Econometrica Supplementary Material

SUPPLEMENT TO "FRONTIER CULTURE: THE ROOTS AND PERSISTENCE OF "RUGGED INDIVIDUALISM" IN THE UNITED STATES"

(*Econometrica*, Vol. 88, No. 6, November 2020, 2329–2368)

SAMUEL BAZZI
Department of Economics, Boston University

MARTIN FISZBEIN
Department of Economics, Boston University

MESAY GEBRESILASSE
Department of Economics, Amherst College

APPENDIX

THIS APPENDIX INCLUDES the following material: Appendix A details the procedure for mapping the frontier; Appendix B presents robustness checks for Sections 3 and 4; Appendix C presents additional results on partisanship; Appendix D describes our instrumental variables results; and Appendix E provides robustness checks for Section 5.2. Additional supplemental material can be found in the replication files and on the authors' websites: Appendix F presents alternative approaches to inference; Appendix G provides further results characterizing historical frontier demographics and institutions; Appendix H presents a case study to illustrate the long-run effects; Appendix I presents an alternative estimate of selective migration; Appendix J presents additional results; and Appendix K describes data sources and construction.

APPENDIX A: MAPPING THE FRONTIER

This section provides a step-by-step description of how we construct the frontier lines for each year between 1790 and 1890.

Step 1. Calculate county level population density per square mile for each year in 1790–1890 using the 2010 county boundaries. First, we harmonize the county-level population data from each year to the 2010 county boundaries using the procedure discussed in Section 2. For intercensal years, we interpolate county-level population by assuming a constant annual population growth rate that matches the decadal growth rate (replacing initial zeros with 0.01 to avoid infinite growth rates). Then, using the 2010 county boundaries shape file, we calculate the county-level population density as the ratio of population over county area in square miles.

Samuel Bazzi: sbazzi@bu.edu
Martin Fiszbein: fiszbein@bu.edu
Mesay Gebresilasse: mgebresilasse@amherst.edu

 https://doi.org/10.3982/ECTA16484

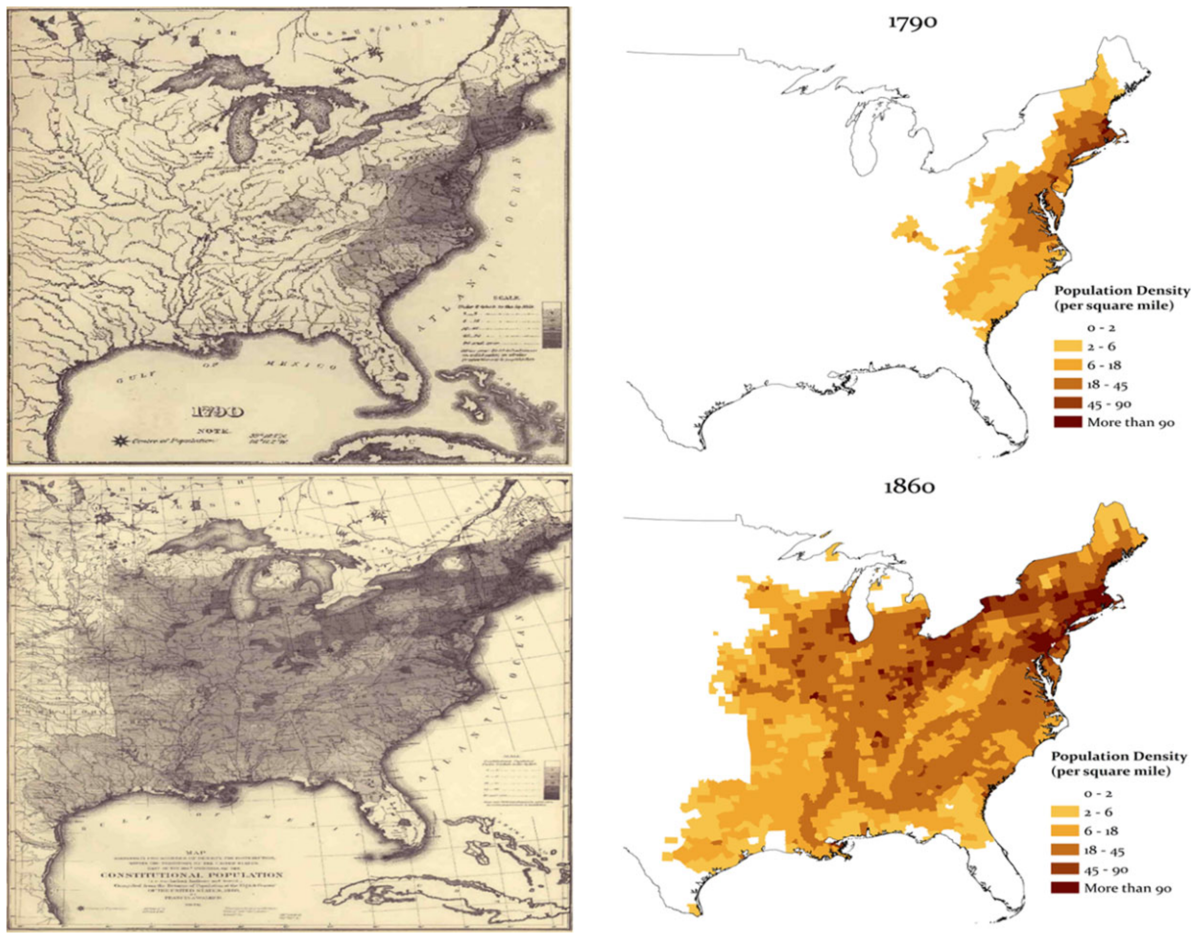

FIGURE A.1.—Population density maps from the 1890 Census report and our maps (1790 and 1860). This figure compares the maps of population density in 1790 and 1860 from the *Progress of the Nation* Census report (on the left) with the maps we constructed for 2010 county boundaries using the procedure described in Section 2 (on the right).

