Online Appendix

Does Equality Persist? Evidence from the Homestead Act

Table of Contents

Data Construction

Given the broad degree of geographic variation in our sample, we collect data on several key geographic covariates. These help us control for potential differences between areas where land was more heavily homesteaded versus sold for cash. We use 30-by-30-meter raster elevation data from the National Elevation Dataset (NED) to calculate the average elevation and topographic ruggedness of each county. We use 30-year "climate normals" from the PRISM climate data set to measure average annual temperature and precipitation over 1980 to 2010 for each county. We also calculate average soil productivity in each county using the Productivity Index grid developed by Schaetzl et al. (2012). We calculate the density of perennial streams in each county (measured as stream miles per acre) using data from the National Hydrography dataset. We also measure the density of railroads (from U.S. Census TIGERLine) and the presence of Indian reservations (also from the Census) in each county, as railroads and reservations affected both land grants and settlement patterns (Allen 2019). Summary statistics are reported in Appendix Table A1

Table A1: Summary Statistics

Variable	(1) N	(2) Mean	(3) Std. Dev.	(4) Min.	(5) Max.	(6) Definition
1900 Endowment Gini	1,286	0.3189	0.1857201	0	0.9312	Gini coefficient for acreage of land patents issued to individuals in a county through 1900
1935 Endowment Gini	1,341	0.3382	0.1650583	0	0.8988	Gini coefficient for acreage of land patents issued to individuals in a county through 1935
1900 Farm Size Gini	1,260	0.3691	0.0781	0.0501	0.7895	Gini coeffcient for average farm size in a county from 1900 Census of Agriculture
1935 Farm Size Gini	1,481	0.4143	0.1021	0.1602	0.8854	Gini coeffcient for average farm size in a county from 1935 Census of Agriculture
1(Beyond Frontier)	1,481	0.5496	0.4977	0	1	Dummy variable =1 if population density is less than 2 people per square mile in 1860
% Homesteads	1,481	28.273	25.852	0	100	% of all private land patents issues as homestead claims
Avg. Elevation	1,481	621.224	623.895	12.673	3329.935	Average of NED 30-meter elevation in county
Ruggedness	1,481	105.038	152.921	1.815	1180.644	Standard deviation of NED 30-meter elevation in county
Avg. Precip.	1,481	848.989	376.472	83.205	2965.638	Average of 800-meter PRISM summer rainfall over 1980-2010
Avg. Temp.	1,481	10.503	3.864	0.4149	22.933	Average of 800-meter PRISM summer temperature over 1980-2010
Share High Quality Soil	1,481	0.3939	0.3550	0	1	Share of all land in a county with Soil Productivity Index > 12
Stream Density	1,481	0.0006	0.0003	3.41e-07	0.0018	Stream miles per acre
Rail Density	1,481	0.0002	0.0001	0	0.0011	Rail miles per acre
Reservation Indicator	1,481	0.2566	0.4369	0	1	Dummy variable =1 if a Native American reservation overlaps a county

Notes: This table reports summary statistics for the main data used throughout the paper. Sources are as follows. Land endowment Ginis: authors' calculations based on records digitized by the General Land Office and made available by the Bureau of Land Management. Farm size Ginis: author's calculations following Galor et al. (2009) using "farms by size" reported in the U.S. Census of Agriculture. Frontier dummy: authors' calculations using data on population density and decadal frontier lines published by Bazzi et al. (2020). Avg. elevation and ruggedness: authors' calculations using 30x30-meter elevation raster data from the National Elevation Dataset. Avg. temperature and precipitation: authors' calculations using 30-year "climate normals" from the PRISM climate data set. Soil quality: authors' calculations using the Soil Productivity Index (PI) developed by Schaetzl et al. (2012). Steam density: authors' calculations using perennial streams extracted from the National Hydrography Dataset. Rail density: authors' calculations using railroads from the U.S. Census TIGERLine database. Reservation indicator: authors' calculations using U.S. Census boundary files delineating Native American reservations.

