

SUPPLEMENTARY APPENDIX

to the paper

**CULTURAL DIVERSITY AND ECONOMIC GROWTH: EVIDENCE FROM THE
US DURING THE AGE OF MASS MIGRATION**

BY PHILIPP AGER AND MARKUS BRÜCKNER

April 2013

Contents

Regression Results	2
Descriptive Statistics	21
Graphs	23
Simulations	24
Data Sources and Variable Description	25

List of Tables

1	Appendix Table 1: Displaying all Controls for Table II – Panel A –	2
2	Appendix Table 2: Displaying all Controls for Table II – Panel B –	3
3	Appendix Table 3: Displaying all Controls for Table II – Panel C –	4
4	Appendix Table 4: Log-Specification	5
5	Appendix Table 5: Controlling for Past Output & Urban Growth	6
6	Appendix Table 6: Controlling for Historical Railway Access 1850-1870	7
7	Appendix Table 7: Different Time Periods 1880-1920	8
8	Appendix Table 8: Different Time Periods 1890-1920	9
9	Appendix Table 9: Different Time Periods 1900-1920	10
10	Appendix Table 10: Different Time Periods 1910-1920	11
11	Appendix Table 11: Panel Estimation 1870-1920	12
12	Appendix Table 12: Controlling for Population Share of American Indians	13
13	Appendix Table 13: Excluding the South	14
14	Appendix Table 14: Controlling for a Quadratic Fractionalization Term	15
15	Appendix Table 15: Using the Change in Control Variables	16
16	Appendix Table 16: Sample Split: Below vs. Above Median Output per Capita	17
17	Appendix Table 17: Excluding Top and Bottom 1 Percentile of Changes in the Fractionalization and Polarization Index	18
18	Appendix Table 18: Excluding Top and Bottom 1 Percentile of the Level of the Fractionalization and Polarization Index	19
19	Appendix Table 19: Linguistic Distances (LD), $\delta = 0.05$	20
20	Descriptive Statistics (1): Average Population Shares	21
21	Descriptive Statistics (2): Average Main Occupation Shares	22
22	Descriptive Statistics (3): Average Main Industry Shares	22

List of Figures

1	Graph (1): Kernel Density Estimation: Fractionalization	23
2	Graph (2): Kernel Density Estimation: Polarization	23
3	Simulation: Fractionalization vs. Polarization	24

Regression Results

Table 1: Displaying all Controls for Table II – Panel A –

	Dependent Variable: $\Delta \ln(\text{Output p.c.})$			
	LS	LS	LS	IV (SP)
ΔPOL	-0.763** (0.335)	-1.199*** (0.342)	-1.062*** (0.320)	-2.687*** (0.555)
$\Delta FRAC$	0.622 (0.535)	1.356** (0.569)	1.038** (0.485)	1.984** (0.832)
<i>Urbanization Rate 1870</i>		0.153 (0.139)	0.0839 (0.200)	0.128 (0.195)
<i>Labor Participation Rate 1870</i>		0.0513* (0.0293)	0.0536* (0.0289)	0.0148 (0.0325)
<i>Population 1870</i>		-0.00379 (0.0295)	0.0488 (0.0296)	0.0642** (0.0302)
<i>Output p.c. 1870</i>		-0.690*** (0.0445)	-0.699*** (0.0477)	-0.702*** (0.0483)
<i>Land Concentration 1870</i>		-0.484*** (0.165)	-0.506*** (0.159)	-0.412** (0.168)
<i>Manufacturing Share 1870</i>		-0.0896 (0.114)	-0.108 (0.108)	-0.138 (0.101)
<i>Rail Access 1870</i>		0.0880*** (0.0255)	0.0984*** (0.0265)	0.0865*** (0.0292)
<i>Population Growth 1870-1920</i>			0.0676 (0.0424)	0.126** (0.0531)
<i>Share Austro-Hungarian 1870</i>			-0.0156 (1.347)	0.843 (1.455)
<i>Share Benelux 1870</i>			0.389 (0.970)	0.261 (0.941)
<i>Share East Europeans 1870</i>			-2.694 (6.987)	-16.59* (8.972)
<i>Share Canadian 1870</i>			-0.0300 (0.699)	0.349 (0.810)
<i>Share Central and South America 1870</i>			-2.006*** (0.563)	-1.379* (0.802)
<i>Share Scandinavia 1870</i>			-0.267 (0.450)	-0.374 (0.617)
<i>Share French 1870</i>			-2.472 (2.559)	-0.926 (2.860)
<i>Share Germans 1870</i>			-0.900 (0.540)	-1.545** (0.775)
<i>Share Irish 1870</i>			-1.652 (1.013)	-1.702 (1.118)
<i>Share Italians 1870</i>			-5.897* (3.319)	-2.691 (4.019)
<i>Share Pacific 1870</i>			-1.159 (0.805)	-0.917 (0.888)
<i>Share Polish 1870</i>			-14.37 (36.04)	-20.15 (39.76)
<i>Share Portuguese 1870</i>			-80.07 (206.3)	-157.9 (202.1)
<i>Share Spanish 1870</i>			-80.31*** (11.62)	-80.06*** (12.24)
<i>Share Swiss 1870</i>			-1.197 (2.451)	-1.460 (2.655)
<i>Share United Kingdom 1870</i>			-1.436 (1.328)	-1.120 (1.385)
<i>Share Asian 1870</i>			-1.700** (0.782)	-1.493 (0.992)
<i>Share Afro-American 1870</i>		-0.387 (0.350)	-1.370** (0.596)	-0.822 (0.850)
<i>Share White US Born 1870</i>		-0.172 (0.398)	-1.114** (0.546)	-0.700 (0.762)
Observations	2160	2160	2160	2160
R^2	0.265	0.546	0.564	-

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In this table we display the control variables (except state fixed effects) that correspond to Table 2, Panel (A).