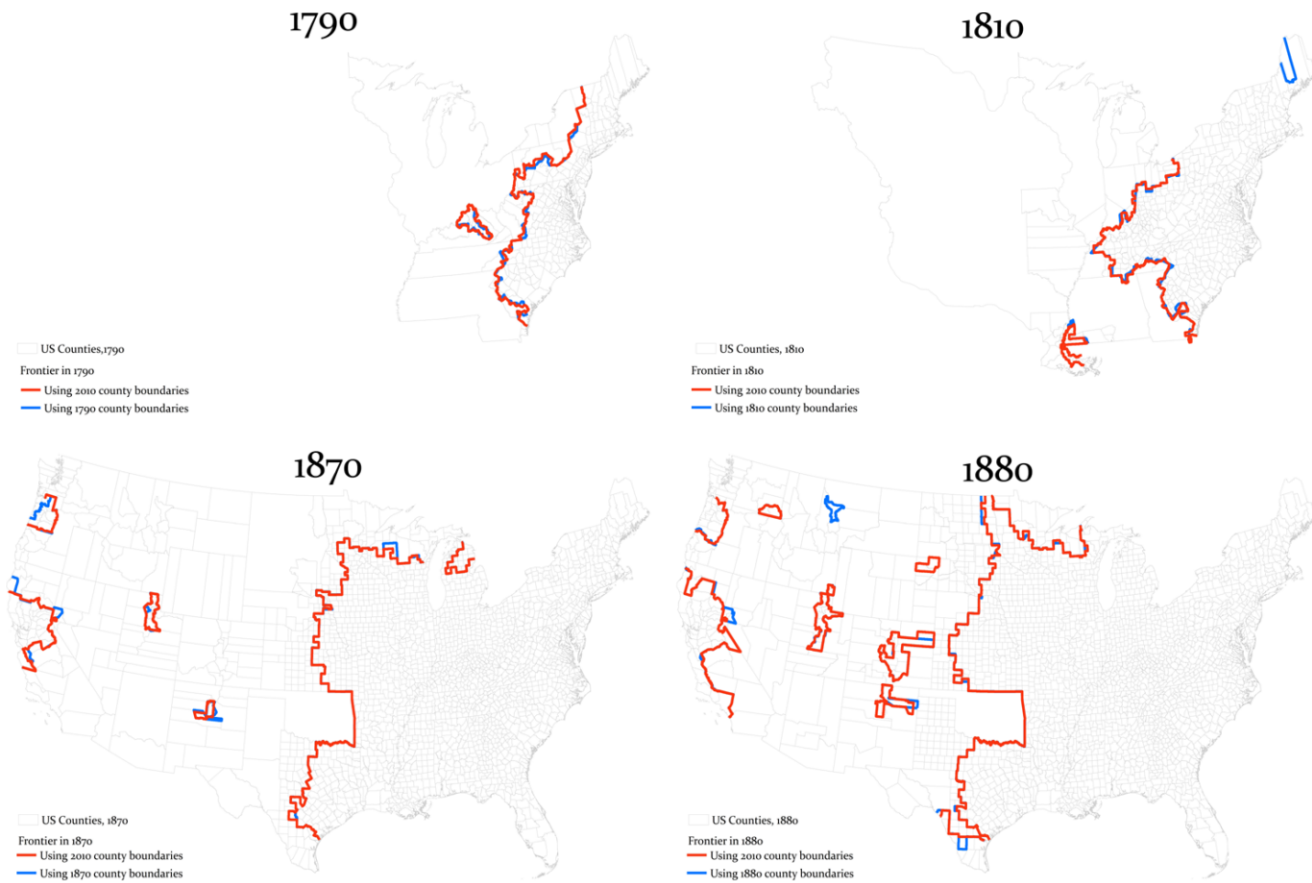

FIGURE A.2.—Frontier lines using contemporaneous versus 2010 county boundaries for selected years. Based on county level Population Census data from 1790 to 1880 and NHGIS county shape files. The figures provide the county boundaries for selected years and the frontier lines for the corresponding years drawn using the contemporaneous county boundaries as well as the 2010 county boundaries. The frontier lines delineate the counties that had population density of two persons or higher. The frontier lines in blue are drawn using the contemporaneous county boundaries, whereas the frontier lines in red are drawn using the 2010 county boundaries (after the data harmonization discussed in Section 2.1).

Step 2. Draw a contour line at population density equal to 2 people per square mile for each year. We use ArcGIS and the 2010 county boundaries. First, for each year, we convert the polygon containing the county level population density data into a raster file using the PolygonToRaster tool and set population density for the given year as the "value field" for the conversion. Then, using the ContourList tool, select the raster file created in the preceding step as an input and set the "contour value" to 2 to create contour lines at population density equal to 2. The resulting lines delineate the counties that have a population density below 2 people per square mile from those counties that have a population density above 2.

Step 3. Clean the contour lines to retain only the significant frontier lines. With the purpose of capturing historical notions of the frontier as "margins of civilization," we discard all contour line segments less than 500 km and also discard isolated pockets of relatively sparse populations within the main area of settled territory. These isolated pockets are the "inner islands" formed by counties with population density below 2 people per square mile surrounded by counties with population density above 2 people per square mile. A second set of frontier lines emerge in the West Coast in mid-19th century. This process of settlement was marked by the Gold Rush and historical forces different than those driving the main east-to-west expansion, so for our baseline analysis we focus on the territory spanned by east-to-west expansion. We do this by keeping only those frontier lines that are east of the westernmost east-to-west frontier line in 1890. In the robustness analysis, we add the West Coast to our baseline sample.

We select line segments based on length and location (e.g., X centroid of the line midpoint) in ArcGIS using the SelectLayerByAttribute tool, and apply CopyFeatures to keep only the selected lines. In the detailed robustness checks in Section 4.4, we also consider various alternatives to the frontier definition such as changing the line cutoffs, restricting to single westernmost frontier line, including the inner island lines, and considering the frontier lines that emerge from the West Coast.

APPENDIX B: FURTHER ROBUSTNESS CHECKS

B.1. *Robustness of the Historical Frontier Differential in Individualistic Names*

TABLE B.I

ROBUSTNESS CHECK ON HISTORICAL NAMES MEASURE IN TABLE I[a]

	(1)	(2)	(3)	(4)
	Raw Reported Name		Metaphone-Adjusted Name	
	Baseline OLS	Foreign-Born NN Matching	Baseline OLS	Foreign-Born NN Matching
Frontier County	0.022	0.014	0.018	0.011
	(0.007)	(0.006)	(0.006)	(0.005)
Number of County–Years	6907	6905	6907	6905
Mean Dep. Var. in Non-Frontier Counties	0.631	0.631	0.602	0.602

[a]Columns 1 and 3 report the baseline estimates from Table I. Columns 2 and 4 report, for these same outcomes, nearest-neighbor (NN) matching estimates of the frontier differential in individualistic names. We match on the county with the most similar foreign-born population share in the given Census year. These estimates are based on the single nearest neighbor. Standard errors in odd columns are clustered using the grid cell approach of Bester, Conley, and Hansen (2011) as described in Section 4.1, and in even columns are bias-adjusted and robust following best practice in the matching literature.

TABLE B.II

TOTAL FRONTIER EXPERIENCE AND CONTEMPORARY COOPERATION VERSUS SELF-RELIANCE[a]

	(1)	(2)	(3)	(4)
Total Frontier Experience	−0.019	−0.025	−0.041	−0.026
	(0.009)	(0.009)	(0.014)	(0.012)
Oster δ for $\beta = 0$	−2.77	−2.61	−15.37	−249.36
Number of Individuals	567	567	567	567
Number of Counties	48	48	48	48
Mean of Dependent Variable	0.549	0.549	0.549	0.549
R^2	0.01	0.02	0.02	0.03
Individual Demographic Controls	✓	✓	✓	✓
Division Fixed Effects		✓	✓	✓
State Fixed Effects			✓	
Geographic/Agroclimatic Controls				✓

[a]This table reports estimates for a dependent variable based on a proxy for individualism in the 1990 round of ANES, covering 567 individuals in 48 counties across 17 states in our sample. The measure asks individuals whether (i) "it is more important to be a cooperative person who works well with others" or (ii) "it is more important to be a self-reliant person able to take care of oneself." The dependent variable equals 1 if they answer (i). We report the same set of specifications in columns 1–4 as in Table II to demonstrate the statistically and economically significant effect of sizes despite the coverage limitations. Standard errors are clustered based on the grid-cell approach of Bester, Conley, and Hansen (2011) as detailed in Section 4.1. The Oster (2019) tests are with reference to a baseline specification with no controls.

B.2. *Alternative Survey-Based Proxy for Contemporary Individualism*

Beyond infrequent names, we draw upon a well suited measure from the ANES data to provide further evidence of the link between TFE and high levels of individualism. Specifically, we use the 1990 ANES round in which respondents were asked whether (i) "it is more important to be a cooperative person who works well with others" or (ii) "it is more important to be a self-reliant person able to take care of oneself." While this question was designed explicitly for studies of American individualism (see Markus (2001)), unfortunately, it was only asked in a single round.

Table B.II provides evidence that self-reliant preferences are stronger today in counties with longer exposure to the frontier historically. Around 55 percent of individuals respond in support of the cooperative answer. However, across different specifications, each decade of additional TFE is associated with around 2–6 percentage points lower support for cooperation over self-reliance. While the results with the full set of controls are noisy, we nevertheless view these findings as at least suggestive of longstanding claims about the rugged individualism pervasive on the frontier. In linking to results elsewhere in the paper, it is worth noting that individuals who identify as Republican in the ANES data are around 15–20 percent more likely to believe that it is better to be a self-reliant than a cooperative person.

Given the small number of counties, we retain this outcome in the appendix rather than in the main tables. Nevertheless, it is reassuring that the results align with our findings for other outcomes with more systematic coverage.

B.3. *Robustness of the Long-Run Effects of TFE on Individualistic Names*

We present below several tables with results discussed in the paper. Tables B.III and B.IV demonstrate robustness of the infrequent names measure to alternative specifications of the reference group (national, division, state, county), reported name versus phonetic sound (metaphone), and top 10 versus top 100 in terms of defining infrequency.