Table A2: Panel Summary Statistics

Variable	(1) N	(2) Mean	(3) Std. Dev.	(4) Min.	(5) Max.	(6) Definition
ln(Gini)	4,673	-1.025382	.2279042	-2.380503	-.143832	Log of farm size gini in year *t*
% Homesteads	4,673	.2892353	.3337617	0	8.403911	Cumulative % of all land patents issued as homestead by year *t*
% △ ln(Gini)	4,673	.0237748	.140355	-1.666158	1.759156	Change in the log of farm size gini from year *t-1* to *t*
% △ Homesteads	4,673	5.18779	4.472693	0	14.72853	% Change number of homestead claims from year *t-1* to *t*
% △ All Land Claims	4,673	6.602698	3.518109	0	14.33368	% Change number of all claims from year *t-1* to *t*

Notes: This table reports summary statistics for the panel data used in Table 2. Sources are as follows. Farm size Ginis: author's calculations following Galor et al. (2009) using "farms by size" reported in the U.S. Census of Agriculture.

Table A3: Full Results for Covariates in Table 1

	(1)	(2)	(3)	(4)	(5)	(6)
	Y= ln(1935 Land Endowment Gini)			Y= ln(1935 Farm Size Gini)		
% Homesteads	-0.00548***	-0.00415***	-0.00347***	-0.00147***	-0.000888***	-0.000607***
	(0.000746)	(0.000742)	(0.000750)	(0.000245)	(0.000241)	(0.000233)
Avg. Elevation		0.0000987			-0.0000139	
		(0.0000612)			(0.0000303)	
Ruggedness		0.0000194			0.000349***	
		(0.000132)			(0.0000838)	
Avg. Precip.		-0.0000441			0.0000289	
		(0.0000482)			(0.0000290)	
Avg. Temp.		-0.0217**			0.0146***	
		(0.00885)			(0.00376)	
% High-Quality Soil		-0.192***			-0.0544***	
		(0.0408)			(0.0165)	
Stream Density		194.6***			54.08***	
		(40.57)			(20.20)	
Rail Density		359.0***			465.7***	
		(83.09)			(44.24)	
Reservation Indicator		0.0866***	0.0301		0.0385***	0.0387***
		(0.0253)	(0.0236)		(0.0129)	(0.0127)
Observations	1,416	1,416	1,416	1,481	1,481	1,481
Adjusted R-squared	0.938	0.943	0.949	0.970	0.974	0.976
Mean Gini	0.349	0.349	0.349	0.344	0.344	0.344
State FE	✓	✓	✓	✓	✓	✓
Arrival Decade FE	✓	✓	✓	✓	✓	✓
Linear Controls		✓			✓	
Binned Controls (by decile)			✓			✓

Notes: This table presents the estimates of the relationship between the percentage of land patents in a county that were issued as homesteads and historic land endowment inequality corresponding to those in Table 3, with the coefficient estimates for all control variables included here. Columns 1 through 3 use patented acres to estimate Gini coefficients for each county as of 1935. Columns 4-6 construct Gini coefficients for Average Farm Size based on the 1935 U.S. Agricultural Census. The binned specifications in Columns 3 and 6 include fixed effects for each decile of elevation, ruggedness, precipitation, temperature, soil quality, stream density, and rail density, effectively comparing counties with very similar resource endowments and initial conditions. Spatial HAC standard errors following Conley (2008) and Hsiang (2010) are reported in paratheses. We use a uniform kernel density and a 150 km cutoff when estimating the spatially correlated standard errors. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table A4: Homesteading along the 1860 Frontier, Omitting Western States

	(1)	(2)	(3)	(4)	(5)	(6)
			Y= % Homesteads			
1(Beyond 1860 Frontier)	12.14***	11.19***	9.377***	8.678***	9.547***	8.600*
	(2.788)	(2.909)	(2.822)	(2.936)	(2.872)	(4.251)
Observations	510	412	323	267	226	100
Adjusted R-squared	0.822	0.805	0.806	0.813	0.820	0.851
Mean Dep. Var.	31.94	30.46	28.91	27.84	26.72	28.88
Distance to Frontier	*200km*	*150km*	*100km*	*75km*	*50km*	*0 km*
State FE	✓	✓	✓	✓	✓	✓
Binned Controls (by decile)	✓	✓	✓	✓	✓	✓