Table 2: Displaying all Controls for Table II – Panel B –

	Dependent Variable: $\Delta \ln(\text{Output p.c.})$			
	LS	LS	LS	IV (SP)
ΔPOL	-0.375*** (0.0891)	-0.424*** (0.134)	-0.497*** (0.138)	-1.737*** (0.390)
<i>Urbanization Rate 1870</i>		0.166 (0.146)	0.0713 (0.202)	0.108 (0.199)
<i>Labor Participation Rate 1870</i>		0.0527 (0.0314)	0.0532* (0.0297)	0.00947 (0.0332)
<i>Population 1870</i>		-0.00565 (0.0303)	0.0541* (0.0297)	0.0766** (0.0301)
<i>Output p.c. 1870</i>		-0.692*** (0.0464)	-0.696*** (0.0487)	-0.696*** (0.0501)
<i>Land Concentration 1870</i>		-0.479*** (0.169)	-0.511*** (0.161)	-0.411** (0.171)
<i>Manufacturing Share 1870</i>		-0.116 (0.119)	-0.119 (0.112)	-0.164 (0.102)
<i>Rail Access 1870</i>		0.0898*** (0.0258)	0.0995*** (0.0264)	0.0872*** (0.0290)
<i>Population Growth 1870-1920</i>			0.0833** (0.0407)	0.164*** (0.0502)
<i>Share Austro-Hungarian 1870</i>			-0.349 (1.311)	0.270 (1.413)
<i>Share Benelux 1870</i>			0.279 (0.960)	0.0241 (0.958)
<i>Share East Europeans 1870</i>			-5.635 (6.973)	-24.14*** (8.706)
<i>Share Canadian 1870</i>			-0.150 (0.628)	0.151 (0.684)
<i>Share Central and South America 1870</i>			-1.843*** (0.474)	-0.978 (0.612)
<i>Share Scandinavia 1870</i>			-0.220 (0.380)	-0.291 (0.472)
<i>Share French 1870</i>			-3.306 (2.619)	-2.428 (2.955)
<i>Share Germans 1870</i>			-0.882* (0.479)	-1.584** (0.632)
<i>Share Irish 1870</i>			-1.732* (0.990)	-1.868* (1.009)
<i>Share Italians 1870</i>			-7.019** (3.232)	-4.580 (3.645)
<i>Share Pacific 1870</i>			-0.979 (0.745)	-0.526 (0.708)
<i>Share Polish 1870</i>			-11.79 (37.68)	-15.62 (42.73)
<i>Share Portuguese 1870</i>			-25.26 (218.5)	-56.35 (227.8)
<i>Share Spanish 1870</i>			-79.27*** (11.67)	-77.93*** (12.37)
<i>Share Swiss 1870</i>			-2.145 (2.356)	-3.405 (2.373)
<i>Share United Kingdom 1870</i>			-1.407 (1.320)	-1.025 (1.370)
<i>Share Asian 1870</i>			-1.878** (0.796)	-1.828* (1.051)
<i>Share Afro-American 1870</i>		-0.0555 (0.293)	-1.097** (0.482)	-0.208 (0.655)
<i>Share White US Born 1870</i>		0.172 (0.318)	-0.844* (0.434)	-0.106 (0.529)
Observations	2160	2160	2160	2160
R^2	0.263	0.541	0.562	-

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In this table we display the control variables (except state fixed effects) that correspond to Table 2, Panel (B).

Table 3: Displaying all Controls for Table II – Panel C –

	Dependent Variable: $\Delta \ln(\text{Output p.c.})$			
	LS	LS	LS	IV (SP)
ΔFRAC	-0.406** (0.176)	-0.423* (0.241)	-0.608*** (0.220)	-1.422** (0.572)
<i>Urbanization Rate 1870</i>		0.153 (0.149)	0.0591 (0.205)	0.0623 (0.201)
<i>Labor Participation Rate 1870</i>		0.0560* (0.0326)	0.0588* (0.0308)	0.0429 (0.0307)
<i>Population 1870</i>		-0.00375 (0.0313)	0.0542* (0.0302)	0.0664** (0.0312)
<i>Output p.c. 1870</i>		-0.688*** (0.0471)	-0.695*** (0.0491)	-0.692*** (0.0495)
<i>Land Concentration 1870</i>		-0.501*** (0.171)	-0.527*** (0.164)	-0.495*** (0.166)
<i>Manufacturing Share 1870</i>		-0.116 (0.121)	-0.119 (0.114)	-0.144 (0.111)
<i>Rail Access 1870</i>		0.0926*** (0.0263)	0.102*** (0.0268)	0.0982*** (0.0257)
<i>Population Growth 1870-1920</i>			0.0816* (0.0416)	0.123*** (0.0474)
<i>Share Austro-Hungarian 1870</i>			-0.628 (1.310)	-0.669 (1.295)
<i>Share Benelux 1870</i>			0.249 (0.979)	0.0714 (0.979)
<i>Share East Europeans 1870</i>			-4.875 (7.359)	-13.80 (9.620)
<i>Share Canadian 1870</i>			-0.261 (0.586)	-0.248 (0.561)
<i>Share Central and South America 1870</i>			-1.864*** (0.459)	-1.426*** (0.529)
<i>Share Scandinavia 1870</i>			-0.182 (0.352)	-0.170 (0.334)
<i>Share French 1870</i>			-3.912 (2.545)	-4.252* (2.559)
<i>Share Germans 1870</i>			-0.777* (0.456)	-1.014** (0.493)
<i>Share Irish 1870</i>			-1.760* (0.997)	-1.871* (0.976)
<i>Share Italians 1870</i>			-8.003** (3.109)	-8.011*** (2.970)
<i>Share Pacific 1870</i>			-0.934 (0.747)	-0.631 (0.693)
<i>Share Polish 1870</i>			-9.764 (37.94)	-9.107 (40.48)
<i>Share Portuguese 1870</i>			11.06 (226.3)	42.93 (239.4)
<i>Share Spanish 1870</i>			-78.85*** (11.68)	-77.55*** (11.45)
<i>Share Swiss 1870</i>			-2.532 (2.337)	-3.727 (2.434)
<i>Share United Kingdom 1870</i>			-1.441 (1.319)	-1.282 (1.321)
<i>Share Asian 1870</i>			-1.989** (0.789)	-2.111** (0.836)
<i>Share Afro-American 1870</i>		-0.0653 (0.334)	-1.056** (0.472)	-0.525 (0.572)
<i>Share White US Born 1870</i>		0.181 (0.366)	-0.785* (0.426)	-0.309 (0.469)
Observations	2160	2160	2160	2160
R^2	0.260	0.536	0.557	-

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In this table we display the control variables (except state fixed effects) that correspond to Table 2, Panel (C).

Table 4: Log-Specification

	$\Delta \ln(\text{Output p.c.})$			
	LS	LS	LS	IV (SP)
$\Delta \ln(POL)$	-1.279** (0.530)	-1.875*** (0.581)	-1.644*** (0.532)	-4.648*** (1.130)
$\Delta \ln(FRAC)$	0.991 (0.758)	1.893** (0.859)	1.393* (0.733)	2.877* (1.557)
Observations	2160	2160	2160	2160
R^2	0.266	0.546	0.565	-
First-stage (Kleibergen-Paap) F-Statistic	-	-	-	28.66
Anderson-Rubin Wald-Test (p-val.)	-	-	-	0.00
Endogeneity Test Statistic POL (p-val.)	-	-	-	0.00
Endogeneity Test Statistic FRAC (p-val.)	-	-	-	0.00
State FE	yes	yes	yes	yes
Initial Controls	no	yes	yes	yes
Population Shares	no	no	yes	yes
Population Growth	no	no	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In columns (1)-(3) the method of estimation is least squares. In column (4) the method of estimation is two-stage least squares. The instrumental variable is the supply-push component of immigrant inflows, IV (Supply Push); also see Section 4 for further details. Initial control variables (1870) are output per capita, the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 5: Controlling for Past Output & Urban Growth

