.e., the sum of year-over-year percentage changes, limited to decline years). Decades are defined as the year one census was enumerated until the year before the next census was enumerated (e.g., 1900 through 1909); this is because censuses were enumerated early in the year. I opt for this measure to separate counties that witnessed severe intra-decade production declines from counties with consistently growing coal mines — the key conceptual distinction described in the theory. As shown in Figure A1, counties with low "shock intensities" experienced consistent growth over the course of a given decade. Counties with high "shock intensities" experienced either absolute declines in production during a decade or sharp negative shocks.

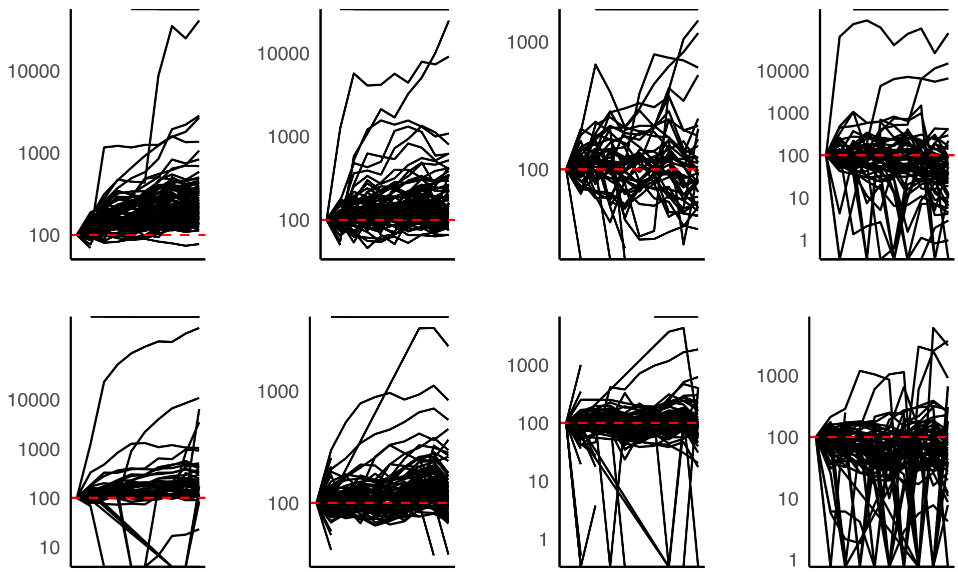

Figure A1: Yearly coal production by county, indexed to production at start of decade (= 100). Divided according to *decade* (1900s- top, 1910s- bottom) and *quartile of shock intensity* (low- left, high- right).

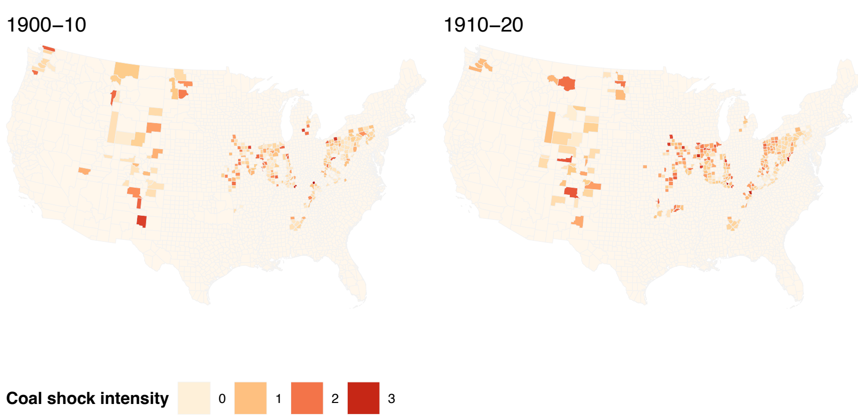

Figure A2: Map of shock intensity by county and decade.

A.1. Year-over-Year Production Changes

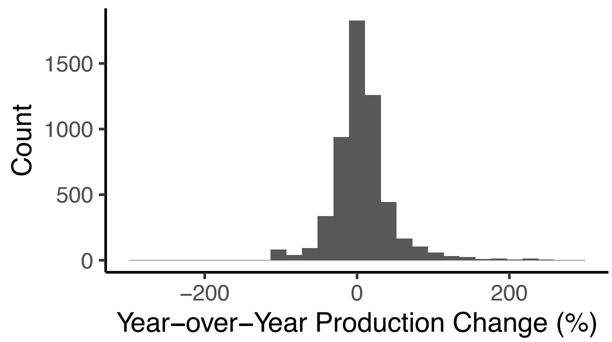

Figure A3: Histogram of year-over-year percentage changes in county-level coal production. The mean absolute value YoY change was 60.3%; the median absolute value change was 17.5%. When mines declined YoY, they typically did so sharply (mean YoY decline of −22.2%, median −15.4%). Slow-and-steady declines were uncommon at this time, as were slow-and-steady periods of growth.

B. Census Data and Group Measurement

Identifying ethnic groups using U.S. census data is not a straightforward task. Immigrant groups may be defined according to individuals' birthplaces or language. But because of changing national borders in Europe and revisions to enumerator instructions in the early 20th century, countries of birth were often defined differently year to year. Relying on country of birth is also not appropriate for ethnically diverse countries. Reliable language information is additionally not available for all immigrants.

To limit these issues, I define immigrant groups according to both birthplace and language. I primarily classify immigrants according to birthplace (`BPL` and `BPLD` in IPUMS USA data) but use information on their mother tongues (`MTONGUE` and `MTONGUED`) to distinguish between ethnic groups from more diverse countries and correct for changes in birthplace definitions over time (e.g., all immigrants listing Polish as their mother tongue are classified as Polish immigrants regardless of birthplace). When birthplace and language information conflict, language is prioritized. The following is a list of the immigrant groups included and the coding process for each, as well as the number of matched immigrants classified into each group.

1. Danish: birthplace (7,006 matched immigrants)
2. Finnish: birthplace (2,277)
3. Icelandic: birthplace (5)
4. Norwegian: birthplace (13,147)
5. Swedish: birthplace (50,853)
6. English: birthplace (103,516)
7. Scottish: birthplace (34,782)
8. Welsh: birthplace (33,975)
9. Irish: birthplace (66,405)
10. Belgian: birthplace (6,695)
11. French: birthplace (8,911)
12. Luxembourgish: birthplace (62)
13. Dutch: birthplace (3,697)
14. Swiss: birthplace or mother tongue listed as "Swiss" (12,120)
15. Greek: birthplace (457)
16. Italian: birthplace (38,173)
17. Portuguese: birthplace (192)
18. Spanish: birthplace (151)
19. Austrian: birthplace or mother tongue listed as "Austrian" (52,852)
20. Bulgarian: birthplace or mother tongue listed as "Bulgarian" (52)
21. Czechoslovak: birthplace; Czech if mother tongue listed as "Czech," Slovak if mother tongue listed as "Slovak" (1,622)

22. German: birthplace or mother tongue listed as “German” (201,228)
23. Hungarian: birthplace or mother tongue listed as “Magyar, Hungarian” (21,487)
24. Polish: birthplace or mother tongue listed as “Polish” (7,829)
25. Rumanian: birthplace or mother tongue listed as “Rumanian” (729)
26. Croatian: birthplace or mother tongue listed as “Croatian” (60)
27. Dalmatian, Montenegrin: mother tongue listed as either “Dalmatian” or “Montenegrin” (0)
28. Serbian: birthplace or mother tongue listed as “Serbian” (29)
29. Slovenian: birthplace or mother tongue listed as “Slovene” (52)
30. Estonian: birthplace (0)
31. Latvian: birthplace (0)
32. Lithuanian: birthplace or mother tongue listed as “Lithuanian” (215)
33. Russian: birthplace (Russian Empire, excluding Baltic states, Armenia, Ukraine) or mother tongue listed as “Russian” (30,228)
34. Ukrainian: birthplace or mother tongue listed as “Ukrainian” (1)
35. Armenian: birthplace or mother tongue listed as “Armenian” (1)
36. Jewish: mother tongue listed as “Yiddish, Jewish” or “Hebrew, Israeli” (0)

B.1. Association between Group Concentration and Future Changes in Coal Production

My identifying assumption is that groups within the same county did not vary in their concentration in local mines according to future production changes in those mines. To test this, I regress shocks to county coal mines on group-specific concentrations in those mines (interaction of county-level group concentrations with group fixed effects).