TABLE B.III

TFE AND ALTERNATIVE MEASURES OF INDIVIDUALISTIC NAMES: COUNTY-LEVEL[a]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Raw Reported Name					Metaphone-Adjusted Name					Nonpatronymic/ -Matronymic	
	Top 10	Top 10	Top 10	Top 10	Top 100	Top 10	Top 10	Top 10	Top 10	Top 100		
	National	Division	State	County	Division	National	Division	State	County	Division	All	First Born
Panel (a): Baseline Specification												
Total Frontier Experience	0.144	0.141	0.149	0.163	0.142	0.150	0.141	0.148	0.159	0.132	0.224	0.241
	(0.021)	(0.021)	(0.022)	(0.021)	(0.021)	(0.022)	(0.021)	(0.024)	(0.024)	(0.022)	(0.029)	(0.030)
Oster δ for $\beta = 0$	−7.93	−13.80	80.78	12.99	−34.76	−8.57	−16.20	98.75	19.66	−30.35	−6.82	−5.63
Number of Counties	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036
R^2	0.60	0.60	0.56	0.55	0.70	0.55	0.58	0.49	0.49	0.69	0.37	0.31
Panel (b): Nearest-Neighbor Matching on Population Density in 1940												
Total Frontier Experience	0.102	0.096	0.097	0.134	0.097	0.102	0.088	0.093	0.119	0.089	0.181	0.183
	(0.022)	(0.022)	(0.024)	(0.024)	(0.023)	(0.022)	(0.022)	(0.025)	(0.025)	(0.023)	(0.027)	(0.028)
Oster δ for $\beta = 0$	3.22	2.63	2.58	5.33	1.97	3.08	2.32	2.93	4.41	2.20	11.63	8.98
Number of Counties	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036
R^2	0.85	0.85	0.83	0.82	0.87	0.84	0.85	0.81	0.80	0.87	0.80	0.79
Panel (c): Nearest-Neighbor Matching on Foreign-Born Share in 1940												
Total Frontier Experience	0.090	0.086	0.098	0.114	0.107	0.095	0.089	0.103	0.119	0.106	0.126	0.139
	(0.018)	(0.018)	(0.019)	(0.020)	(0.017)	(0.020)	(0.019)	(0.022)	(0.022)	(0.017)	(0.025)	(0.027)
Oster δ for $\beta = 0$	2.16	1.96	2.73	2.85	3.02	2.52	2.41	4.02	4.50	4.81	2.96	3.45
Number of Counties	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036
R^2	0.86	0.87	0.85	0.84	0.90	0.84	0.85	0.81	0.81	0.90	0.81	0.80

[a]This table reports estimates in panel (a) based on the column 2 specification in Table II, in panel (b) based on the column 3 specification in Table II, and in panel (c) based on the column 4 specification in Table II. The outcomes from Table II are in columns 2 and 7. The other columns are based on alternative specifications of the dependent variable as listed at the top of the table. All other specification details are as in Table II. Standard errors are clustered based on the grid-cell approach of Bester, Conley, and Hansen (2011) as detailed in Section 4.1. Panels (b) and (c) additionally two-way cluster on the nearest-neighbor pair.

TABLE B.IV

TFE AND ALTERNATIVE MEASURES OF INDIVIDUALISTIC NAMES: INDIVIDUAL LEVEL[a]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Raw Reported Name					Metaphone-Adjusted Name					Nonpatronymic/ -Matronymic	
	Top 10	Top 10	Top 10	Top 10	Top 100	Top 10	Top 10	Top 10	Top 10	Top 100		
	National	Division	State	County	Division	National	Division	State	County	Division	All	First Born
Panel (a): FE: Age, State, Birth Order, Gender												
Total Frontier Experience	0.015	0.014	0.013	0.011	0.018	0.014	0.013	0.011	0.010	0.013	0.011	0.018
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Number of Individuals	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	9,986,891	3,134,130
Dep. Var. Mean	0.735	0.732	0.723	0.710	0.353	0.702	0.693	0.679	0.665	0.227	0.925	0.888
R^2	0.03	0.03	0.03	0.02	0.04	0.02	0.02	0.02	0.01	0.03	0.06	0.09
Panel (b): FE: Panel (a) + Neighbor Population Density Pair												
Total Frontier Experience	0.005	0.004	0.004	0.004	0.007	0.005	0.003	0.003	0.003	0.005	0.004	0.007
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Number of Individuals	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	9,986,891	3,134,130
Dep. Var. Mean	0.735	0.732	0.723	0.710	0.353	0.702	0.693	0.679	0.665	0.227	0.925	0.888
R^2	0.04	0.03	0.03	0.03	0.05	0.02	0.02	0.02	0.02	0.03	0.07	0.10
Panel (c): FE: Panel (a) + Neighbor Foreign-Born Share Pair												
Total Frontier Experience	0.005	0.005	0.005	0.004	0.009	0.005	0.005	0.004	0.004	0.007	0.003	0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Number of Individuals	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	10,036,304	9,986,891	3,134,130
Dep. Var. Mean	0.735	0.732	0.723	0.710	0.353	0.702	0.693	0.679	0.665	0.227	0.925	0.888
R^2	0.04	0.03	0.03	0.03	0.05	0.02	0.02	0.02	0.02	0.03	0.07	0.10
Panel (d): FE: Panel (a) + Last Name												
Total Frontier Experience	0.014	0.013	0.012	0.011	0.017	0.014	0.012	0.010	0.009	0.012	0.010	0.016
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Number of Individuals	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,803,398	2,891,127
Dep. Var. Mean	0.735	0.732	0.724	0.710	0.353	0.703	0.694	0.680	0.665	0.228	0.926	0.889
R^2	0.07	0.07	0.07	0.06	0.08	0.06	0.06	0.06	0.05	0.07	0.10	0.14

(*Continues*)

TABLE B.IV—*Continued*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Raw Reported Name					Metaphone-Adjusted Name					Nonpatronymic/ -Matronymic	
	Top 10	Top 10	Top 10	Top 10	Top 100	Top 10	Top 10	Top 10	Top 10	Top 100		
	National	Division	State	County	Division	National	Division	State	County	Division	All	First Born
	Panel (e): FE: Panel (b) + Last Name											
Total Frontier Experience	0.005	0.005	0.004	0.004	0.006	0.005	0.003	0.003	0.003	0.004	0.004	0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Number of Individuals	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,850,916	9,803,398	2,891,127
Dep. Var. Mean	0.735	0.732	0.724	0.710	0.353	0.703	0.694	0.680	0.665	0.228	0.926	0.889
R^2	0.08	0.08	0.07	0.07	0.09	0.07	0.06	0.06	0.06	0.07	0.11	0.15

[a]This table reports analogous individual–child-level regressions of Table B.III. This allows for the inclusion child age, gender, and birth order fixed effects. Panels (d) and (e) additionally include fixed effects for family surname. Standard errors are clustered based on the grid-cell approach of Bester, Conley, and Hansen (2011) as detailed in Section 4.1. The sample size in column 11 is smaller than in prior columns because it excludes all boys with no father and girls with no mother present in the household.

TABLE B.V

PERSISTENCE OF THE EFFECT OF TFE ON INDIVIDUALISTIC NAMES: 1910–1940[a]

	1910 (1)	1920 (2)	1930 (3)	1940 (4)
	Panel (a): Infrequent Names			
Total Frontier Experience	0.170	0.157	0.138	0.141
	(0.025)	(0.023)	(0.023)	(0.021)
Number of Counties	2036	2036	2036	2036
R^2	0.50	0.53	0.56	0.60
	Panel (b): Infrequent Names Metaphone-Adjusted			
Total Frontier Experience	0.169	0.171	0.140	0.141
	(0.027)	(0.024)	(0.023)	(0.021)
Number of Counties	2036	2036	2036	2036
R^2	0.41	0.48	0.51	0.58

[a]This table reports analogous estimates of Table II but for each year since 1910. Standard errors are clustered based on the grid-cell approach of Bester, Conley, and Hansen (2011) as detailed in Section 4.1.

These tables also report results for the non-patronymic/-matronymic measure, including a version that is based solely on first-born children of each gender.

Table B.III reports county-level results. Panel (a) reports estimates of the specification in column 2 of Table II for the different outcomes listed at the top of each column. Panels (b) and (c) report analogous estimates for columns 3 and 4 of Table II which include fixed effects for county pairs with, respectively, the most similar population density and foreign-born population shares in 1940.

Table B.IV reports individual-level results rather than county-level mean outcomes. Panels (a)–(c) are as in Table B.III with added fixed effects for child age, birth order, and gender. Panel (d) augments the panel (a) specification with nearly 400,000 fixed effects for family surnames. Panel (e) adds those surname fixed effects to the panel (b) specification.

Finally, Table B.V shows that the baseline results for individualistic names look similar in each decade before 1940 but after the closing of the frontier.

B.4. *Robustness: Additional Controls, Regional Heterogeneity, and Extended Frontier Era*

Table B.VI reports the checks on omitted confounders discussed in Section 4.4. The full elaboration of coefficients on the control variables can be seen in Appendix Table J.3.

Table B.VII reports the regional heterogeneity and extended time-frame results described in Section 4.4. We begin by adding West Coast frontier counties to our sample. These 105 counties were settled starting in the mid-19th century and were located to the west of the major frontier line on the West Coast in 1890 (the year in which the Census declared the frontier closed). As shown in column 1, for all key outcomes, the estimated effects of TFE remain effectively unchanged.