	$\Delta \ln(\text{Output p.c.})$				
	(1)	(2)	(3)	(4)	(5)
ΔPOL	-2.745*** (0.713)	-2.606*** (0.972)	-2.801*** (0.633)	-2.951*** (0.829)	-3.758* (1.925)
$\Delta FRAC$	3.113** (1.269)	2.080 (1.983)	2.141** (0.953)	2.201 (1.717)	3.183 (3.260)
<i>Output Growth 1860-1870</i>	0.276*** (0.0571)				
<i>Output Growth 1850-1870</i>		0.0879*** (0.0286)			
<i>Urban Growth 1860-1870</i>			-0.0154 (0.175)		
<i>Urban Growth 1850-1870</i>				-0.148 (0.194)	
<i>Urban Growth 1820-1870</i>					0.810** (0.371)
Observations	1983	1606	2001	1606	751
First-stage (Kleibergen-Paap) F-Statistic	17.04	15.46	26.75	22.18	11.14
Anderson-Rubin Wald-Test (p-val.)	0.00	0.00	0.00	0.00	0.00
Endogeneity Test Statistic POL (p-val.)	0.05	0.01	0.00	0.00	0.01
Endogeneity Test Statistic FRAC (p-val.)	0.17	0.02	0.01	0.00	0.02
State FE	yes	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1. In columns (1)-(5) the method of estimation is two-stage least squares. The instrumental variable for the polarization and fractionalization index is the supply-push component of immigrant inflows, respectively; also see Section 4 for further details. The instrumental variable for past output growth is 1860 output per capita in column (1) and 1850 output per capita in column (2). Initial control variables (1870) are output per capita (in columns (3)-(5)), the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 6: Controlling for Historical Railway Access 1850-1870

	$\Delta \ln(\text{Output p.c.})$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔPOL	-2.745*** (0.713)	-2.738*** (0.710)	-3.250*** (0.882)	-3.246*** (0.882)	-2.606*** (0.972)	-2.685*** (0.942)	-2.646*** (0.938)	-2.689*** (0.933)
$\Delta FRAC$	3.113** (1.269)	3.113** (1.258)	3.823* (1.983)	3.835** (1.955)	2.080 (1.983)	2.275 (1.921)	2.128 (1.919)	2.271 (1.897)
<i>Railway Access 1870</i>	0.0339 (0.0252)			0.00754 (0.0322)	0.0304 (0.0291)			0.0132 (0.0313)
<i>Railway Access 1860</i>		0.0409 (0.0308)		0.00892 (0.0312)		0.0486* (0.0283)		0.0273 (0.0306)
<i>Railway Access 1850</i>			0.0569* (0.0333)	0.0512 (0.0338)			0.0656** (0.0322)	0.0502 (0.0316)
<i>Output Growth 1860-1870</i>	0.276*** (0.0571)	0.273*** (0.0573)	0.205* (0.118)	0.205* (0.118)				
<i>Output Growth 1850-1870</i>					0.0879*** (0.0286)	0.0885*** (0.0288)	0.0885*** (0.0295)	0.0891*** (0.0289)
Observations	1983	1983	1603	1603	1606	1605	1605	1604
First-stage (Kleibergen-Paap) F-Statistic	17.04	17.35	17.01	17.20	15.46	16.38	15.43	16.24
Anderson-Rubin Wald-Test (p-val.)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
State FE	yes	yes	yes	yes	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1. In columns (1)-(8) the method of estimation is two-stage least squares. The instrumental variable for the polarization and fractionalization index is the supply-push component of immigrant inflows, respectively; also see Section 4 for further details. The instrumental variable for past output growth is 1860 output per capita in columns (1)-(4) and 1850 output per capita in columns (5)-(8). Initial control variables (1870) are the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 7: Different Time Periods 1880-1920

	Dependent Variable: $\Delta \ln(\text{Output p.c.})$			
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL	-1.098** (0.437)	-2.753** (1.085)	-5.208*** (1.771)	-3.512*** (1.055)
$\Delta FRAC$	1.268* (0.669)	2.636 (1.863)	6.728** (3.258)	3.801** (1.877)
<i>Output Growth 1850-1880</i>				0.0304 (0.0212)
Observations	1607	1607	1607	1607
R^2	0.405	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	28.69	9.73	21.18
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.21
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
State FE	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1. In column (1) the method of estimation is least squares. In columns (2)-(5) the method of estimation is two-stage least squares. In column (2), the instrumental variable is the supply-push component of immigrant inflows, IV (Supply Push); also see Section 4 for further details. In column (3) the instrumental variables are the initial polarization and fractionalization index (IV (Initial)). In columns (4)-(5), the instrumental variables are the supply-push component of immigrant inflows and the initial polarization and fractionalization index (IV (Supply Push and Initial)). The instrumental variable for past output growth is 1850 output per capita in column (5). Initial control variables (1880) are output per capita (in columns (1)-(4)), the urbanization rate, land concentration, the manufacturing share, population size, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 8: Different Time Periods 1890-1920

	Dependent Variable: $\Delta \ln(\text{Output p.c.})$			
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL	-1.709*** (0.355)	-2.575*** (0.617)	-3.697*** (0.804)	-2.942*** (0.561)
$\Delta FRAC$	2.372*** (0.683)	2.070* (1.090)	4.080** (1.757)	2.166* (1.180)
<i>Output Growth 1860-1890</i>				0.0985*** (0.0344)
Observations	1970	1970	1970	1970
R^2	0.478	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	39.99	57.15	29.79
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.33
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
State FE	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1. In column (1) the method of estimation is least squares. In columns (2)-(5) the method of estimation is two-stage least squares. In column (2), the instrumental variable is the supply-push component of immigrant inflows, IV (Supply Push); also see Section 4 for further details. In column (3) the instrumental variables are the initial polarization and fractionalization index (IV (Initial)). In columns (4)-(5), the instrumental variables are the supply-push component of immigrant inflows and the initial polarization and fractionalization index (IV (Supply Push and Initial)). The instrumental variable for past output growth is 1860 output per capita in column (5). Initial control variables (1890) are output per capita (in columns (1)-(4)), the urbanization rate, land concentration, the manufacturing share, population size, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 9: Different Time Periods 1900-1920

	Dependent Variable: $\Delta \ln(\text{Output p.c.})$			
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL	-1.505*** (0.483)	-4.998*** (1.172)	-4.387** (1.706)	-4.719*** (1.282)
$\Delta FRAC$	1.822** (0.818)	6.823*** (2.292)	5.961 (3.688)	6.659** (2.748)
<i>Output Growth 1870-1900</i>				-0.0437 (0.0328)
Observations	2221	2221	2221	2221
R^2	0.343	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	30.35	34.65	35.64
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.47
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
State FE	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1. In column (1) the method of estimation is least squares. In columns (2)-(5) the method of estimation is two-stage least squares. In column (2), the instrumental variable is the supply-push component of immigrant inflows, IV (Supply Push); also see Section 4 for further details. In column (3) the instrumental variables are the initial polarization and fractionalization index (IV (Initial)). In columns (4)-(5), the instrumental variables are the supply-push component of immigrant inflows and the initial polarization and fractionalization index (IV (Supply Push and Initial)). The instrumental variable for past output growth is 1870 output per capita in column (5). Initial control variables (1900) are output per capita (in columns (1)-(4)), the urbanization rate, land concentration, the manufacturing share, population size, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 10: Different Time Periods 1910-1920