Notes: This table presents the estimates of the difference in the percentage of land patents in a county that were issued as homesteads for counties that were just beyond the frontier of settlement in 1860 (based on Bazzi et al. 2020 and Turner. This table omits counties along smaller frontiers in Western states, focusing on the main frontier of settlement in the Midwest depicted in Figures A2 through A7. All columns include state fixed effects fixed effects for each decile of elevation, ruggedness, precipitation, temperature, soil quality, stream density, and rail density, effectively comparing counties with very similar resource endowments and initial conditions. Moving from left to right, each column includes progressively smaller samples of counties that are closer to the 1860 frontier, beginning with a 200-kilometer cutoff in column 1 and ending with directly adjacent counties in column 6. We omit decade fixed effects from these models because—by construction—the samples consistent of counties that were settled just before vs. just after 1860 only. Spatial HAC standard errors following Conley (2008) and Hsiang (2010) are reported in parentheses. We use a uniform kernel density and a 150 km cutoff when estimating the spatially correlated standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Homesteading and Historical Land Inequality along the 1860 Frontier, Omitting Western States

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:						
	Y= ln(1900 Land Endowment Gini)					
% Homesteads	-0.00432***	-0.00721***	-0.00679***	-0.00582***	-0.00577***	-0.00573***
	(0.00141)	(0.00132)	(0.00133)	(0.00123)	(0.00127)	(0.00162)
Observations	483	386	301	248	213	95
Adjusted R-squared	0.942	0.943	0.947	0.956	0.956	0.970
Mean Gini	0.306	0.321	0.328	0.327	0.326	0.317
Panel B:						
	Y= ln(1900 Farm Size Gini)					
% Homesteads	-0.000712**	-0.00106***	-0.00123***	-0.000928**	-0.00104**	-0.00182***
	(0.000318)	(0.000316)	(0.000368)	(0.000392)	(0.000452)	(0.000640)
Observations	449	370	293	245	212	94
Adjusted R-squared	0.991	0.993	0.993	0.993	0.992	0.995
Mean Gini	0.342	0.340	0.340	0.340	0.341	0.357
Distance to Frontier	*200km*	*150km*	*100km*	*75km*	*50km*	*0 km*
State FE	✓	✓	✓	✓	✓	✓
Binned Controls (by decile)	✓	✓	✓	✓	✓	✓

Notes: This table presents the estimates of the relationship between the percentage of land patents in a county that were issued as homesteads and historic land endowment inequality in progressively smaller subsets of counties near the frontier of settlement in 1860 (based on Bazzi et al. 2020 and Turner). This table omits counties along smaller frontiers in Western states, focusing on the main frontier of settlement in the Midwest depicted in Figures A2 to A7. All columns include state fixed effects fixed effects for each decile of elevation, ruggedness, precipitation, temperature, soil quality, stream density, and rail density, effectively comparing counties with very similar resource endowments and initial conditions. Moving from left to right, each column includes progressively smaller samples of counties that are closer to the 1860 frontier, beginning with a 200-kilometer cutoff in column 1 and ending with directly adjacent counties in column 6. We omit decade fixed effects from these models because—by construction—the samples consistent of counties that were settled just before vs. just after 1860 only. Panel A uses the natural log of the 1900 land endowment Gini as the dependent variable and panel B uses the natural log of the 1900 farm size Gini as the dependent variable. Spatial HAC standard errors following Conley (2008) and Hsiang (2010) are reported in parentheses. We use a uniform kernel density and a 150 km cutoff when estimating the spatially correlated standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Homesteading along the 1860 Frontier, Omitting Oklahoma

	(1)	(2)	(3)	(4)	(5)	(6)
	Y= % Homesteads					
1(Beyond 1860 Frontier)	9.064***	8.786***	8.659***	8.514***	9.038***	8.995**
	(2.554)	(2.538)	(2.604)	(2.644)	(2.791)	(4.363)
Observations	465	379	299	247	213	95
Adjusted R-squared	0.852	0.856	0.858	0.855	0.858	0.840
Mean Dep. Var.	29.73	29.14	28.09	26.87	26.49	28.97
Distance to Frontier	*200km*	*150km*	*100km*	*75km*	*50km*	*0 km*
State FE	✓	✓	✓	✓	✓	✓
Binned Controls (by decile)	✓	✓	✓	✓	✓	✓