Then we split the sample by Census region and show that the effects of TFE hold separately in the Midwest (column 2), the South (column 3), and the West (column 4). The coefficient estimates are generally smaller and noisier in the West, which can be explained in part by the small sample size (152 counties). In subsequent columns 5–8, we extend the frontier time period through 1950, incorporating in our sample counties that experienced

TABLE B.VI

ADDITIONAL CONTROLS[a]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Baseline Controlling for ...														
Ruggedness		✓												✓
Rainfall Risk			✓											✓
Distance to Nearest Portage Site				✓										✓
Distance to Nearest Mine					✓									✓
Distance to Nearest Indian Battle						✓								✓
Slave Population Share, 1860							✓							✓
Sex Ratio, 1890								✓						✓
Immigrant Share, 1890									✓					✓
Scottish and Irish Immigrant Share, 1890										✓				✓
Birthplace Diversity, 1890											✓			✓
Years Connected to Railroad by 1890												✓		✓
Manufacturing Employment Share, 1890													✓	✓
Panel (a): Infrequent Name Share in 1940 (Standardized)														
Total Frontier Experience	0.141	0.143	0.139	0.143	0.141	0.145	0.106	0.142	0.134	0.107	0.127	0.094	0.151	0.089
	(0.021)	(0.020)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)	(0.019)	(0.020)	(0.022)	(0.020)	(0.019)
Oster δ for $\beta = 0$	−13.80	−11.03	−40.85	−8.28	−14.43	−6.38	1.23	−13.21	10.07	2.03	3.86	1.16	−6.57	1.61
Number of Counties	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036
R^2	0.60	0.62	0.60	0.60	0.61	0.60	0.62	0.61	0.61	0.67	0.61	0.66	0.66	0.76
Panel (b): Infrequent Name Share in 1940, Metaphone-Adjusted (Standardized)														
Total Frontier Experience	0.141	0.143	0.139	0.143	0.141	0.144	0.107	0.141	0.135	0.106	0.128	0.094	0.150	0.091
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.021)	(0.020)	(0.020)	(0.021)	(0.022)	(0.021)	(0.019)
Oster δ for $\beta = 0$	−16.20	−12.62	−36.47	−8.50	−17.00	−8.30	1.49	−15.37	16.69	2.21	5.20	1.30	−7.25	1.93
Number of Counties	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036
R^2	0.58	0.60	0.58	0.58	0.58	0.58	0.59	0.58	0.58	0.65	0.59	0.63	0.63	0.73
Panel (c): Mean Government Preferences Outcomes (CCES), 2006–2016														
Total Frontier Experience	0.014	0.013	0.014	0.014	0.014	0.013	0.014	0.014	0.011	0.011	0.010	0.012	0.011	0.009
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)

(*Continues*)

TABLE B.VI—*Continued*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Baseline Controlling for ...														
Ruggedness		✓												✓
Rainfall Risk			✓											✓
Distance to Nearest Portage Site				✓										✓
Distance to Nearest Mine					✓									✓
Distance to Nearest Indian Battle						✓								✓
Slave Population Share, 1860							✓							✓
Sex Ratio, 1890								✓						✓
Immigrant Share, 1890									✓					✓
Scottish and Irish Immigrant Share, 1890										✓				✓
Birthplace Diversity, 1890											✓			✓
Years Connected to Railroad by 1890												✓		✓
Manufacturing Employment Share, 1890													✓	✓
Oster δ for $\beta = 0$	6.49	6.23	7.07	7.00	6.34	5.46	7.00	6.64	3.73	3.36	3.33	4.20	3.71	2.26
Number of Counties	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759	112,759
Mean of Dependent Variable	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
R^2	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	Panel (d): County Property Tax Rate in 2010													
Total Frontier Experience	−0.034	−0.034	−0.034	−0.033	−0.034	−0.033	−0.033	−0.034	−0.028	−0.026	−0.027	−0.023	−0.036	−0.023
	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Oster δ for $\beta = 0$	−27.45	−7.11	−3.23	2.54	−30.95	9.19	8.59	−30.17	0.55	0.48	0.49	0.28	−2.08	0.42
Number of Counties	2029	2029	2029	2029	2029	2029	2029	2029	2029	2029	2029	2029	2029	2029
Mean of Dependent Variable	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
R^2	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.83	0.83	0.83	0.83	0.83	0.85
	Panel (e): Republican Vote Share, Average 2000–2016													
Total Frontier Experience	2.055	2.050	2.115	2.095	2.055	2.172	1.399	2.060	1.715	1.717	1.689	1.640	2.137	0.931
	(0.349)	(0.349)	(0.338)	(0.344)	(0.350)	(0.351)	(0.361)	(0.347)	(0.328)	(0.340)	(0.327)	(0.361)	(0.350)	(0.316)

(*Continues*)

TABLE B.VI—*Continued*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Baseline Controlling for …														
Ruggedness		✓												✓
Rainfall Risk			✓											✓
Distance to Nearest Portage Site				✓										✓
Distance to Nearest Mine					✓									✓
Distance to Nearest Indian Battle						✓								✓
Slave Population Share, 1860							✓							✓
Sex Ratio, 1890								✓						✓
Immigrant Share, 1890									✓					✓
Scottish and Irish Immigrant Share, 1890										✓				✓
Birthplace Diversity, 1890											✓			✓
Years Connected to Railroad by 1890												✓		✓
Manufacturing Employment Share, 1890													✓	✓
Oster δ for $\beta = 0$	−8.55	−8.93	−6.43	−6.75	−8.56	−5.10	2.34	−8.92	7.33	7.40	6.72	4.53	−7.81	1.35
Number of Counties	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036	2036
Mean of Dependent Variable	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04	60.04
R^2	0.33	0.33	0.34	0.33	0.33	0.34	0.38	0.34	0.38	0.38	0.39	0.36	0.36	0.49
State Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

[a]This table augments the baseline specification, reproduced in column 1, with additional controls. The variables are defined in Section 4.4 and at the end of Appendix K, but we note here that the measure in column 5 is based on the known mining sites pre-1890. Standard errors are clustered based on the grid-cell approach of Bester, Conley, and Hansen (2011) as detailed in Section 4.1. The Oster (2019) tests are with reference to a baseline specification with no controls.

TABLE B.VII

WEST COAST, EXTENDED TIME FRAME: REGIONAL HETEROGENEITY[a]

Frontier Time Frame:	Baseline (1790–1890)				Extended (1790–1950)			
	Baseline + West Coast	Only Midwest	Only South	Only West	Extended Sample	Only Midwest	Only South	Only West
Regional sample restriction:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel (a): Infrequent Name Share in 1940 (Standardized)							
Total Frontier Experience	0.138	0.247	0.163	0.103	0.087	0.124	0.114	0.068
	(0.020)	(0.044)	(0.030)	(0.069)	(0.014)	(0.034)	(0.026)	(0.020)
Number of Counties	2141	987	936	152	2500	1038	1074	322
	Panel (b): Infrequent Name Share in 1940, Metaphone-Adjusted (Standardized)							
Total Frontier Experience	0.138	0.242	0.148	0.112	0.085	0.125	0.093	0.071
	(0.021)	(0.043)	(0.029)	(0.072)	(0.014)	(0.033)	(0.024)	(0.019)
Number of Counties	2141	987	936	152	2500	1038	1074	322
	Panel (c): Mean Government Preferences Outcomes (CCES), 2006–2016							
Total Frontier Experience	0.015	0.019	0.010	0.026	0.013	0.017	0.010	0.013
	(0.003)	(0.004)	(0.003)	(0.012)	(0.002)	(0.004)	(0.003)	(0.004)
Number of Individuals	140,715	49,218	52,285	32,319	158,403	49,479	55,462	46,569
Mean of Dependent Variable	0.40	0.40	0.43	0.36	0.41	0.40	0.43	0.38
	Panel (d): County Property Tax Rate in 2010							
Total Frontier Experience	−0.031	−0.051	−0.027	−0.006	−0.025	−0.042	−0.031	−0.009
	(0.006)	(0.014)	(0.007)	(0.013)	(0.004)	(0.012)	(0.006)	(0.004)
Number of Counties	2134	981	935	152	2491	1029	1074	322
Mean of Dependent Variable	1.01	1.24	0.75	0.76	0.98	1.23	0.78	0.72
	Panel (e): Average Republican Vote Share over 2000–2016							
Total Frontier Experience	2.070	1.882	2.458	1.459	1.302	1.515	1.429	1.197
	(0.332)	(0.414)	(0.396)	(0.890)	(0.256)	(0.350)	(0.422)	(0.274)
Number of Counties	2141	987	936	152	2500	1038	1074	322
Mean of Dependent Variable	59.43	59.15	61.78	48.81	60.49	59.43	63.18	56.10
State Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓	✓	✓	✓	✓

[a]Focusing on five key outcomes across panels (a)–(e), this table extends our baseline sample of counties and examines region-by-region sample splits. Column 1 adds 105 counties along the secondary West Coast frontier (see Figure 3). Column 2 restricts to counties in the Midwest Census region, column 3 restricts to the South region, and column 4 restricts to the West, which includes the 105 counties added in column 1 plus 47 others in states in the West region but falling inside the 1890 main east-to-west frontier line. Column 5 expands the column 1 sample to include counties beyond the (main and secondary) 1890 frontier lines but inside the eventual frontier line realized by 1950. Columns 6–8 then proceed with the same region-by-region sample splits. Standard errors are clustered based on the grid-cell approach of Bester, Conley, and Hansen (2011) as detailed in Section 4.1.

frontier conditions beyond 1890. Here, the effects of TFE are economically and statistically significant across all regions.