	Dependent Variable: $\Delta \ln(\text{Output p.c.})$			
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL	-0.0465 (0.851)	-1.810*** (0.691)	-4.117*** (1.551)	-2.360*** (0.786)
$\Delta FRAC$	-0.139 (0.787)	2.604* (1.500)	7.161* (3.890)	3.555* (2.032)
<i>Output Growth 1880-1910</i>				-1.899*** (0.699)
				3.045* (1.824)
				0.0361 (0.0272)
Observations	2401	2401	2401	2401
R^2	0.460	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	34.43	25.78	45.23
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.16
Anderson-Rubin Wald-Test (p-val.)	-	0.03	0.00	0.02
Initial Controls	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1. In column (1) the method of estimation is least squares. In columns (2)-(5) the method of estimation is two-stage least squares. In column (2), the instrumental variable is the supply-push component of immigrant inflows, IV (Supply Push); also see Section 4 for further details. In column (3) the instrumental variables are the initial polarization and fractionalization index (IV (Initial)). In columns (4)-(5), the instrumental variables are the supply-push component of immigrant inflows and the initial polarization and fractionalization index (IV (Supply Push and Initial)). The instrumental variable for past output growth is 1880 output per capita in column (5). Initial control variables (1910) are output per capita (in columns (1)-(4)), the urbanization rate, land concentration, the manufacturing share, population size, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 11: Panel Estimation 1870-1920

	Dependent Variable: $\Delta \ln(\text{Output p.c.})$			
	LS	LS	LS	IV (SP)
ΔPOL	-0.770*** (0.238)	-0.715*** (0.240)	-0.677*** (0.245)	-1.309** (0.539)
$\Delta FRAC$	1.216*** (0.369)	1.206*** (0.377)	1.268*** (0.387)	1.598** (0.777)
<i>Lagged Output Growth</i>				0.111*** (0.0249)
Observations	11748	11748	11748	11748
R^2	0.672	0.683	0.687	-
First-stage (Kleibergen-Paap) F-Statistic	-	-	-	23.93
Anderson-Rubin Wald-Test (p-val.)	-	-	-	0.00
State \times Year FE	yes	yes	yes	yes
Lagged Controls	no	yes	yes	yes
Lagged Population Shares	no	no	yes	yes
Population Growth	no	no	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the county level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In columns (1)-(3) the method of estimation is least squares. In column (4) the method of estimation is two-stage least squares. The instrumental variable is the supply-push component of immigrant inflows, IV (Supply Push); also see Section 4 for further details. In column (4) the instrumental variable for lagged output growth is the second lag of output per capita. Lagged control variables are output per capita, the urbanization rate, land concentration, the manufacturing share, population size, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 12: Controlling for Population Share of American Indians

	$\Delta \ln(\text{Output p.c.})$			
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL	-1.055*** (0.325)	-2.672*** (0.545)	-2.332*** (0.453)	-2.206*** (0.423)
$\Delta FRAC$	1.024** (0.492)	1.957** (0.820)	0.810 (1.191)	1.195 (0.890)
Observations	2160	2160	2160	2160
R^2	0.564	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	28.00	12.94	17.20
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.31
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
State FE	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In column (1) the method of estimation is least squares. In columns (2)-(4) the method of estimation is two-stage least squares. In column (2), the instrumental variable is the supply-push component of immigrant inflows, IV (Supply Push); also see Section 4 for further details. In column (3) the instrumental variables are the initial 1870 polarization and fractionalization index (IV (Initial)). In column (4), the instrumental variables are the supply-push component of immigrant inflows and the initial 1870 polarization and fractionalization index (IV (Supply Push and Initial)). Initial control variables (1870) are output per capita, the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 13: Excluding the South

	$\Delta \ln(\text{Output p.c.})$			
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL	-1.532*** (0.313)	-2.191*** (0.421)	-1.944*** (0.406)	-2.008*** (0.402)
$\Delta FRAC$	1.707*** (0.509)	0.673 (0.962)	-0.222 (0.926)	0.140 (0.886)
Observations	1125	1125	1125	1125
R^2	0.614	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	20.99	34.95	18.01
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.30
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
State FE	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In column (1) the method of estimation is least squares. In columns (2)-(4) the method of estimation is two-stage least squares. In column (2), the instrumental variable is the supply-push component of immigrant inflows, IV (Supply Push); also see Section 4 for further details. In column (3) the instrumental variables are the initial 1870 polarization and fractionalization index (IV (Initial)). In column (4), the instrumental variables are the supply-push component of immigrant inflows and the initial 1870 polarization and fractionalization index (IV (Supply Push and Initial)). Initial control variables (1870) are output per capita, the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 14: Controlling for a Quadratic Fractionalization Term

	$\Delta \ln(\text{Output p.c.})$							
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL	-0.440 (0.324)	-2.342** (0.988)	-1.916** (0.953)	-1.455** (0.738)	-2.146*** (0.456)	-5.403*** (0.933)	-5.895*** (0.964)	-5.010*** (0.894)
$\Delta FRAC$	-1.153 (0.832)	0.564 (2.650)	-0.893 (2.628)	-1.497 (2.120)	2.628*** (0.626)	5.443*** (1.018)	4.005*** (1.443)	4.564*** (1.124)
$\Delta FRAC^2$	2.095*** (0.695)	1.344 (2.084)	1.374 (2.304)	2.315 (1.752)				
Observations	2160	2160	2160	2160	2160	2160	2160	2160
R^2	0.569	-	-	-	0.568	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	23.50	16.05	13.00	-	34.98	17.41	19.08
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00	-	0.00	0.00	0.00
State FE	yes	yes	yes	yes	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1. In columns (1) and (5) the method of estimation is least squares. In columns (2)-(4) and (6)-(8) the method of estimation is two-stage least squares. In columns (2) and (6) the instrumental variable is the supply-push component of immigrant inflows (IV (SP)); also see Section 4 for further details. In columns (3) and (7) the instrumental variables are the initial 1870 polarization and fractionalization index (IV (Initial)). In columns (4) and (8), the instrumental variables are the supply-push component of immigrant inflows and the initial 1870 polarization and fractionalization index (IV (SP and Initial)). Initial control variables (1870) are output per capita, the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 15: Using the Change in Control Variables