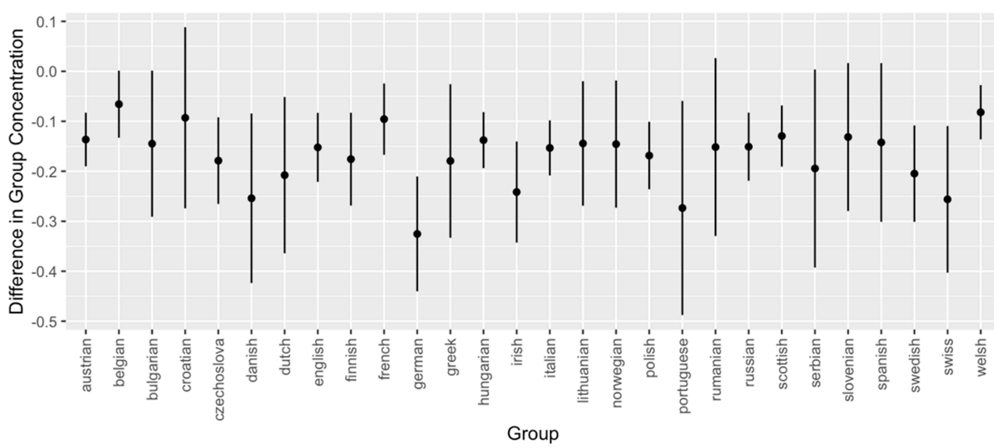

Figure B1: Regression of shocks to county coal (shock intensity measure) on group-specific concentrations in those mines. Coefficients on group concentration variable plotted with 95% confidence intervals.

Figure B1 indicates that while groups were generally less concentrated in coal mines on the precipice of decline, this tendency was consistent across groups — note the similar coefficient magnitudes and overlapping confidence intervals. There is little evidence that some groups sorted into or out of coal more than others according to future shock to the local mines. Within-county comparisons (regressions with county fixed effects) across groups should accordingly be valid.

B.2. Groups per County

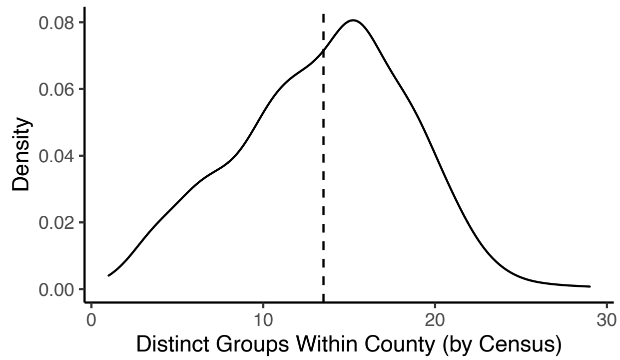

Figure B2: Number of European ethnic groups in the linked dataset by county and census. Dashed vertical line indicates the mean number of groups (13.5). Across censuses, 99.4% of counties include more than one European ethnic group in the analysis dataset.

C. Additional Calculation Details

C.1. Estimated Income from Wages

Income information was not collected prior to the 1940 census. I accordingly use data from that census (`INCWAGE` in IPUMS) to estimate individuals' pre-tax income from wages and salary. Limiting the 1940 census to all non-agricultural workers, I calculate the median income for every permutation of state of residence, occupation, birthplace, race, and sex, which I then apply to individuals in earlier censuses. This approach is based on what scholars have done previously (e.g., Abramitzky et al., 2021, "Intergenerational Mobility of Immigrants in the United States over Two Centuries," *American Economic Review*.).

C.2. Bartik Estimate of Non-Coal Shocks

I record for each county, at the start of each decade, the proportion of workers employed in each of six major industry categories: (1) durable goods manufacturing, (2) non-durable goods manufacturing, (3) agriculture, (4) petroleum and natural gas extraction, (5) metallic mining, and (6) non-metallic mining (note the exclusion of coal mining). I identified workers in each industry category using `IND1950` classifications in IPUMS datasets. I then draw on annual national production data for each industry category, gathering this from U.S. Bureau of the Census (1975). Due to its availability each year, I use data on the value of production for each industry category; for agriculture, I use a wholesale price index. For each industry category, I compute the same "shock intensity" measure as I did for coal. I then calculate a weighted average across these categories using counties' initial employment shares, taking the square root due to a rightward skew.

This calculation may be written as

$$\text{Non-coal shock intensity}_{c(t\to t+9)} = \sqrt{\frac{\sum_i^6 \frac{\text{industry workers}_{ict}}{\text{all workers}_{ct}} \times \text{shock intensity}_{i(t\to t+9)}}{6}}$$

where *i* indexes the six industry categories listed above, *c* indexes counties, and *t* indexes years.

C.3. Congressional Election Data

I collected Democratic and Republican congressional vote share data from Clubb, Flanigan, and Zingale, 2006, "Electoral Data for Counties in the United States: Presidential and Congressional

Races, 1840–1972," ICPSR. Republican strongholds are those where the Republican vote share exceeded the Democratic share by at least 12.6 points (median difference in sample).

C.4. Observation Weights

I estimate a probit model to predict individuals' probabilities of being matched as a function of their age, sex, literacy, urban or rural residence, immigrant status, and occupational prestige. This follows the recommendation of Abramitzky, Boustan, and Rashid, 2020, "Census Linking Project: Version 1.0 [dataset]." Occupational prestige reflects occupation-specific incomes in 1950 (OCCSCORE in IPUMS USA data). Following Abramitzky, Boustan, and Rashid, the age variable is collapsed into ten bins (0–9, 10–19, ..., 90–) and the occupation variable is collapsed into six bins (0–9, 10–19, ..., 50–). Weights are calculated according to the first year of a given range (e.g., 1900 for 1900–10). The following describes the weight calculation for individual *i* and matching range *t*:

$$\text{weight}_{it} = \frac{1 - \text{predicted probability of having been matched}_{it}}{\text{predicted probability of having been matched}_{it}} \times \frac{\%\ \text{matched}_t}{1 - \%\ \text{matched}_t}$$

D. Summary Statistics and Information on Matched Immigrants

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Group concentration	698,546	0.242	0.266	0.000	0.022	0.441	1.000
Coal shock intensity	698,809	0.648	0.480	0.000	0.279	0.875	3.332
General shock intensity (Bartik)	698,809	0.007	0.003	0.002	0.006	0.009	0.026
Naturalized (outcome)	698,809	0.647	0.478	0	0	1	1
Marriage to white citizen (outcome)	698,809	0.086	0.281	0	0	0	1
Marriage to non-coethnic immig. (outcome)	698,809	0.213	0.409	0	0	0	1
Speaks English (outcome)	684,337	0.925	0.264	0.000	1.000	1.000	1.000
Immigration year	698,809	1,883.110	14.400	1,815	1,874	1,892	1,910
Income from employment	296,065	1,160.914	495.469	20.000	839.500	1,450.000	50,000.000
Male	698,809	0.570	0.495	0	0	1	1
Living with spouse	698,809	0.851	0.356	0	1	1	1
Living with family in coal	698,809	0.104	0.306	0	0	0	1
Coethnic pop. as % county	698,809	0.032	0.028	0.00000	0.011	0.048	0.217
County coal reliance	698,809	0.136	0.139	0.0002	0.027	0.266	0.599
Rural pop. as % county	698,809	0.491	0.259	0.089	0.219	0.707	1.000
Black pop. as % county	698,809	0.021	0.036	0.000	0.004	0.034	0.404

Table D1: Summary statistics for full dataset. Statistics do not include variable transformations (e.g., square root) used in regression analyses. Due to sharp rightward skews, the following variables are transformed by square root in these analyses: group concentration, coal shock intensity, general shock intensity (Bartik), coethnic pop. as % county, county coal reliance, and Black pop. as % county. These are transformed by square root due to being bounded between 0 and 1 (or, in the case of coal shocks, largely distributed in that range). The income from employment variable is log-transformed.