APPENDIX C: FURTHER RESULTS ON VOTING AND PARTISANSHIP

Table C.I reports estimates for the effects of TFE on partisan legislative speech associated with opposition to big government and redistribution. We compute measures of speech intensity based on the corpus of legislative speech and associated bigrams pro-

TABLE C.I

TFE AND OPPOSITION TO BIG GOVERNMENT IN REPUBLICAN LEGISLATOR SPEECH[a]

	1902–1930 (1)	1932–1960 (2)	1962–1990 (3)	1992–2016 (4)
	Dep. Var.: Share of Legislator Speech With Bigrams Including [. . .]			
	Panel (a): Big Government			
Total Frontier Experience	−0.000	−0.008	0.001	0.021
	(0.000)	(0.015)	(0.009)	(0.010)
Number of Counties	2100	1510	1447	1638
R^2	0.01	0.02	0.04	0.06
	Panel (b): Taxation			
Total Frontier Experience	−0.009	−0.003	0.079	0.122
	(0.013)	(0.055)	(0.105)	(0.136)
Number of Counties	2100	1510	1447	1638
R^2	0.04	0.04	0.12	0.10
	Panel (c): Budget			
Total Frontier Experience	−0.034	0.099	0.067	0.079
	(0.070)	(0.097)	(0.056)	(0.036)
Number of Counties	2100	1510	1447	1638
R^2	0.02	0.03	0.09	0.06

[a]This table reports estimates of equation (C.1) across four different time periods: 1902–1930 (column 1), 1932–1960 (column 2), 1962–1990 (column 3), and 1992–2016 (column 4). Other details on the specifications can be found in the discussion above.

TABLE C.II

TOTAL FRONTIER EXPERIENCE AND PREFERENCES OVER PARTISAN POLICY ISSUES[a]

Dependent variable:	Opposes Affordable Care Act (1)	Opposes Increasing Minimum Wage (2)	Opposes Banning Assault Rifles (3)	Opposes Regulation of CO_2 Emissions (4)
Total Frontier Experience	0.022	0.023	0.015	0.016
	(0.004)	(0.008)	(0.004)	(0.004)
Oster δ for $\beta = 0$	7.33	11.07	9.44	10.77
Number of Individuals	29,446	5134	29,404	29,215
Number of Counties	1728	1066	1723	1718
Mean of Dependent Variable	0.53	0.31	0.37	0.32
R^2	0.06	0.06	0.09	0.08
Survey Wave Fixed Effects	✓	✓	✓	✓
Individual Demographic Controls	✓	✓	✓	✓
Statf Fixed Effects	✓	✓	✓	✓
Geographic/Agroclimatic Controls	✓	✓	✓	✓

[a]This table reports estimates of equation (4) for four measures of support for conservative issues that are particularly relevant to the frontier setting in historical accounts. The dependent variables are all binary indicators based on questions in the CCES across different years. The measure in column 1 equals 1 if the individual in 2014 believes that the Affordable Care Act (ACA) should be repealed, in column 2 equals 1 if the individual in 2007 opposes an increase in the minimum wage, in column 3 equals 1 if the individual in 2014 opposes a ban on assault rifles, and in column 4 equals 1 if the individual in 2014 opposes regulation of pollution by the Environmental Protection Agency (EPA). The set of specifications are otherwise the same as in Table III; see the notes therein for details. Standard errors are clustered based on the grid-cell approach of Bester, Conley, and Hansen (2011) as detailed in Section 4.1. The Oster (2019) tests are with reference to a baseline specification that only includes state fixed effects.

vided in Gentzkow, Shapiro, and Taddy (2019). In particular we estimate the specification

$$\left(\frac{\text{bigram } b}{\text{all words}}\right)_{d(\ell)t} = \alpha + \beta \,\text{total frontier experience}_d + \mathbf{x}_d'\boldsymbol{\gamma} + \mathbf{FE} + \varepsilon_{d(\ell)t}, \tag{C.1}$$

where the dependent variable captures the share of bigrams related to topic b in all words used by Republican legislator ℓ from congressional district d in congress year t.[1] We include fixed effects **FE** for the Census division in which d lies as well as the congress year. Standard errors are clustered at the congressional district level.

We consider three topics b that are particularly relevant to the other anti-statist outcomes we consider in the paper. These include, across panels (a) Big Government based on the "big govern" bigram, (b) Taxation based on the top four Republican-leaning bigrams on this topic with highest average partisanship across all sessions identified by Gentzkow, Shapiro, and Taddy (2019) (tax increas, rais tax, tax relief, american taxpay), and (c) Budget based on the top four Republican-leaning bigrams on this topic with highest average partisanship across all sessions identified by Gentzkow, Shapiro, and Taddy (2019) (govern spend, feder spend, intern revenu, treasuri depart). TFE is computed at the congressional district level rather than the county level as in our core specifications in the paper.

The estimates in Table C.I suggest that TFE amplified the supply of political opposition to big government among Republican legislators, especially beginning in the mid-1990s. While partisanship around these themes and issues grew for politicians everywhere around this time, our estimates suggest that this growth may have been differential in regions with greater TFE. While some of the differences with earlier periods are noisy (e.g., in panels (b) and (c)), this supply-side pattern is consistent with the growing demand-side differential seen in the Republican presidential vote shares in Figure 5. These findings paint a consistent picture of TFE capturing latent cultural attitudes that can be activated around salient political themes.

As noted in Section 4.3, these time patterns could be due to changes in the type of elected representative or to a change in the type of speech used by representatives who would have been elected otherwise. What is important here is that the patterns line up with the strong voter demand for attention to such issues seen in Table III.

APPENDIX D: INSTRUMENTAL VARIABLES STRATEGY

This section presents a suite of results corresponding to our discussion of the IV specification in Section 4.4. Table D.I presents IV estimates for the same five primary outcomes as in Table V. In panel (a), we find large, significant effects of TFE that are slightly larger but statistically indistinguishable from the OLS estimates. Panel (b) shows similar results when using predicted rather than actual migrant flows in the IV construction. Overall, the IV exercises help clarify the identifying variation in TFE. The first-stage results in Table D.II show that there are various geoclimatic predictors of TFE, but national immigration inflows, which are unrelated to local conditions of any given county, also account for a sizable amount of variation in TFE. Coupled with the robustness checks in the previous section, the similarity of the IV and OLS results in Table D.I suggests that our findings are not driven by local conditions determining both TFE and outcomes of interest today. Below, we further develop the IV and its identifying assumptions.

[1] We multiply the dependent variable by 1000 for presentational purposes.

TABLE D.I

INSTRUMENTAL VARIABLES (IV) ESTIMATES FOR SUMMARY OUTCOMES[a]

	Infrequent Name		Mean	County	Republican
	Raw	Metaphone	Gov. Prefs.	Property	Vote Share
	Standardized		CCES	Tax Rate	Avg.
	1940	1940	2006–2016	2010	2000–2016
Dependent variable:	(1)	(2)	(3)	(4)	(5)
	Panel (a): IV = Log Average Actual National Migration Inflows Over 30 Years				
Total Frontier Experience	0.193	0.215	0.010	−0.045	3.407
	(0.039)	(0.060)	(0.004)	(0.014)	(0.585)
Number of Observations	2036	2036	112,759	2029	2036
Mean of Dependent Variable	0.00	0.00	0.41	1.02	60.04
First-Stage F Statistic	193.64	193.64	40.34	194.13	193.64
	Panel (b): IV = Log Average Predicted National Migration Inflows Over 30 Years				
Total Frontier Experience	0.223	0.232	0.007	−0.049	3.177
	(0.044)	(0.062)	(0.004)	(0.014)	(0.624)
Number of Observations	2036	2036	112,759	2029	2036
Mean of Dependent Variable	0.00	0.00	0.41	1.02	60.04
First-Stage F Statistic	195.84	195.84	44.56	196.31	195.84
State fixed effects	✓	✓	✓	✓	✓
Geographic/agroclimatic controls	✓	✓	✓	✓	✓

[a]This table reports instrumental variables estimates of equation (4) based on the instruments described in Section 4.4. We again report results for the four summary outcomes examined in prior tables, and total frontier experience is measured in decades. Panel (a) reports the IV estimates for the baseline sample and specification using the log of the average national annual actual migration inflows over the 30 years since the frontier is within 110 km from the county centroid. Panel (b) reports the estimates using the IV constructed based on annual migration inflows to the United States predicted by weather shocks in Europe. The details on the construction of both instrumental variables are presented in the Appendix D. The first-stage F statistics are cluster-robust, and standard errors are clustered based on the grid-cell approach of Bester, Conley, and Hansen (2011) as detailed in Section 4.1.