	$\Delta \ln(\text{Output p.c.})$			
	(1) LS	(2) IV (SP)	(3) IV (Initial)	(4) IV (SP and Initial)
ΔPOL	-1.110*** (0.260)	-2.292*** (0.561)	-1.834*** (0.366)	-1.701*** (0.352)
$\Delta FRAC$	1.158** (0.491)	2.062*** (0.719)	0.530 (1.063)	0.928 (0.835)
Observations	2109	2109	2109	2109
R^2	0.642	-	-	-
Kleibergen-Paap F-Statistic	-	19.49	10.51	15.78
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.20
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
State FE	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes
Change Controls	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes
Change Population Shares	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In column (1) the method of estimation is least squares. In columns (2)-(4) the method of estimation is two-stage least squares. In column (2) the instrumental variable is the supply-push component of immigrant inflows (IV (A)); also see Section 4 for further details. In column (3) the instrumental variables are the initial 1870 polarization and fractionalization index (IV (B)). In column (4), the instrumental variables are the supply-push component of immigrant inflows and the initial 1870 polarization and fractionalization index (IV (C)). Initial control variables (1870) are output per capita, the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of Afro-Americans (estimates not reported in the table). The change of control variables (1870-1920) included are the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 16: Sample Split: Below vs. Above Median Output per Capita

	Above Median Output p.c.				Below Median Output p.c.			
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL	-1.744*** (0.417)	-2.669*** (0.697)	-2.233*** (0.525)	-2.301*** (0.571)	-0.171 (0.441)	-2.622*** (0.635)	-2.565*** (0.615)	-2.318*** (0.600)
$\Delta FRAC$	2.355*** (0.591)	2.346* (1.327)	1.287 (1.335)	1.693 (1.215)	-0.676 (0.692)	2.089** (0.913)	1.393 (1.789)	1.368 (1.336)
Observations	1080	1080	1080	1080	1080	1080	1080	1080
R^2	0.537	-	-	-	0.576	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	21.01	9.57	14.41	-	24.96	16.48	15.48
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00	-	0.00	0.00	0.00
State FE	yes	yes	yes	yes	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** p<0.01, ** p<0.05, * p<0.1. In columns (1) and (5) the method of estimation is least squares. In columns (2)-(4) and (6)-(8) the method of estimation is two-stage least squares. In columns (2) and (6) the instrumental variable is the supply-push component of immigrant inflows (IV (SP)); also see Section 4 for further details. In columns (3) and (7) the instrumental variables are the initial 1870 polarization and fractionalization index (IV (Initial)). In columns (4) and (8), the instrumental variables are the supply-push component of immigrant inflows and the initial 1870 polarization and fractionalization index (IV (SP and Initial)). Initial control variables (1870) are output per capita, the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of African-Americans (estimates not reported in the table).

Table 17: Excluding Top and Bottom 1 Percentile of Changes in Fractionalization and Polarization

PANEL A: Between 1 and 99 Percentile Δ FRAC Index				
Dependent Variable: $\Delta \ln(\text{Output p.c.})$				
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL	-0.967*** (0.313)	-2.720*** (0.573)	-2.382*** (0.495)	-2.186*** (0.452)
$\Delta FRAC$	0.797 (0.500)	1.773** (0.838)	0.491 (1.263)	0.986 (0.904)
Observations	2126	2126	2126	2126
R^2	0.573	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	26.09	10.88	18.21
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.17
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
PANEL B: Between 1 and 99 Percentile Δ POL Index				
ΔPOL	-1.142*** (0.316)	-3.012*** (0.615)	-2.651*** (0.476)	-2.405*** (0.437)
$\Delta FRAC$	1.119** (0.496)	2.388*** (0.863)	1.066 (1.179)	1.509* (0.863)
Observations	2124	2124	2124	2124
R^2	0.563	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	24.42	10.49	17.29
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.21
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
PANEL C: Between 1 and 99 Percentile Output Growth				
ΔPOL	-0.940*** (0.308)	-2.244*** (0.538)	-1.985*** (0.435)	-1.797*** (0.394)
$\Delta FRAC$	0.892* (0.460)	1.623** (0.734)	0.589 (1.056)	0.855 (0.844)
Observations	2128	2128	2128	2128
R^2	0.531	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	24.50	12.52	14.00
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.13
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In column (1) the method of estimation is least squares. In columns (2)-(4) the method of estimation is two-stage least squares. In column (2) the instrumental variable is the supply-push component of immigrant inflows (IV (SP)); also see Section 4 for further details. In column (3) the instrumental variables are the initial 1870 polarization and fractionalization index (IV (Initial)). In column (4) the instrumental variables are the supply-push component of immigrant inflows and the initial 1870 polarization and fractionalization index (IV (SP and Initial)). Initial control variables (1870) are output per capita, the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of African-Americans. We further include state fixed effects, population growth and population shares as additional control variables (estimates not reported in the table).

Table 18: Excluding Top and Bottom 1 Percentile of the Level of Fractionalization and Polarization

PANEL A: Between 1 and 99 Percentile FRAC Index				
Dependent Variable: $\Delta \ln(\text{Output p.c.})$				
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL	-1.022*** (0.334)	-2.508*** (0.581)	-2.224*** (0.463)	-2.110*** (0.433)
$\Delta FRAC$	0.979* (0.504)	1.855** (0.844)	0.883 (1.217)	1.194 (0.936)
Observations	2041	2041	2041	2041
R^2	0.569	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	30.96	14.01	19.81
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.45
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
PANEL B: Between 1 and 99 Percentile POL Index				
ΔPOL	-1.084*** (0.320)	-2.612*** (0.544)	-2.299*** (0.446)	-2.202*** (0.420)
$\Delta FRAC$	1.058** (0.484)	1.970** (0.820)	0.934 (1.160)	1.271 (0.881)
Observations	2115	2115	2115	2115
R^2	0.564	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	32.65	15.29	19.27
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.38
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
PANEL C: Between 1 and 99 Percentile POL and FRAC Index				
ΔPOL	-1.041*** (0.331)	-2.454*** (0.566)	-2.183*** (0.451)	-2.094*** (0.427)
$\Delta FRAC$	0.997* (0.502)	1.840** (0.838)	0.938 (1.198)	1.217 (0.932)
Observations	2010	2010	2010	2010
R^2	0.569	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	34.44	16.08	20.66
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.49
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In column (1) the method of estimation is least squares. In columns (2)-(4) the method of estimation is two-stage least squares. In column (2) the instrumental variable is the supply-push component of immigrant inflows (IV (SP)); also see Section 4 for further details. In column (3) the instrumental variables are the initial 1870 polarization and fractionalization index (IV (Initial)). In column (4) the instrumental variables are the supply-push component of immigrant inflows and the initial 1870 polarization and fractionalization index (IV (SP and Initial)). Initial control variables (1870) are output per capita, the urbanization rate, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of African-Americans. We further include state fixed effects, population growth and population shares as additional control variables (estimates not reported in the table).