	Matched immigrants	Unmatched immigrants	Immigrants elsewhere
Percentage	29.1%	70.9%	n/a
Age	41.6	41.7	43.0
Female	43.0%	35.8%	44.5%
Can read and write	89.2%	78.0%	87.8%
Rural	46.9%	45.4%	29.9%
Occupational prestige	11.7	13.0	11.4

Table D2: Mean characteristics of matched and unmatched European immigrants in coal counties, compared to European immigrants elsewhere in the country.

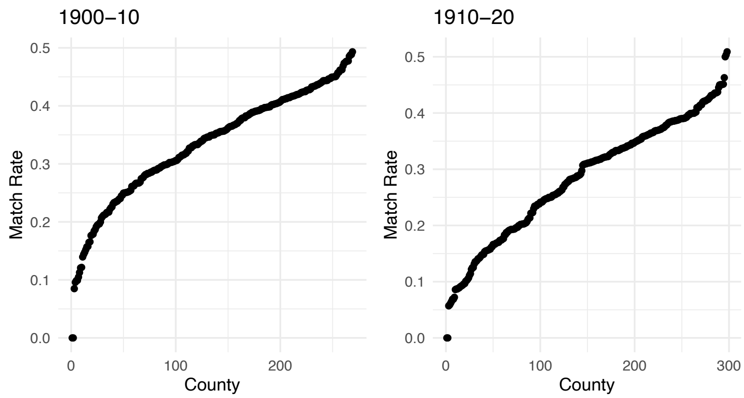

Figure D1: Match rates among European immigrants by county, in ascending order.

In Figure D2, I consider the possibility of immigrants moving back to Europe. One limitation to census linking analyses is that instances of international return migration are not directly observed; only individuals living in the U.S. during the enumeration of two censuses can be matched. To evaluate this, I test whether the interaction of group concentration and coal shocks predicts whether European immigrants in coal counties were successfully matched to the next census. One reason why individuals are not matched is departure from the country, making this an approximate measure of return migration. While match rates were generally lower for groups more concentrated in coal, I find no evidence that this relationship varied with shocks to local coal mines.

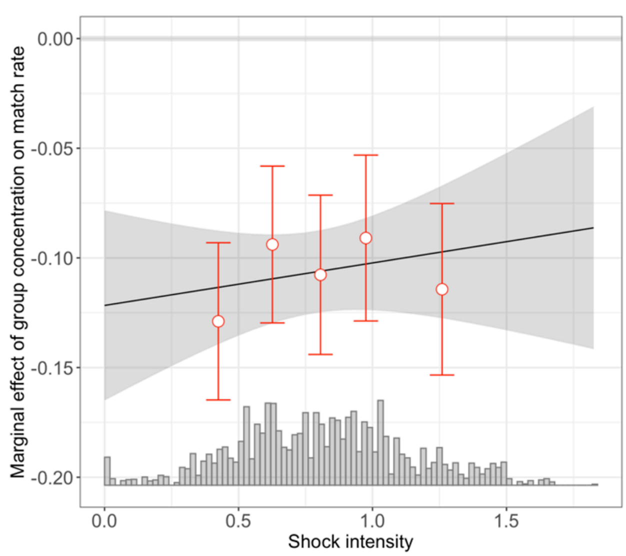

Figure D2: Regression of group match rate (at county-year level) on interaction of group concentration and shock intensity (n = 7,382). State-year and county fixed effects are included; standard errors are clustered at group-county level. Points indicate results of binned estimation (Hainmueller, Mummolo, and Xu 2019).

E. Regression Tables

	Naturalized	
	Model 1	Model 2
Group concentration	−0.652***	−0.587***
	(0.103)	(0.107)
Coal shock intensity	−0.852***	3.312
	(0.177)	(2.063)
General shock intensity (Bartik)	−6.327***	5.147
	(1.911)	(3.463)
Coethnics as % county		−0.235
		(0.259)
County coal reliance		1.377***
		(0.247)
Rural population as % county		−0.059
		(0.194)
Black population as % county		−0.979*
		(0.569)
Year of emigration		−0.007***
		(0.001)
Estimated income (ln)		0.220***
		(0.038)
Living with spouse		0.053***
		(0.016)
Coal shock × concentration	0.428***	0.487***
	(0.098)	(0.112)
Coal shock × Bartik	7.731***	−5.413*
	(1.734)	(3.121)
Coal shock × coethnic pop.		−0.176
		(0.267)
Coal shock × coal reliance		−1.172***
		(0.294)
Coal shock × rural pop.		0.077
		(0.191)
Coal shock × Black pop.		−0.359
		(0.497)
Coal shock × immig. year		−0.001
		(0.001)
Coal shock × income		−0.041
		(0.043)
Coal shock × spouse		−0.007
		(0.020)
Counties	311	298
N	75056	59503
Adj. R-squared	0.122	0.170

***p < .01; **p < .05; *p < .1

	Naturalized
Group concentration	−0.343***
	(0.073)
Coal shock intensity	−0.105**
	(0.043)
General shock intensity (Bartik)	1.543
	(1.936)
Coal shock × concentration	0.190***
	(0.074)
Group as % county × (1900–10)	0.122
	(0.153)
Group as % county × (1910–20)	−0.474***
	(0.092)
Coal reliance × (1900–10)	0.603***
	(0.165)
Coal reliance × (1910–20)	0.730***
	(0.142)
Rural pop. × (1900–10)	0.037
	(0.131)
Rural pop. × (1910–20)	−0.092
	(0.125)
Black pop. × (1900–10)	−1.343***
	(0.496)
Black pop. × (1910–20)	−1.323***
	(0.491)
Immig. year × (1900–10)	−0.008***
	(0.0004)
Immig. year × (1910–20)	−0.008***
	(0.0004)
Income × (1900–10)	0.100***
	(0.018)
Income × (1910–20)	0.230***
	(0.016)
w/ spouse × (1900–10)	0.095***
	(0.014)
w/ spouse × (1910–20)	0.032***
	(0.006)
N	59503
Adj. R-squared	0.171

***p < .01; **p < .05; *p < .1

Table E1: *Left*- Full regression table, corresponding to Figure 3 in the main text. *Right*- Replication of main regression analysis, interacting beginning-of-period covariates with year fixed effects.

	Naturalized	
	High exposure	Low exposure
Coal shock intensity	0.058**	0.010
	(0.027)	(0.022)
General shock intensity (Bartik)	0.520	−1.661***
	(1.170)	(0.618)
Year of emigration	−0.008***	−0.008***
	(0.0004)	(0.001)
Estimated income (ln)	0.191***	0.213***
	(0.017)	(0.037)
Living with spouse	0.043***	0.050***
	(0.006)	(0.008)
Coethnics as % county	−0.704***	−0.100
	(0.096)	(0.075)
County coal reliance	0.287**	0.046
	(0.117)	(0.062)
Rural population as % county	−0.161***	−0.032
	(0.053)	(0.042)
Black population as % county	−0.162	−0.349**
	(0.115)	(0.166)
N	32561	26942
Adj. R-squared	0.124	0.195

***p < .01; **p < .05; *p < .1

Table E2: Models regressing naturalization on coal shock intensity, differentiating immigrants with high group exposure (at least the median of 22.7%) from those with low exposure (below 22.7%). State-year fixed effects, errors clustered by county. Note that the coefficients on "coal shock intensity" are not distinct at a statistically significant level.