Notes: This table presents the estimates of the difference in the percentage of land patents in a county that were issued as homesteads for counties that were just beyond the frontier of settlement in 1860 (based on Bazzi et al. 2020 and Turner. This table omits counties in Oklahoma and counties along smaller frontiers in Western states, focusing on the main frontier of settlement in the Midwest depicted in Figures A2 through A7. All columns include state fixed effects fixed effects for each decile of elevation, ruggedness, precipitation, temperature, soil quality, stream density, and rail density, effectively comparing counties with very similar resource endowments and initial conditions. Moving from left to right, each column includes progressively smaller samples of counties that are closer to the 1860 frontier, beginning with a 200-kilometer cutoff in column 1 and ending with directly adjacent counties in column 6. We omit decade fixed effects from these models because—by construction—the samples consistent of counties that were settled just before vs. just after 1860 only. Spatial HAC standard errors following Conley (2008) and Hsiang (2010) are reported in parentheses. We use a uniform kernel density and a 150 km cutoff when estimating the spatially correlated standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Homesteading and Historical Land Inequality along the 1860 Frontier, Omitting Oklahoma

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:						
	Y= ln(1900 Land Endowment Gini)					
% Homesteads	-0.00634***	-0.00694***	-0.00658***	-0.00582***	-0.00577***	-0.00573***
	(0.00123)	(0.00130)	(0.00134)	(0.00123)	(0.00127)	(0.00162)
Observations	464	379	299	247	213	95
Adjusted R-squared	0.947	0.944	0.948	0.956	0.956	0.970
Mean Gini	0.315	0.324	0.328	0.326	0.326	0.317
Panel B:						
	Y= ln(1900 Farm Size Gini)					
% Homesteads	-0.000712**	-0.00106***	-0.00123***	-0.000928**	-0.00104**	-0.00182***
	(0.000318)	(0.000316)	(0.000368)	(0.000392)	(0.000452)	(0.000640)
Observations	449	370	293	245	212	94
Adjusted R-squared	0.991	0.993	0.993	0.993	0.992	0.995
Mean Gini	0.342	0.340	0.340	0.340	0.341	0.357
Distance to Frontier	*200km*	*150km*	*100km*	*75km*	*50km*	*0 km*
State FE	✓	✓	✓	✓	✓	✓
Binned Controls (by decile)	✓	✓	✓	✓	✓	✓

Notes: This table presents the estimates of the relationship between the percentage of land patents in a county that were issued as homesteads and historic land endowment inequality in progressively smaller subsets of counties near the frontier of settlement in 1860 (based on Bazzi et al. 2020 and Turner). This table omits counties in Oklahoma and counties along smaller frontiers in Western states, focusing on the main frontier of settlement in the Midwest depicted in Figures A2 to A7. All columns include state fixed effects fixed effects for each decile of elevation, ruggedness, precipitation, temperature, soil quality, stream density, and rail density, effectively comparing counties with very similar resource endowments and initial conditions. Moving from left to right, each column includes progressively smaller samples of counties that are closer to the 1860 frontier, beginning with a 200-kilometer cutoff in column 1 and ending with directly adjacent counties in column 6. We omit decade fixed effects from these models because—by construction—the samples consistent of counties that were settled just before vs. just after 1860 only. Panel A uses the natural log of the 1900 land endowment Gini as the dependent variable and panel B uses the natural log of the 1900 farm size Gini as the dependent variable. Spatial HAC standard errors following Conley (2008) and Hsiang (2010) are reported in parentheses. We use a uniform kernel density and a 150 km cutoff when estimating the spatially correlated standard errors. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table A8: Homesteading along the 1860 Frontier, Omitting Reservation Counites

	(1)	(2)	(3)	(4)	(5)	(6)
	Y= % Homesteads					
1(Beyond 1860 Frontier)	11.14***	11.34***	10.83***	10.34***	11.51***	10.86**
	(3.118)	(3.028)	(3.037)	(3.092)	(3.072)	(4.169)
Observations	382	310	241	194	169	74
Adjusted R-squared	0.853	0.863	0.874	0.877	0.864	0.840
Mean Dep. Var.	28.66	28.66	27.70	26.44	25.28	26.99
Distance to Frontier	*200km*	*150km*	*100km*	*75km*	*50km*	*0 km*
State FE	✓	✓	✓	✓	✓	✓
Binned Controls (by decile)	✓	✓	✓	✓	✓	✓