Figure D.1 shows immigration inflows to the United States over the study period.

Figure D.2(a) then shows the strong positive correlation between these inflows by decade and the speed of westward expansion, proxied by the east-to-west distance traveled by the country's population centroid (the green dot in Figure A.1(b) for 1860). Figure D.2(b) shows that the scale of native-born migration to the frontier is greater in years with more immigrants arriving to the United States. We identify migrants moving to the frontier using the same procedure based on differences in children's birthplaces as detailed in Section 5.

These scatter plots help visualize the process by which immigrants arriving in the United States pushed the edges of settlement westward, which in turn hastened the onward march of the frontier line. In periods with low immigrant inflows, this push slowed down, leading some counties to remain part of the frontier for longer than those that just happened to be getting closer to the frontier line at a time of rapid inflows into the United States. Table D.II demonstrates the strong first stage in our main IV regressions from Table D.I.

TABLE D.II

FIRST-STAGE RESULTS FOR THE INSTRUMENTAL VARIABLES ESTIMATES IN TABLE D.I[a]

	Total Frontier Experience	
Dependent Variable (in First Stage):	(1)	(2)
Log Average Actual National Migration Inflows	−1.016	
	(0.073)	
Log Average Predicted National Migration Inflows		−2.010
		(0.144)
Log County Area	0.234	0.232
	(0.072)	(0.073)
Latitude	−0.091	−0.076
	(0.079)	(0.075)
Longitude	−0.153	−0.176
	(0.029)	(0.030)
Mean Annual Temperature	−0.144	−0.102
	(0.071)	(0.068)
Mean Annual Rainfall	−0.002	−0.002
	(0.001)	(0.001)
Mean of Median Altitude	−0.001	−0.001
	(0.000)	(0.000)
Distance to Coast	−0.001	−0.001
	(0.000)	(0.000)
Distance to Rivers	0.003	0.004
	(0.001)	(0.001)
Distance to Lakes	−0.000	−0.000
	(0.000)	(0.000)
Average Agricultural Suitability	2.686	0.876
	(0.788)	(0.752)
Number of Counties	2036	2036
First-Stage F Statistic	193.64	195.84

[a]This table reports the first-stage results corresponding to the baseline IV regressions presented in Table D.I.

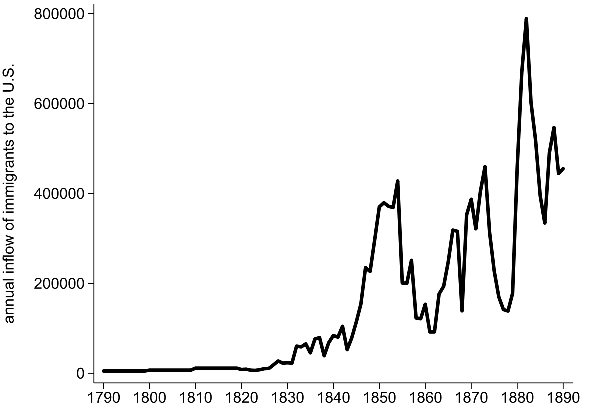

FIGURE D.1.—Annual migration inflows. This figure plots the total number of migrants entering the United States, 1790–1890. The data for 1820–1890 are available from the Migration Policy Institute (2016), while the data for 1790–1819 is imputed from Tucker (1843).

(a) Center-of-Population Shift

(b) Migration of Native-Born Children to the Frontier

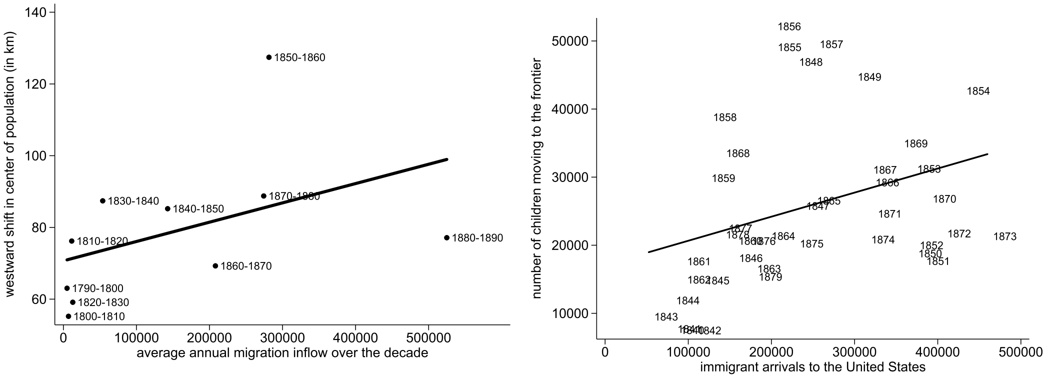

FIGURE D.2.—Immigration and westward expansion. Graph (a) plots the length of the decadal westward shift of the center of population (in km) against the average annual immigrant inflow during the decade. The center of population is the point at which weights of equal magnitude corresponding to the location of each person in an imaginary flat surface representing the United States would balance out. Graph (b) plots the relationship between the number of immigrant arrivals to the United States in a given year and the number of children brought to the frontier with their parents, a sample that we use throughout Section 5.

Section 4.4 shows that the main results hold in an instrumental variable specification exploiting time-series variation in national migration inflows. To address concerns regarding the excludability of the baseline instrument due to pull factors associated with immigrant inflows, we show in panel (b) of Table D.I that the IV results are qualitatively unchanged when using an instrument based on push factors unrelated to frontier conditions. For this version of the IV, we draw on the approach in Sequeira, Nunn, and Qian (2020), using country–year-level data on migrant inflows from 16 European countries to the United States from 1820 to 1890 and constructing predicted migration outflows induced by weather shocks. First, using country-specific regressions, we predict the annual migrant outflows from each country to the United States as a function of country-specific shocks to temperature and rainfall in the prior year (see Sequeira, Nunn, and Qian (2020), for details on these measures). Second, we aggregate across countries to obtain the total predicted migrant inflows to the United States for each year. Analogous to our baseline instrument, we then construct the IV for each county in our sample by calculating the average annual predicted migrant inflow to the United States over the 30 years starting from the first year in which the given county is just west of the frontier. Figure D.3 shows how the predicted inflows, which isolate push factors, compare to the actual inflows, which naturally include both push and pull. While the data on migrant inflows from Europe to the United States are available only starting in 1820, we retain the full sample of counties in the IV regressions by imputing the inflows for 1790–1819 using linear extrapolation of the post-1819 predicted inflows.[2]

[2]Restricting the sample to counties just west of the frontier after 1820—for which the IV is solely based on predicted flows without extrapolation—delivers similar results, though the estimates are noisier due to the smaller sample size.

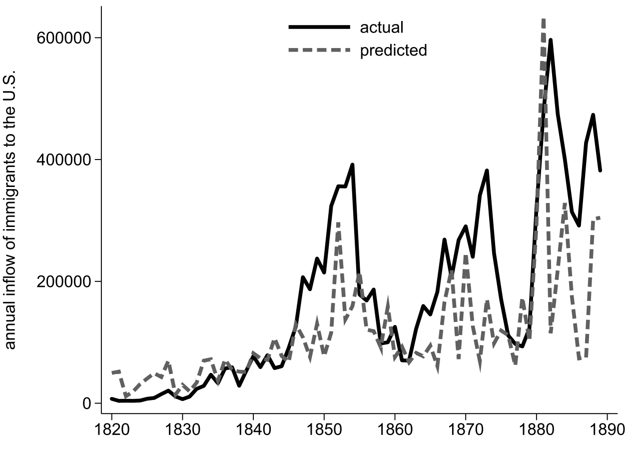

FIGURE D.3.—Actual versus predicted immigration inflows from Europe to the United States. This figure compares the actual migration inflows from Europe from 1820 to 1890 to the predicted flows based on the total country-specific predicted outflows using the climatic shocks approach in Sequeira, Nunn, and Qian (2020) as described above.