	Naturalized			
	Immigrant miners		Coal-adjacent immigrants	
	Model 1	Model 2	Model 3	Model 4
Group concentration	−0.833***	−0.789***	−0.764***	−0.751***
	(0.137)	(0.146)	(0.166)	(0.158)
Coal shock intensity	−0.717***	10.112***	−1.350***	1.065
	(0.204)	(3.564)	(0.262)	(2.480)
General shock intensity (Bartik)	−6.049***	10.302*	−10.653***	3.129
	(2.122)	(5.385)	(2.915)	(3.923)
Coethnics as % county		−0.459		0.158
		(0.313)		(0.362)
County coal reliance		1.607***		1.538***
		(0.347)		(0.294)
Rural population as % county		−0.467*		−0.061
		(0.243)		(0.227)
Black population as % county		−1.445*		−1.150
		(0.794)		(0.735)
Year of emigration		−0.005***		−0.008***
		(0.001)		(0.001)
Estimated income (ln)		0.053		0.302***
		(0.089)		(0.048)
Living with spouse		0.094***		0.042**
		(0.021)		(0.020)
Coal shock × concentration	0.536***	0.599***	0.679***	0.678***
	(0.151)	(0.170)	(0.153)	(0.162)
Coal shock × Bartik	3.333*	−11.490***	12.580***	−1.293
	(1.928)	(4.421)	(2.495)	(3.698)
Coal shock × coethnic pop.		0.079		−0.543
		(0.421)		(0.348)
Coal shock × coal reliance		−1.730***		−0.959***
		(0.382)		(0.354)
Coal shock × rural pop.		0.394		0.147
		(0.291)		(0.213)
Coal shock × Black pop.		0.614		−0.347
		(0.655)		(0.603)
Coal shock × immig. year		−0.005***		−0.0001
		(0.002)		(0.001)
Coal shock × income		0.062		−0.110**
		(0.108)		(0.052)
Coal shock × spouse		−0.079**		0.014
		(0.031)		(0.026)
N	23415	23374	50552	35503
Adj. R-squared	0.106	0.132	0.120	0.186

***p < .01; **p < .05; *p < .1

	Naturalized			
	More competitive		Republican stronghold	
	Model 1	Model 2	Model 3	Model 4
Group concentration	−0.792***	−0.612***	−0.604***	−0.491***
	(0.099)	(0.112)	(0.178)	(0.145)
Coal shock intensity	−1.254***	6.772***	−0.062	0.160
	(0.254)	(2.571)	(0.739)	(3.430)
General shock intensity (Bartik)	−5.486*	−1.080	−1.227	14.613*
	(3.247)	(6.455)	(6.990)	(8.879)
Coethnics as % county		−0.290		−0.163
		(0.283)		(0.455)
County coal reliance		0.947**		1.847***
		(0.427)		(0.432)
Rural population as % county		−0.152		0.002
		(0.351)		(0.413)
Black population as % county		0.271		−1.960
		(0.898)		(1.457)
Year of emigration		−0.005***		−0.009***
		(0.001)		(0.002)
Estimated income (ln)		0.220***		0.214***
		(0.046)		(0.059)
Living with spouse		0.024		0.101***
		(0.021)		(0.024)
Coal shock × concentration	0.689***	0.615***	0.200	0.236
	(0.112)	(0.131)	(0.162)	(0.152)
Coal shock × Bartik	10.778***	3.157	−1.535	−19.478**
	(2.578)	(5.671)	(7.241)	(8.291)
Coal shock × coethnic pop.		−0.155		−0.272
		(0.325)		(0.459)
Coal shock × coal reliance		−0.761		−1.624***
		(0.510)		(0.522)
Coal shock × rural pop.		−0.164		0.131
		(0.316)		(0.430)
Coal shock × Black pop.		−0.938		−1.264
		(0.634)		(1.326)
Coal shock × immig. year		−0.003***		0.001
		(0.001)		(0.002)
Coal shock × income		−0.058		−0.002
		(0.054)		(0.061)
Coal shock × spouse		0.037		−0.075**
		(0.026)		(0.033)
N	33942	27060	37402	29747
Adj. R-squared	0.129	0.177	0.115	0.162

***p < .01; **p < .05; *p < .1

Table E3: *Left*- Full regression table, corresponding to Figure 4 in the main text. *Right*- Full regression table, corresponding to Figure 5 in the main text.

	Central Appalachia + West		Naturalized Elsewhere		Elsewhere (matched n)
	Model 1	Model 2	Model 3	Model 4	Model 5
Group concentration	−0.230	−0.299*	−0.714***	−0.612***	−0.758***
	(0.199)	(0.166)	(0.114)	(0.119)	(0.248)
Coal shock intensity	0.509	−5.169	−0.958***	3.649*	12.116**
	(0.354)	(7.841)	(0.192)	(2.138)	(5.932)
General shock intensity (Bartik)	5.612	1.783	−7.585***	4.618	12.899
	(3.783)	(5.170)	(2.142)	(4.398)	(8.713)
Year of emigration		−0.010***		−0.007***	−0.003
		(0.003)		(0.001)	(0.003)
Estimated income (ln)		0.202**		0.220***	0.085
		(0.100)		(0.040)	(0.114)
Living with spouse		0.070		0.051***	0.041
		(0.046)		(0.016)	(0.048)
Coethnics as % county		−0.157		−0.235	−0.406
		(0.633)		(0.269)	(0.412)
County coal reliance		0.507		1.372***	1.600***
		(0.538)		(0.295)	(0.592)
Rural population as % county		0.527		−0.062	−0.363
		(0.427)		(0.213)	(0.452)
Black population as % county		0.036		−1.186*	−3.866***
		(1.081)		(0.668)	(1.349)
Coal shock × concentration	−0.079	0.090	0.496***	0.527***	0.703***
	(0.221)	(0.212)	(0.108)	(0.123)	(0.265)
Coal shock × Bartik	−2.801	3.324	8.520***	−5.219	−16.623**
	(3.730)	(6.733)	(1.879)	(3.738)	(7.238)
Coal shock × immig. year		0.003		−0.001	−0.006*
		(0.004)		(0.001)	(0.003)
Coal shock × income		−0.042		−0.040	0.069
		(0.116)		(0.045)	(0.138)
Coal shock × spouse		−0.035		−0.004	0.013
		(0.058)		(0.022)	(0.065)
Coal shock × coethnic pop.		−0.079		−0.197	−0.037
		(0.794)		(0.276)	(0.431)
Coal shock × coal reliance		0.177		−1.191***	−1.888***
		(0.704)		(0.347)	(0.659)
Coal shock × rural pop.		−0.647		0.078	0.073
		(0.430)		(0.218)	(0.432)
Coal shock × Black pop.		0.304		−0.401	−0.829
		(1.049)		(0.550)	(1.195)
N	7736	5494	67320	54009	6144
Adj. R-squared	0.135	0.175	0.121	0.169	0.154

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$

Table E4: Full regression table, corresponding to Figure 6 in the main text. **NB:** One challenge with these analyses is that the sample size for the Central Appalachian and Western coal areas is substantially smaller than that for other areas. To evaluate whether the null is due to a lack of statistical power, I draw a random sample of the non-Appalachian/Western observations to approximate the sample size of the Appalachian/Western tests (Model 5). Significant results are still found for non-Appalachian/Western areas with this deflated sample, indicating that the null for Central Appalachia/West is not simply due to a lack of statistical power.