Table 19: Pol and Frac Index with Linguistic Distances (LD), $\delta = 0.05$

	$\Delta \ln(\text{Output p.c.})$			
	LS	IV (SP)	IV (Initial)	IV (SP and Initial)
ΔPOL_{LD}	-148.0*** (37.81)	-212.3*** (58.77)	-204.1*** (57.78)	-219.0*** (58.19)
$\Delta FRAC_{LD}$	50.33*** (15.74)	59.41*** (18.85)	50.64*** (19.57)	68.93*** (19.97)
Observations	2161	2161	2161	2161
R^2	0.552	-	-	-
First-stage (Kleibergen-Paap) F-Statistic	-	21.47	13.17	8.70
Hansen-J statistic (p-val.)	-	n.a.	n.a.	0.12
Anderson-Rubin Wald-Test (p-val.)	-	0.00	0.00	0.00
State FE	yes	yes	yes	yes
Initial Controls	yes	yes	yes	yes
Population Shares	yes	yes	yes	yes
Population Growth	yes	yes	yes	yes

Huber robust standard errors (shown in parentheses) are clustered at the state level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In column (1) the method of estimation is least squares. In columns (2)-(4) the method of estimation is two-stage least squares. In column (2) the instrumental variable is the supply-push component of immigrant inflows (IV (Supply Push)); also see Section 4 for further details. In column (3) the instrumental variables are the initial 1870 polarization and fractionalization index (IV (Initial)). In column (4) the instrumental variables are the supply-push component of immigrant inflows and the initial 1870 polarization and fractionalization index (IV (Supply Push and Initial)). Initial control variables (1870) are the urbanization share, output per capita, land concentration, the manufacturing share, population size, labor participation rate, counties' rail access, the share of native-born white and the share of African-Americans (estimates not reported in the table). We construct our measure of fractionalization with distances following Greenberg (1956) as: $FRAC_{LD} = \sum_{i=1}^N \sum_{j=1}^N \pi_i \pi_j d_{ij}$, where subindices for counties and states are left out for simplicity. The corresponding measure of polarization follows Esteban and Ray (1994, 1999) and is constructed as: $POL_{LD} = \sum_{i=1}^N \sum_{j=1}^N \pi_i^2 \pi_j d_{ij}$. Both indices include a measure of inter-group distances d_{ij} . To proxy for inter-group distances, we follow Fearon (2003), Desmet et al. (2009a) and Esteban et al. (2010) and use information on linguistic groups compiled by the Ethnologue project to construct a measure of linguistic distances between any two groups as $d_{ij} = 1 - b_{ij}^\delta$ (See <http://www.ethnologue.com>, Desmet et al. (2009a) and Esteban et al. (2010) for further information on the Ethnologue project). The parameter b_{ij} is the ratio of the number of shared branches between i and j to the maximum number of branches between any two languages and $\delta \in (0, 1]$ represents a sensitivity parameter determining how fast the distance declines as the number of shared branches increases (see Desmet et al., 2009a) (We used the language trees reported by Ethnologue to construct the parameter b_{ij} . See also Desmet et al. (2009b) for a detailed discussion on the construction of such a language tree). The abbreviation LD denotes the use of language distances for the metric d_{ij} . We use the representative language of each country of origin to construct the linguistic distances. Thus, if two groups speak the same representative language we set $b_{ij} = 1$. We compute linguistic distance as Desmet et al. (2009a) using $\delta = 0.05$.

Descriptive Statistics

Table 20: Average Population Shares: Years 1870, 1920

Year	Share 1	Share 2	Share 3	Share 4	Share 5	Share 6	Share 7
1870	.0022978	.0015995	.0013146	.014865	.0432896	.0131676	.0022796
1920	.005545	.0017838	.0062861	.0139002	.0136935	.0161658	.0008786

Year	Share 8	Share 9	Share 10	Share 11	Share 12	Share 13	Share 14
1870	.0276525	.	.0233431	.0005533	.0012711	.0001813	.0002078
1920	.0121973	.0008646	.003241	.0050011	.0020155	.0029711	.0065188

Year	Share 15	Share 16	Share 17	Share 18	Share 19	Share 20	Share 21
1870	.0005284	.001831	.0163596	.0480108	.0057655	.1503518	.7333189
1920	.001228	.0011946	.0064377	n.a.	.0096398	.1165702	.80195

Share 1: Austro-Hungarian; Share 2: Benelux; Share 3: East Europe; Share 4: Canada; Share 5: Central and South America; Share 6: Scandinavia; Share 7: France; Share 8: Germany; Share 9: Greece; Share 10: Ireland; Share 11: Italy; Share 12: Pacific; Share 13: Poland; Share 14: Portugal; Share 15: Spain; Share 16: Switzerland; Share 17: United Kingdom (England, Wales and Scotland); Share 18: Asia; Share 19: Others; Share 20: Afro-Americans and Share 21: White US Native Born.

Table 21: Average Main Occupation Shares: Years 1870, 1920

Year	Share 1	Share 2	Share 3	Share 4	Share 5	Share 6	Share 7
1870	.0269179	.3906677	.0314494	.0049546	.0120077	.0764527	.0723302
1920	.0479377	.3469686	.0539587	.0358102	.0316605	.087921	.0852661

Year	Share 8	Share 9	Share 10	Share 11
1870	.0451425	.0099688	.2524161	.0801304
1920	.0293113	.0256638	.1559264	.0998232

Share 1: Professionals; Share 2: Farmers; Share 3: Managers; Share 4: Clerical; Share 5: Sales Workers; Share 6: Craftsmen; Share 7: Operatives; Share 8: Pr. Household; Share 9: Service; Share 10: Farm Laborers; Share 11: Other Laborers.

Table 22: Average Main Industry Shares: Years 1870, 1920

Year	Share 1	Share 2	Share 3	Share 4	Share 5	Share 6	Share 7
1870	.7032434	.1542317	.0714427	.1141836	.0759488	.0861482	.019187
1920	.5209154	.1082717	.0499704	.1500325	.0839766	.0922905	.0267744

Year	Share 8	Share 9	Share 10	Share 11
1870	.0126641	.1089207	.0522843	.0606334
1920	.0150387	.084784	.0570265	.0375212

Share 1: Agriculture; Share 2: Mining; Share 3: Construction; Share 4: Manufacturing; Share 5: Transportation & Communication; Share 6: Trade; Share 7: Finance; Share 8: Business Service; Share 9: Service (other); Share 10: Professional Service; Share 11: Public Administration.