Probing Instrument Validity

While this IV strategy addresses some concerns about omitted variables, the exclusion restriction may not hold. In particular, the immigration flows underlying the IV affect both the scale and composition of migrants to the frontier.[3] During periods of greater immigration (lower predicted TFE) frontier settlers may include relatively more foreigners and non-individualistic native-born. For example, in periods with many immigrants arriving to the East coast, a large number of native-born Americans flowed westward, many of whom could have been non-individualistic types. By a similar logic, large immigration inflows out of Europe (induced by weather shocks) may lead to a greater stock of non-individualistic, foreign-born arriving on the frontier. These population flows would directly lower TFE and reduce the prevalence of individualism. While these types of selective migration could invalidate the IV, neither seems pervasive during the frontier era as we argue here.

We explore this potential bias in the latter half of the 1800s when it is possible to separately relate scale and composition of frontier migrant flows to the national immigration shocks underlying our IV. We identify migrants moving to the frontier using the same procedure based on differences in children's birthplaces as detailed in Section 5. We measure individualism based on the names of children born prior to moving.

In certain contexts, one might expect the prevalence of individualists to be inversely related to the scale of migration. However, Figure D.4(a) suggests that this is not the case when looking at native-born migration flows to the frontier. Each point on the graph reflects a given year's number of children under the age of 20 migrating to the frontier with their parents and the mean prevalence of individualistic names among those children. The lack of correlation between scale and individualism suggests that there is not a mechanical relationship between the two in our setting.

[3]In a standard Roy–Borjas model of migration, the size and composition of migration flows are in general jointly determined and not independent (see, e.g., Grogger and Hanson (2011)).

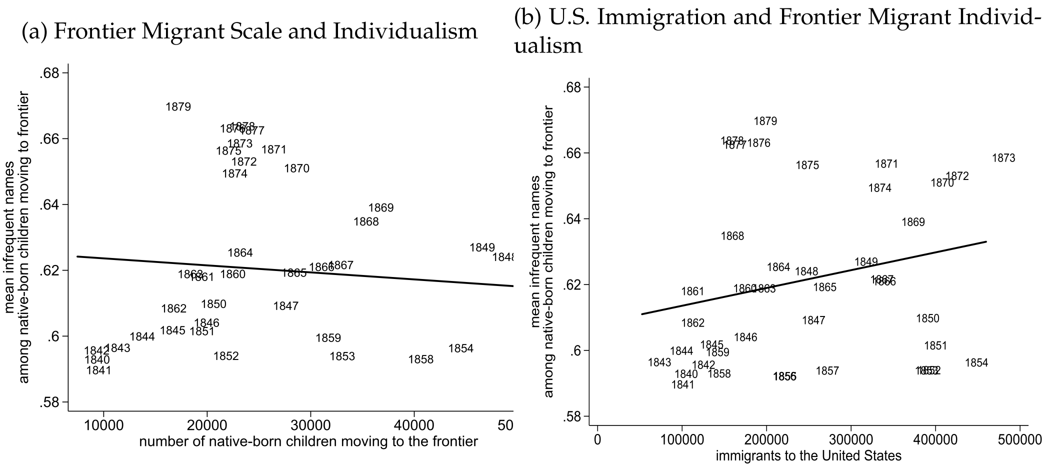

FIGURE D.4.—U.S. immigrant arrivals and native-born frontier migration scale and individualism. Graph (a) plots the relationship between the number of children brought to the frontier with their parents and the prevalence of individualistic names among them. Each point is a given year of migration computed based on the procedure used in Section 5 based on the 1850–1880 Censuses. Graph (b) plots the relationship between the number of immigrant arrivals to the United States in a given year and the prevalence of individualistic children's names among frontier migrants.

Figure D.4(b) shows that the prevalence of individualism among frontier migrants is not lower in years with more immigrants arriving to the United States, but rather weakly higher. This goes against the intuition that selective migration of individualists would be weaker in periods with greater push factors in settled areas. One explanation might be that non-individualists have stronger social networks that allow them to deal with adverse labor market shocks and ultimately remain in settled areas. Without deep social networks, individualists might be more readily pushed to move to the frontier. In any case, the observed patterns tend to alleviate a salient concern about the exclusion restriction in the IV estimation.[4]

APPENDIX E: FURTHER ROBUSTNESS CHECKS ON THE EXPOSURE EFFECTS IN SECTION 5.2

E.1. *Additional Results: Adulthood Exposure Event Study*

[4]Note that each of the graphs in Figure D.4 looks similar when allowing for (cumulative) lags in the number of immigrants to the United States and/or when using the alternative, predicted immigrant flows based on weather shocks in Europe.

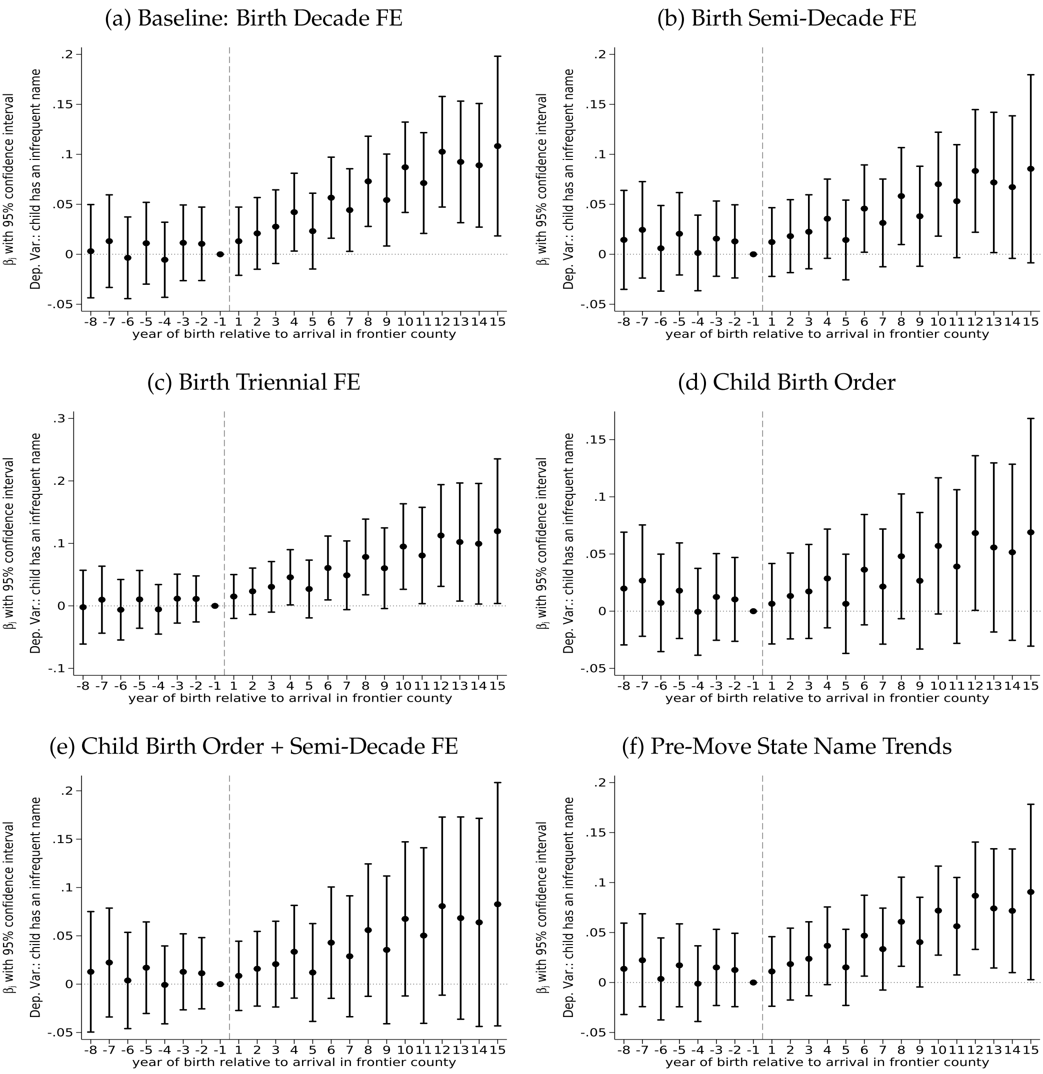

FIGURE E.1.—Identifying exposure effects: Adulthood exposure (I). These graphs report estimates of $\beta_j +/- 95$ percent confidence intervals in equation (6) for $j = -8, \ldots, 15$ (with other j suppressed for presentational purposes). Each β_j can be interpreted as the differential likelihood of an infrequent names being given to children born j years before/after their parents moved to the frontier, relative to children born one year prior to moving. The sample includes 57,097 children born to 16,901 families headed by white, native-born parents who moved with at least one child to a frontier county as we observe them in the Census in 1850, 1860, 1870, or 1880. All estimates control for household fixed effects and child gender. Graph (a) additionally includes child birth decade FE, (b) includes 5-yearly birth cohort, (c) includes 3-yearly birth cohort FE, (d) controls for child birth order, (e) controls for child birth order, and (f) controls for the mean gender-specific infrequent name share in each child birth year in the state from which each family migrated before arriving on the frontier. Standard errors are clustered by contemporaneous county.