	Naturalized			
	More coethnic coworkers		Fewer coethnic coworkers	
	Model 1	**Model 2**	**Model 3**	**Model 4**
Group concentration	-1.451^{***}	-1.220^{***}	-0.277^{**}	-0.209
	(0.290)	(0.189)	(0.109)	(0.139)
Coal shock intensity	-2.459^{***}	0.835	-0.578^{***}	1.495
	(0.482)	(3.543)	(0.199)	(2.860)
General shock intensity (Bartik)	-15.967^{***}	5.499	-5.568^{**}	-10.378^{*}
	(4.757)	(4.828)	(2.530)	(5.786)
Year of emigration		0.697		-0.264
		(0.431)		(0.371)
Estimated income (ln)		1.918^{***}		0.431
		(0.405)		(0.344)
Living with spouse		-0.041		0.275
		(0.279)		(0.304)
Coethnics as % county		-1.000		-0.825
		(1.019)		(0.911)
County coal reliance		-0.008^{***}		-0.007^{***}
		(0.002)		(0.001)
Rural population as % county		0.372^{***}		0.165^{***}
		(0.064)		(0.050)
Black population as % county		0.045^{**}		0.031
		(0.019)		(0.037)
Coal shock × concentration	1.292^{***}	1.102^{***}	0.193^{*}	0.162
	(0.272)	(0.194)	(0.110)	(0.152)
Coal shock × Bartik	22.200^{***}	-0.945	6.191^{***}	7.748
	(4.605)	(5.165)	(1.983)	(5.068)
Coal shock × immig. year		-0.684^{*}		-0.239
		(0.412)		(0.397)
Coal shock × income		-1.250^{***}		0.017
		(0.472)		(0.428)
Coal shock × spouse		0.218		-0.478
		(0.271)		(0.298)
Coal shock × coethnic pop.		0.026		-0.708
		(0.905)		(0.613)
Coal shock × coal reliance		0.0002		-0.001
		(0.002)		(0.002)
Coal shock × rural pop.		-0.170^{**}		-0.006
		(0.068)		(0.057)
Coal shock × Black pop.		0.009		0.015
		(0.027)		(0.044)
N	25260	21840	25292	13663
Adj. R-squared	0.169	0.215	0.079	0.131

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$

Table E5: Full regression table, corresponding to Figure 7 in the main text.

F. Social Assimilation

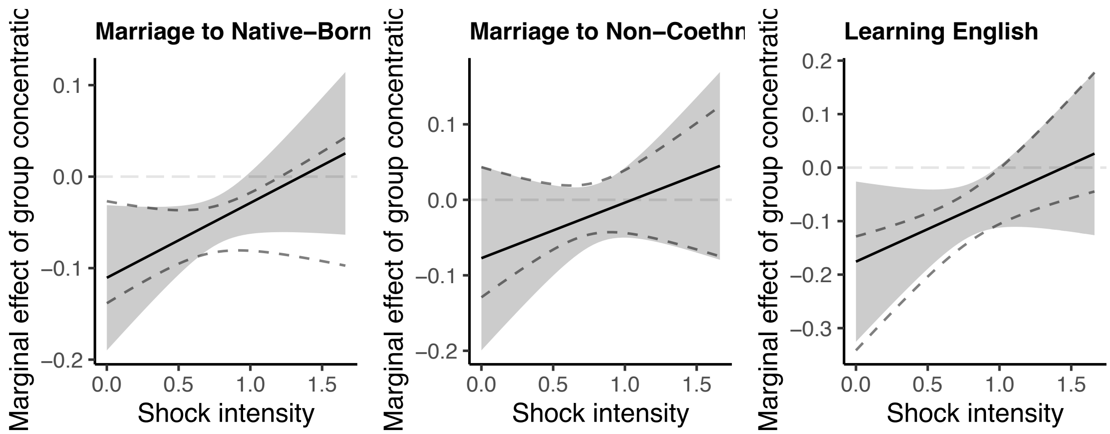

Figure F1: Interaction plots for marriage to native-born white citizen (left), marriage to a non-coethnic immigrant (center), and learning English (right) outcomes. Note that the income covariate is excluded from these tests due to limited availability for immigrant women.

G. Additional Tests

G.1. Long-Term vs. Recent Immigrants

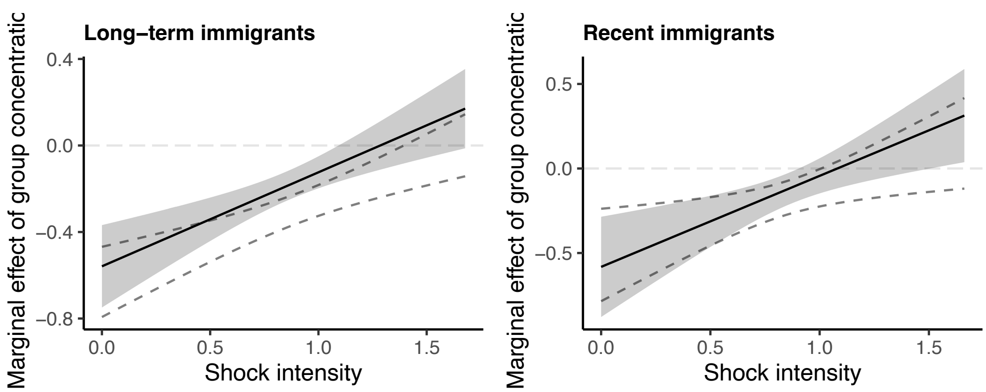

Figure G1: Interaction of group concentration and shock intensity on naturalization. Recent immigrants first arrived in the U.S. within five years of census enumeration; long-term immigrants arrived earlier.

G.2. Entry of Coal-Adjacent Immigrants into Coal

In these tests, I consider the possibility of coal-adjacent immigrants entering the coal industry during a given decade. I first limit the sample of coal-adjacent immigrants to those of at least 41 years of age, who would have been older than 75% of all coal miners upon entering the industry. I then limit the sample to coal-adjacent immigrants still not working in coal at the end of a decade.

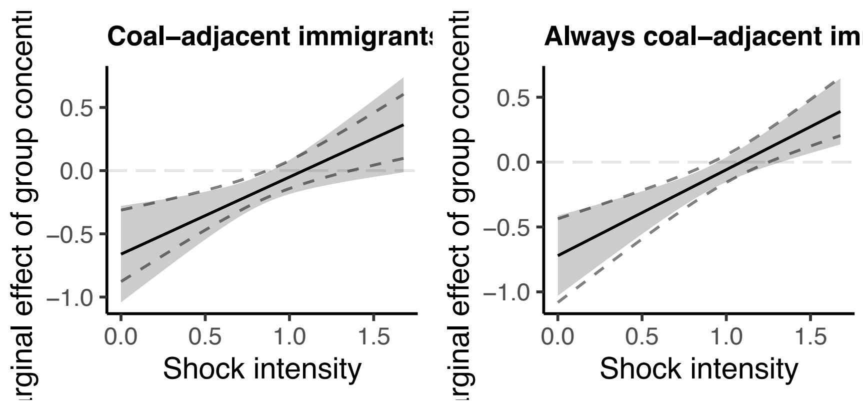

Figure G2: *Left*- Interaction of group concentration and shock intensity on naturalization, limited to coal-adjacent immigrants of at least 41 years of age (n = 14,950). *Right*- Interaction of group concentration and shock intensity on naturalization, limited to coal-adjacent immigrants still not working in coal at end of decade (n = 45,731).