Graphs

Figure 1: Kernel Density Estimation: Change of Fractionalization, 1870 - 1920

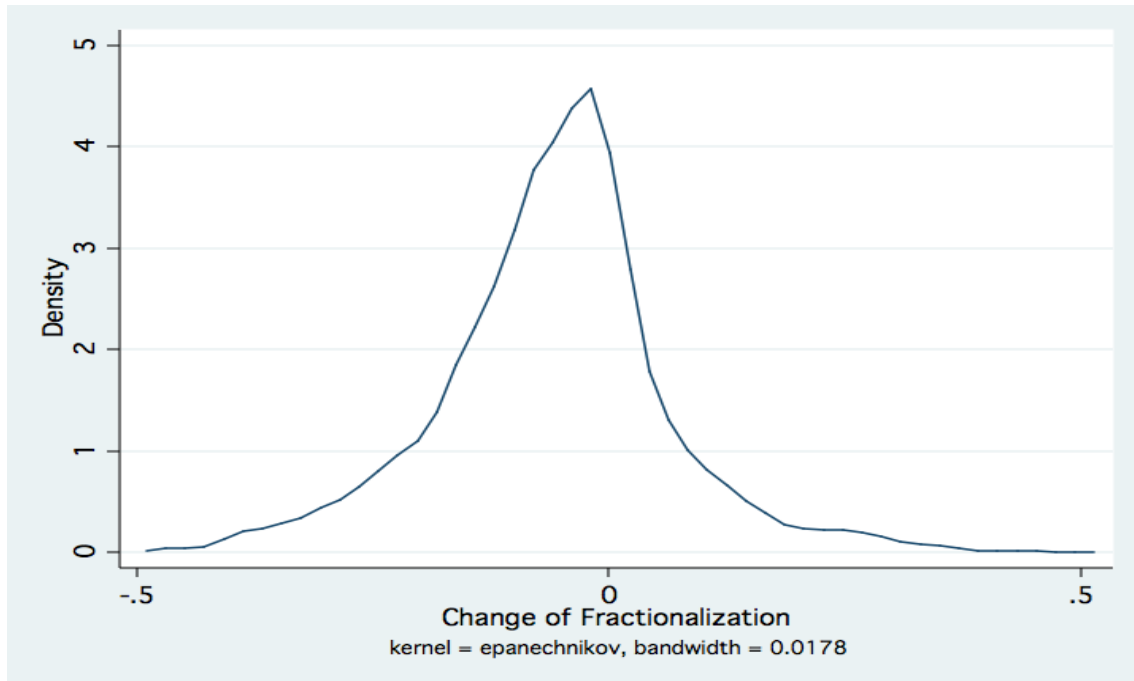
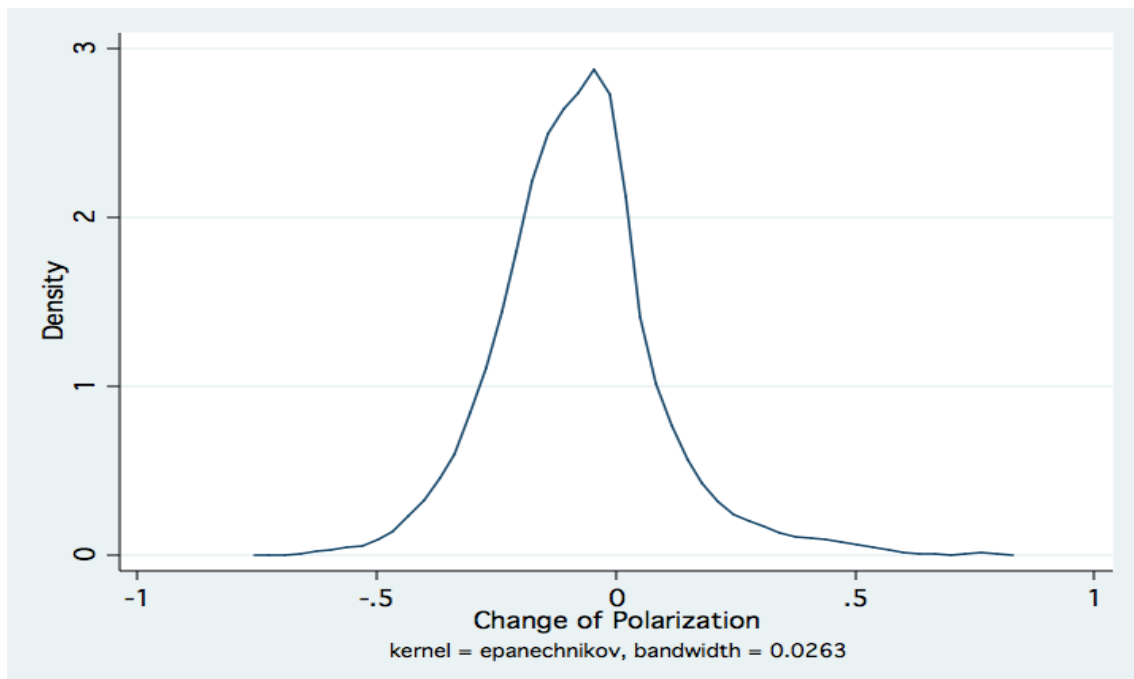


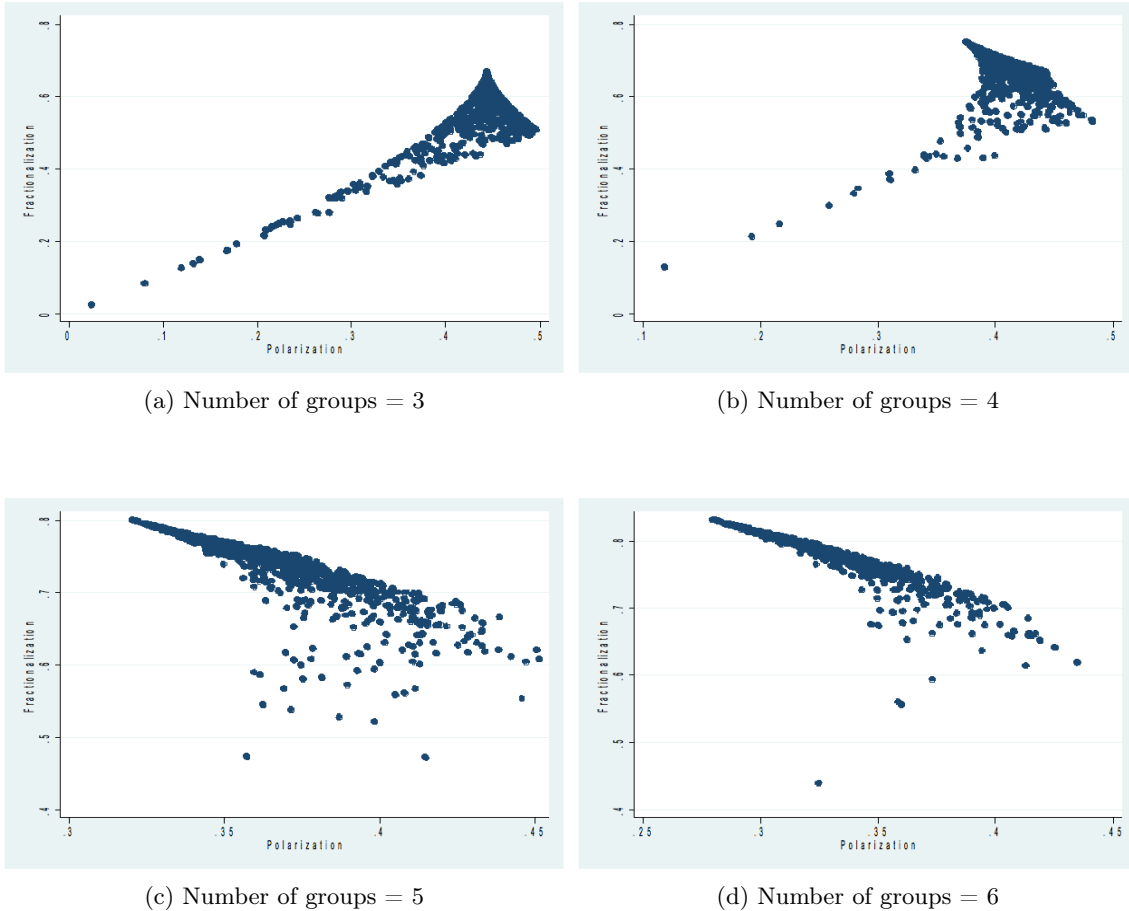
Figure 2: Kernel Density Estimation: Change of Polarization, 1870 - 1920