E.2. *Additional Results: Childhood Exposure Age-at-Move*

As discussed in Section 5.2, our results are also robust to accounting for measurement error in the linking procedure. First, in Figure E.2(a) and panel (a) of Table E.I below,

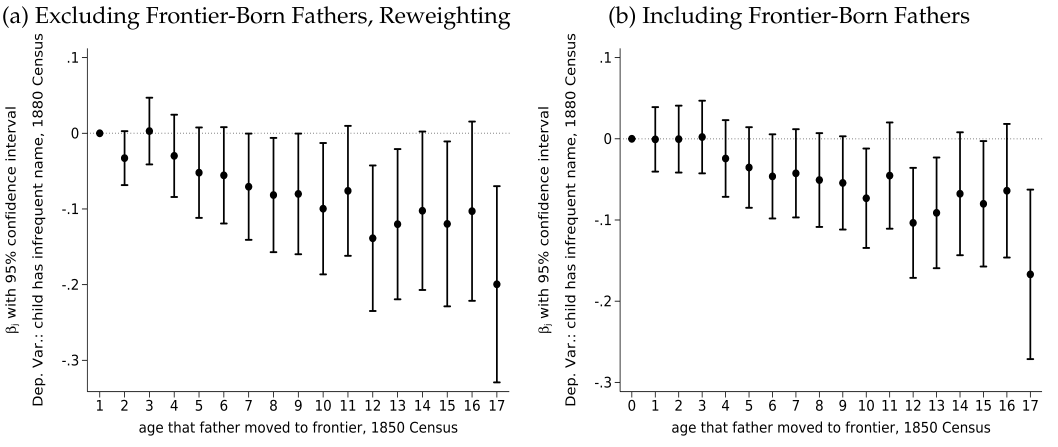

FIGURE E.2.—Identifying exposure effects: Adulthood exposure (I). Graph (a) reports a reweighted estimate of Figure 7. We reweight each child observation by the estimated odds (inverse probability weights) that the father was successfully linked across Census rounds. These weights are estimated as a function of the interaction of father's age in 1850 and whether or not the father has an infrequent name. In graph (b), the estimates are with respect to children born to fathers who were themselves born on the frontier. The sample in graph (a) (graph (b)) consists of 81,823 (146,085) children age 0–20 in the 1880 Census with fathers hailing from 17,778 (28,776) families observed in the 1850 Census and where at least two brothers (one brother) were born before the family moved to the frontier.

TABLE E.I

IDENTIFYING EXPOSURE EFFECTS: CHILDHOOD EXPOSURE (II)[a]

	(1)	(2)	(3)	(4)	(5)
			(a) Excluding Frontier-Born Fathers and Reweighting by Link Probability		
Age at Move to Frontier	−0.009	−0.008	−0.007	−0.006	−0.006
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	81,823	81,823	81,823	81,823	81,823
Number of Families	17,778	17,778	17,778	17,778	17,778
Mean of Dependent Variable	0.68	0.68	0.68	0.68	0.68
R^2	0.27	0.27	0.27	0.27	0.27
			Panel (b): Including Frontier-Born Fathers		
Age-at-Move to Frontier	−0.007	−0.006	−0.005	−0.004	−0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	146,085	146,085	146,085	146,085	146,085
Number of Families	28,776	28,776	28,776	28,776	28,776
Mean of Dependent Variable	0.69	0.69	0.69	0.69	0.69
R^2	0.25	0.25	0.25	0.25	0.25
Extended Family (1850 Household) FE	✓	✓	✓	✓	✓
State of Residence FE in 1880	–	✓	–	–	–
Child Birth Cohort FE	–	–	Decade	5-Yearly	3-Yearly

[a]This table reports a reweighted estimate of panel (a) in Table VIII. We reweight each child observation by the estimated odds (inverse probability weights) that the father was successfully linked across Census rounds. These weights are estimated as a function of the interaction of father's age in 1850 and whether or not the father has an infrequent name. Standard errors are two-way clustered by 1850 family and 1880 county.

we reweight the baseline sample by the odds of a successful link estimated as a flexible function of the father's age interacted with whether or not the father himself has an infrequent name. Following Bailey, Cole, and Massey (2020), we use these propensity scores to construct inverse probability weights. The results are very similar and in some specifications more precisely estimated than the baseline in Figure 7 and Table VIII. This helps to rule out a sample selection bias wherein fathers with more individualistic names and, hence, greater inherited individualism, might be more likely to be linked across Censuses. Second, we can further restrict our baseline sample to the children of fathers with unique matches between the 1850 and 1880 Census. This substantially cuts the sample by more than half, which leads to sizable reductions in statistical power. The resulting estimates of continuous age-at-move effects range from −0.010 (0.007) to −0.007 (0.007) across the analogous specifications 1 and 5 in Table VIII. In other words, while discarding the considerable information in non-unique matches we use in the baseline (which includes a dummy indicator for such matches), we find estimates that are quantitatively similar but noisy. We cannot reject that the coefficients equal zero, but we also cannot reject that they equal the baseline estimates.

Moreover, the age-at-move estimates are robust to including children whose fathers were born on the frontier. We generalize equation (8) to include all brothers who were born after their parents moved to the frontier, normalizing their ages at move to $j = 0$. Doing so in Figure E.2(b) and panel (b) of Table E.I suggests similar patterns, despite the substantial increase in the sample size.[5] We omit the children of fathers born on the frontier from the baseline in Figure 7 since they may only appear in the sample as a result of selective fertility among parents who found high returns to individualism after arrival to the frontier. This need not introduce a source of bias per se as these fathers would also have experienced more years of frontier conditions than their older siblings. However, they do introduce a source of sample selectivity just as they did in the event study. That the results look similar with and without these additional fathers suggests that this type of sample selectivity is not a first-order concern.

REFERENCES

BAILEY, M., C. COLE, AND C. MASSEY (2020): "Simple Strategies for Improving Inference With Linked Data: A Case Study of the 1850–1930 IPUMS Linked Representative Historical Samples," *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 53 (2), 80–93. [23]

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. [4-6,8,9,12-14,16]

GENTZKOW, M., J. M. SHAPIRO, AND M. TADDY (2019): "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," *Econometrica*, 87 (4), 1307–1340. [15]

GROGGER, J., AND G. H. HANSON (2011): "Income Maximization and the Selection and Sorting of International Migrants," *Journal of Development Economics*, 95 (1), 42–57. [19]

MARKUS, G. B. (2001): "American Individualism Reconsidered," in *Citizens and Politics: Perspectives From Political Psychology*, ed. by J. H. Kuklinski. New York: Cambridge University Press, 401–432. [5]

MIGRATION POLICY INSTITUTE (2016): "Tabulations of U.S. Department of Homeland Security, Office of Immigration Statistics, Yearbook of Immigration Statistics." [17]

OSTER, E. (2019): "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 37 (2), 187–204. [5,12,14]

[5]This procedure adds a large number of 1850 households for whom a first son was born prior to moving while a second (and higher-order) son was born after moving to the frontier.

SEQUEIRA, S., N. NUNN, AND N. QIAN (2020): "Immigrants and the Making of America," *The Review of Economic Studies*, 87 (1), 382–419. [18,19]

TUCKER, G. (1843): "Progress of Population and Wealth in the United States, in Fifty Years, as Exhibited by the Decennial Census Taken in That Period," *The Merchants' Magazine and Commercial Review*, 9, 136–144. [17]

Co-editor Fabrizio Zilibotti handled this manuscript.

Manuscript received 5 July, 2018; final version accepted 16 June, 2020; available online 14 July, 2020.