G.3. Leading Economic Conditions (Placebo Test)

	Naturalized
Group concentration	-0.819^{***}
	(0.151)
Coal shock intensity	-0.734^{***}
	(0.189)
General shock intensity (Bartik)	-2.180
	(2.388)
Coal shock intensity (LEADING)	0.197^{*}
	(0.115)
General shock intensity (LEADING)	-0.281
	(1.122)
Coal shock $\times$ concentration	0.429^{***}
	(0.112)
Coal shock $\times$ Bartik	8.088^{***}
	(1.778)
Concentration $\times$ coal shock (LEADING)	0.145
	(0.104)
Coal shock (LEADING) $\times$ Bartik (LEADING)	-1.099^{*}
	(0.662)
N	72676
Adj. R-squared	0.124

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$

Table G1: Regressions on naturalization. Includes standard shock variables, alongside *leading* shock indicators (shock in the forthcoming decade; e.g., shocks between 1920–1929 for immigrants matched from 1910–1919). Null results are expected and found for these leading indicators in interaction with group concentration. Significant results remain for the standard shock variables.

G.4. Growth Measures

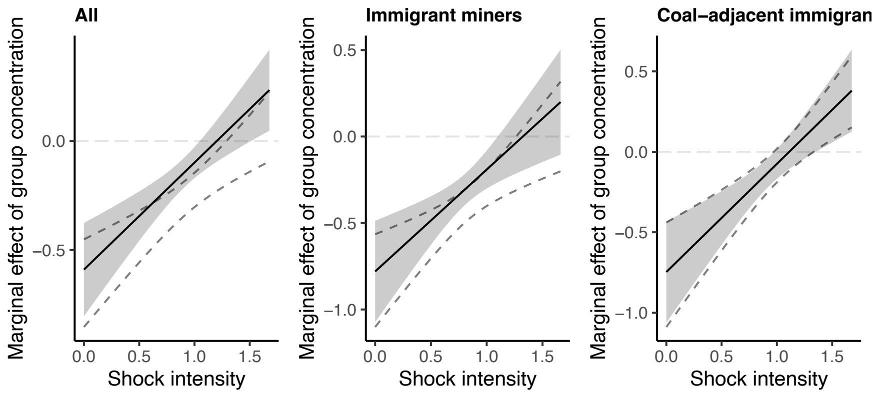

Figure G3: Baseline models, including additional controls for (a) coal and Bartik growth (sum of year-over-year percentage increases) and (b) the sum of absolute percentage changes in coal production and Bartik-estimated production.

G.5. Rail-Connected Counties

For this test, group concentration and shock intensity (both coal and general) are calculated by aggregating all counties to which a given immigrant's county had a direct railroad connection (excluding that origin county). I gathered historical railroad shapefiles from Atack, 2016, "Historical Geographic Information Systems (GIS) database of U.S. Railroads for 1899 and 1909." I isolated all railroads in operation by 1899 or 1909 and identified those that traversed coal-producing counties in 1900 or 1910 respectively. I approximate the most direct rail connections by identifying the coal-producing counties that were on the same rail line (same railroad name in the Atack data).

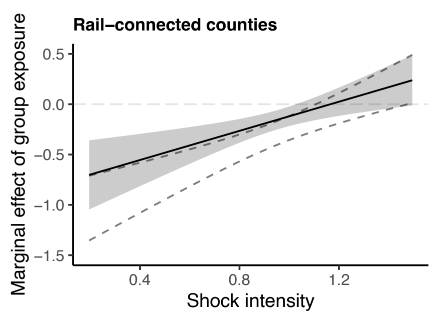

Figure G4: Interaction of rail-connected group concentration and shock intensity on naturalization.

G.6. Declarations of Intention to Nature

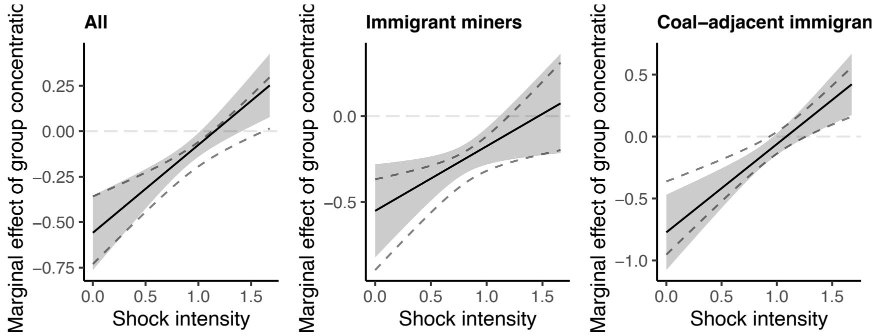

Figure G5: Interaction of group concentration and shock intensity on declarations of intention to naturalize. This is a less precise indicator of political assimilation as many immigrants who declared an intent to naturalize ultimately did not do so.

G.7. Clustering Standard Errors at County Level

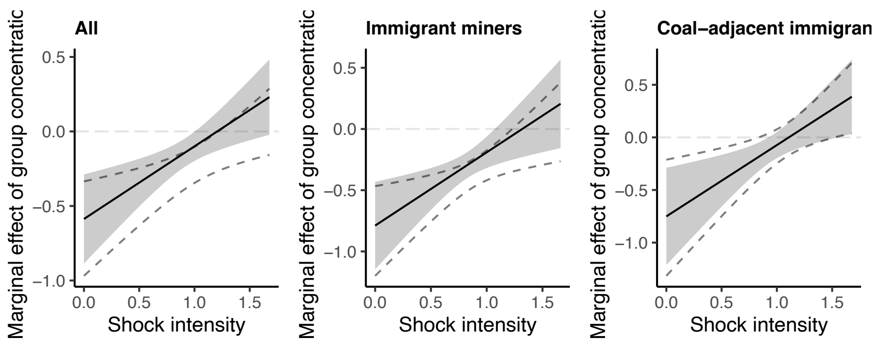

Figure G6: Interaction of group concentration and shock intensity on naturalization. Standard errors clustered by county.

G.8. Binned Estimator Results for Main Tests

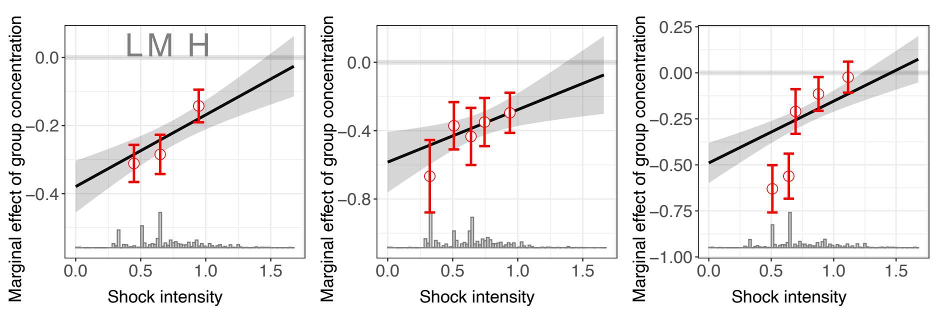

Figure G7: *L*- baseline; *C*- immigrant miners; *R*- coal-adjacent. By calculating coefficients by quantile (tercile for baseline tests; quintile for subsamples), this estimator avoids issues of linear extrapolation to parts of the domain without common support in the moderator (Hainmueller, Mummolo, and Xu 2019). Quantile estimates affirm main returns, mirroring pattern found with traditional linear estimator. Differences between highest/lowest quantile coefficient are statistically significant at $\alpha = 0.05$ level for each test.