Simulations

To illustrate and provide some intuition for the relationship between the fractionalization and polarization index, we simulated data on the population share of n groups by drawing n times from a uniform $[0,1]$ distribution and dividing each draw by the total size of the realizations drawn. We repeated this 1000 times, constructing at each replication the fractionalization and polarization index. Appendix Figure 3 shows the results for $n = 3, 4, 5, 6$. The key message is that, when the number of groups is small the polarization and fractionalization index are positively correlated and as the number of groups increase the correlation turns negative. We do not show the results for $n = 2$ because in this case the fractionalization and polarization index are exactly the same (see Montalvo and Reynal-Querol (2005a,b)).

Figure 3: Simulation: Fractionalization (y-axis) vs. Polarization (x-axis)



Data Sources and Variable Description

Our main data source is the Inter-University Consortium for Political and Social Research (ICPSR) 2896 data file. The ICPSR 2896 data file contains detailed decennial US county and state level data on demographic, economic, and social variables which were collected by the US Bureau of Census for the period 1790-2000.¹ One key advantage of the ICPSR data set is that it enables us to exploit the underlying cultural heterogeneity in the United States at the county level. In particular, the ICPSR data set comprises – from 1870 onwards – detailed information about the country of origin of foreign-born, white US-born citizens and African-Americans which is necessary to calculate the cultural diversity indices described in Section 3.

As a further database, we use the Integrated Public Use Microdata Series (IPUMS-USA). The IPUMS is a public project and data are freely available. For more information see: <http://usa.ipums.org/usa/index.shtml>. The IPUMS gives us the possibility to exploit individual level data and construct aggregate data at the county level whenever these variables are missing in the ICPSR data set, but available at the IPUMS. We used the IPUMS database to construct the fractionalization index of occupations, the industry diversity index and the share of public sector employment (see Section 5). We use the IPUMS benchmark occupation classification variable *occ1950* and the benchmark industry classification variable *ind1950* to construct the fractionalization index of occupations and the industry diversity index, respectively.² In the supplementary online appendix we provide a detailed description of the variables used in our empirical analysis (if not further specified, variables are selected from the ICPSR 2896 data file).

¹More information about the data set (i.e. scope of study, data collection and data source) can be found at <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/02896>.

²The occupation classification code of 1950 assigns an occupation code to the individual's reported occupation and is a reference for all Census occupation data available at the IPUMS. Census years with different occupation coding schemes (as the 1870 and 1920 Census) are converted by the IPUMS into the occupation classification of 1950 (i.e. the variable *occ1950*) to make occupations over time comparable. The same applies to the IPUMS industrial classification variable *ind1950*, which contains coded information about the industry an individual worked in. For more information see IPUMS at <http://usa.ipums.org/usa/index.shtml>.

Dependent Variables

VARIABLE	YEARS	DESCRIPTION
<i>Output growth</i>	1870 - 1920	Total output in per capita terms is formed as the sum of manufacturing value added and agricultural value added. The growth variable is calculated as the change in logarithmic units.
<i>Manufacturing Value Added</i>	1870 - 1920	We use manufacturing value added in per capita terms. We calculate manufacturing value added as the difference between manufacturing output and the cost of materials used in manufacturing. For 1910, there are no manufacturing Census data available at the county level.
<i>Agricultural Value Added</i>	1870 - 1920	Agricultural value added is in per capita terms. We calculate agricultural value added as the difference between agricultural output and the cost of inputs used in agriculture. We use the variable <i>farmout</i> , which contains the estimated value of farm products, as measure for agricultural output for the years 1870 - 1900. For 1910 - 1920 we use as agricultural output the sum of values of crops, value of dairy products, value of chickens and eggs produced, value of honey and wax produced and the value of wool produced. As a proxy of the input costs in agriculture, we use expenditure for fertilizer (available 1880 - 1920) and for feed (available 1910 - 1920). See the ICPSR 2896 codebook for more details.
<i>Urban growth</i>	1870 - 1920	Change in the population share living in urban counties. The Census declared a county population as urban, if at least 2500 inhabitants lived in urban places.
<i>Population growth</i>	1870 - 1920	Change in the county population over time. The growth variable is calculated as the change in logarithmic units.

Dependent Variables (CONTINUED)

<i>Change in Occupational Diversity Index</i>	1870 - 1920	We take the occupation variable <i>occ1950</i> from the IPUMS to construct the occupational diversity index. See Section 5.2 for more details.
<i>Change in Industry Diversity Index</i>	1870 - 1920	We take the industry classification variable <i>ind1950</i> from the IPUMS to construct the industry diversity index. See Section 5.2 for more details.
<i>Change in the Tax Ratio</i>	1870 - 1920	Taxes collected by counties as a fraction of output per capita. The growth variable is calculated as the change in logarithmic units. See Rhode and Strumpf (2003) for more information.
<i>Change in Public Sector Employment</i>	1870 - 1920	We calculate public sector employment using the IPUMS industry classification variable (category: public administration) <i>ind1950</i> . Public sector employment is in per capita terms. The growth variable is calculated as the change in logarithmic units.

ADDITIONAL CONTROLS

VARIABLE	YEARS	DESCRIPTON
<i>Land Concentration</i>	1870	Gini coefficient of farm size distribution, calculated as in Galor et al. (2009, p. 175).
<i>Manufacturing Share</i>	1870	Share of manufacturing output over the sum of manufacturing and agricultural output in 1870.
<i>Labor Participation Rate</i>	1870	Share of individuals in the labor force in 1870. The labor force classification and status is taken from the IPUMS.
<i>Population</i>	1870	Total population in US counties in 1870.
<i>Urbanization</i>	1870	Population share living in urban counties. The Census declared a county population as urban, if at least 2500 inhabitants lived in urban places in 1870.
<i>Rail Access</i>	1870	Indicator variable that is equal to one if a county has access to a railroad within its borders in 1870, and zero otherwise. See Atack et al. (2008) for more information on the construction of the railroad database.
<i>Share of Native-Born White</i>	1870	Share of native-born white out of the total population in 1870.
<i>Share of Afro-Americans</i>	1870	Share of Afro-Americans (declared by the historical US Census as negro population) out of the total population in 1870.