G.9. Iteratively Dropping Sets of Observations

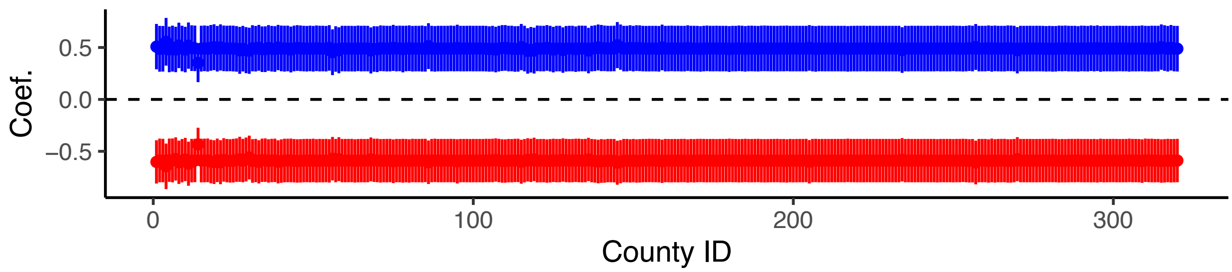

Figure G8: Replications of main model of interaction of group concentration and shock intensity on naturalization, including all covariates. Each replication drops a **single county** from the dataset to check for the presence of highly influential outlier counties. Coefficients and 95% confidence intervals are plotted for the interaction of group concentration and shock intensity (top) and group concentration main effect (bottom).

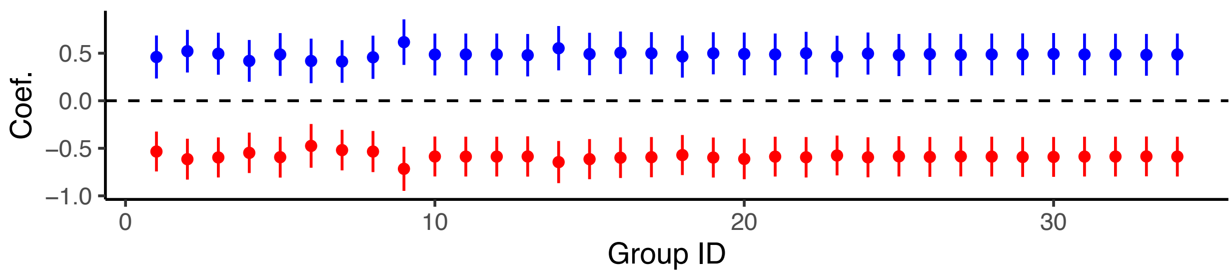

Figure G9: Replications of main model of interaction of group concentration and shock intensity on naturalization, including all covariates. Each replication drops a **single ethnic group** from the dataset to check for the presence of highly influential outlier groups. Coefficients and 95% confidence intervals are plotted for the interaction of group concentration and shock intensity (top) and group concentration main effect (bottom).

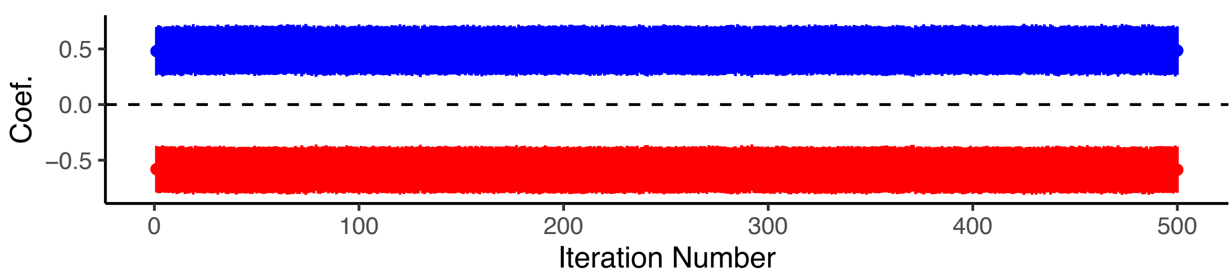

Figure G10: Replications of main model of interaction of group concentration and shock intensity on naturalization, including all covariates. Each replication randomly drops **1% of all observations** to check for highly influential outliers (similar to the suggestion of Broderick, Giordano, and Meager, "An Automatic Finite-Sample Robustness Metric," <arxiv.org/abs/2011.14999>). Coefficients and 95% confidence intervals are plotted for the interaction of group concentration and shock intensity (top) and group concentration main effect (bottom).

G.10. County-Year Fixed Effects

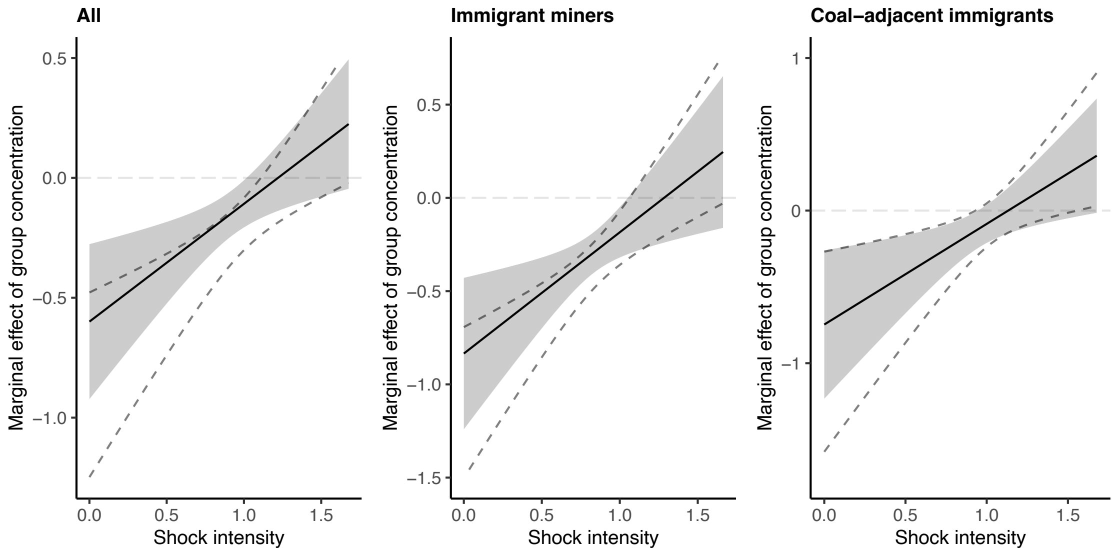

Figure G11: *L*- baseline; *C*- immigrant miners; *R*- coal-adjacent. Main models with county-census fixed effects instead of county and state-year fixed effects.

G.11. Excluding Major Strike Years

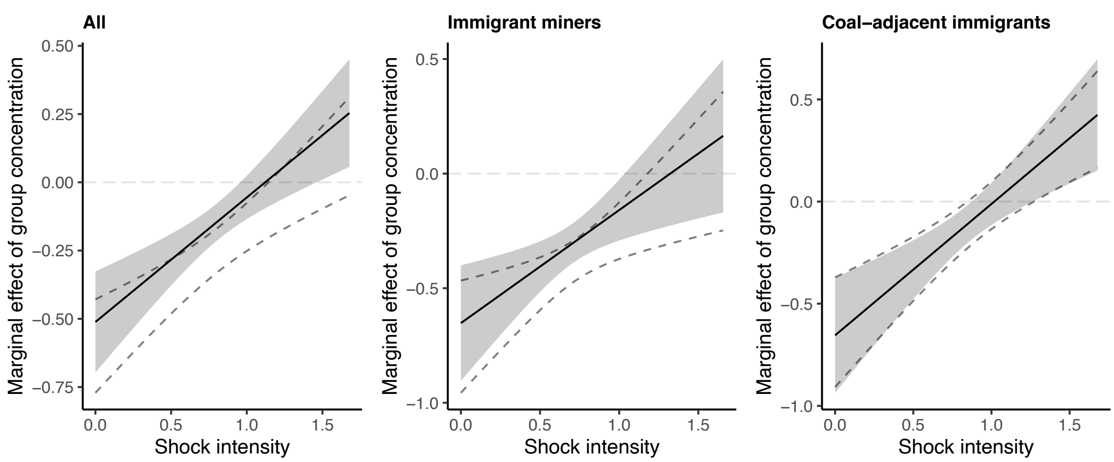

Figure G12: *L*- baseline; *C*- immigrant miners; *R*- coal-adjacent. County-years with major strike activity, as listed in U.S. Geological Survey *Mineral Resources of the United States* reports, are excluded from calculation of coal shock intensity. Excluded county-years: all United States (1919); all Pennsylvania (1902); Westmoreland County, PA (1910–11); Kanawha County, WV (1912–13); Las Animas County, CO (1914).

G.12. Excluding States with Non-Citizen Suffrage

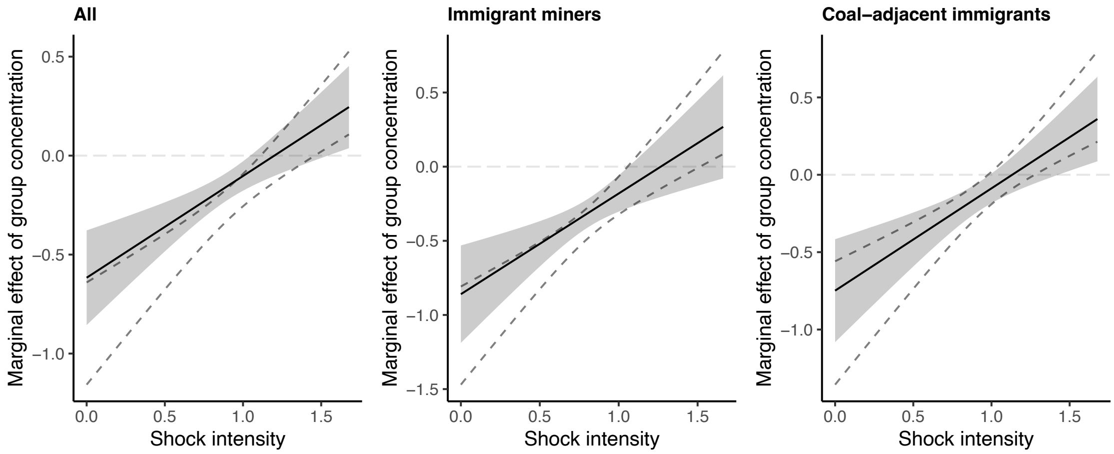

Figure G13: *L*- baseline; *C*- immigrant miners; *R*- coal-adjacent. Models exclude immigrants in states that granted “alien suffrage” for the majority of a given decade, per Kleppner 1987 (p. 166): Indiana (1900s, 1910s), Kansas (1900s, 1910s), Missouri (1900s, 1910s), Nebraska (1900s, 1910s), Oregon (1900s), South Dakota (1900s, 1910s), Wisconsin (1900s). Remaining specifications are the same as those in the main models.

G.13. Including Only States with Non-Citizen Suffrage

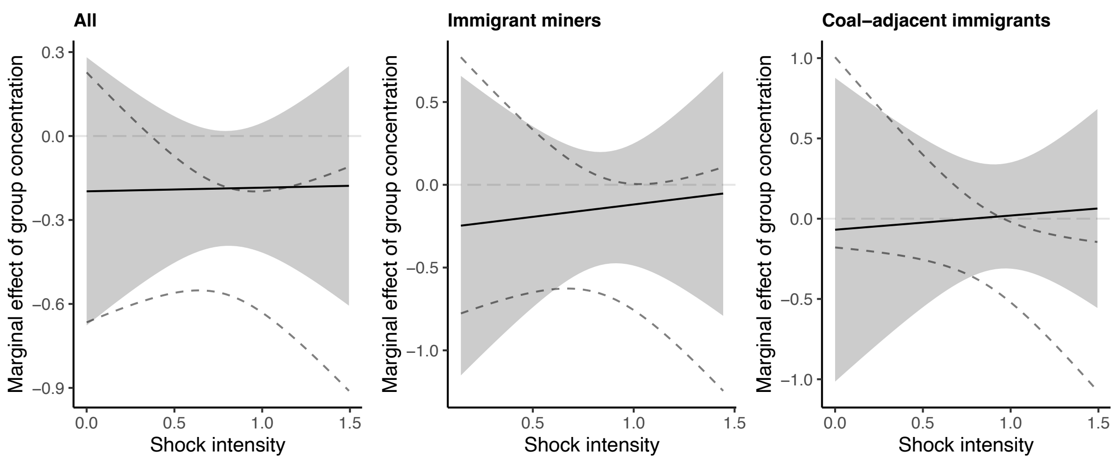

Figure G14: *L*- baseline; *C*- immigrant miners; *R*- coal-adjacent. Models include *only* immigrants in states that granted “alien suffrage” for the majority of a given decade, per Kleppner 1987 (p. 166): Indiana (1900s, 1910s), Kansas (1900s, 1910s), Missouri (1900s, 1910s), Nebraska (1900s, 1910s), Oregon (1900s), South Dakota (1900s, 1910s), Wisconsin (1900s). Remaining specifications are the same as those in the main models. **Interpretation:** These null results are anticipated; in these states, immigrants would have been able to access political rents without naturalizing, given their ability to vote before becoming citizens. Though note that the sample sizes for these tests are limited (*L*- 2294 observations in model without covariates; *C*- 771; *R*- 1492).

G.14. Alternative Regression Models

	Naturalized
Group concentration	-0.323^{***}
	(0.045)
N	75056
R-squared	0.124
Adj. R-squared	0.120

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$

Table G2: Bivariate model, regressing naturalization on group concentration. Remaining specifications are the same as in the main models.

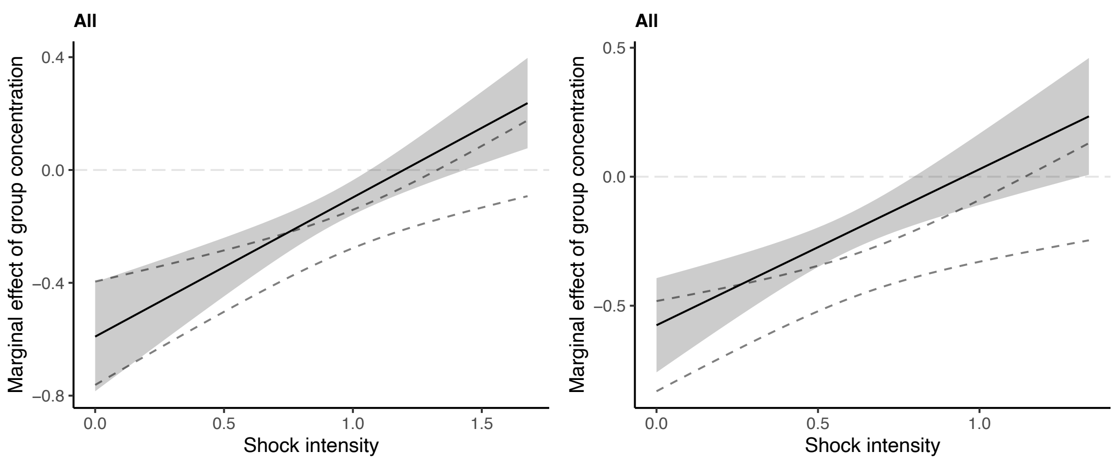

Figure G15: Replication of primary model, encompassing both immigrant miners and coal-adjacent immigrants. Plot on the *left* excludes the observation weights. Plot on the *right* converts all square root transformations to logarithmic transformations.