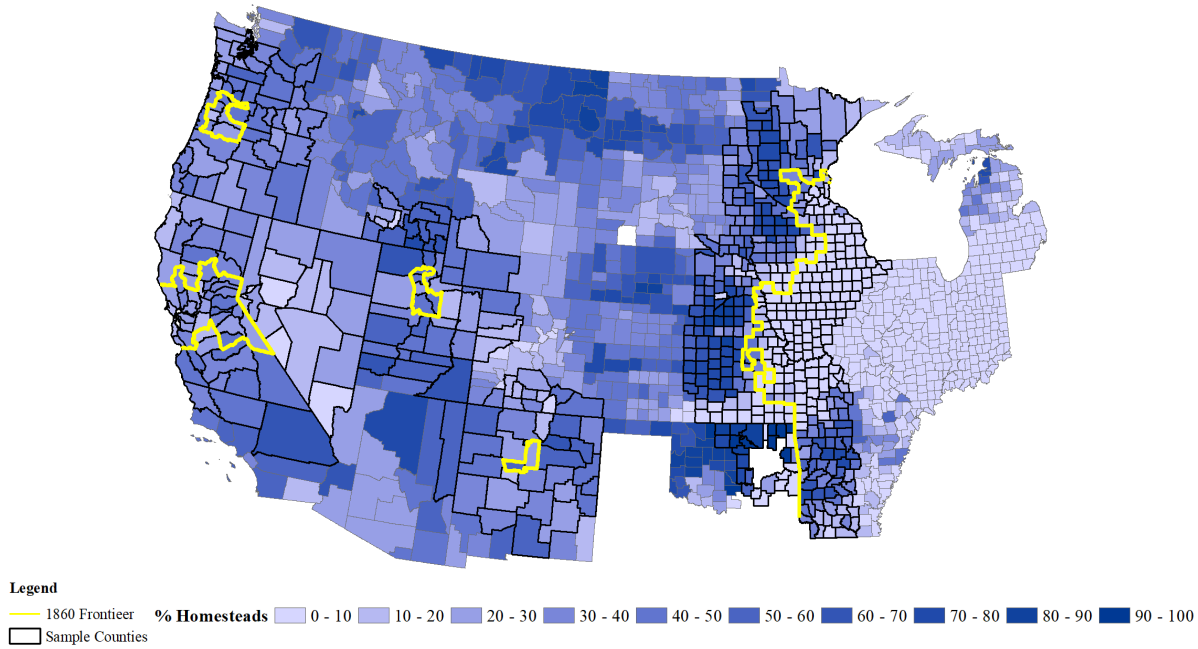
Notes: This table presents the estimates of the difference in the percentage of land patents in a county that were issued as homesteads for counties that were just beyond the frontier of settlement in 1860 (based on Bazzi et al. 2020 and Turner. This table omits counties that overlap with Native American Reservations and counties along smaller frontiers in Western states, focusing on the main frontier of settlement in the Midwest depicted in Figures A2 through A7. All columns include state fixed effects fixed effects for each decile of elevation, ruggedness, precipitation, temperature, soil quality, stream density, and rail density, effectively comparing counties with very similar resource endowments and initial conditions. Moving from left to right, each column includes progressively smaller samples of counties that are closer to the 1860 frontier, beginning with a 200-kilometer cutoff in column 1 and ending with directly adjacent counties in column 6. We omit decade fixed effects from these models because—by construction—the samples consistent of counties that were settled just before vs. just after 1860 only. Spatial HAC standard errors following Conley (2008) and Hsiang (2010) are reported in parentheses. We use a uniform kernel density and a 150 km cutoff when estimating the spatially correlated standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Homesteading and Historical Land Inequality along the 1860 Frontier, Omitting Reservation Counties

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:						
	Y= ln(1900 Land Endowment Gini)					
% Homesteads	-0.00512***	-0.00709***	-0.00665***	-0.00602***	-0.00653***	-0.00789***
	(0.00136)	(0.00152)	(0.00159)	(0.00147)	(0.00156)	(0.00170)
Observations	382	310	241	194	169	74
Adjusted R-squared	0.950	0.944	0.946	0.958	0.957	0.974
Mean Gini	0.322	0.328	0.331	0.328	0.327	0.309
Panel B:						
	Y= ln(1900 Farm Size Gini)					
% Homesteads	-0.000942***	-0.000921***	-0.000790**	-0.000301	-0.0000928	-0.00133**
	(0.000298)	(0.000336)	(0.000386)	(0.000449)	(0.000474)	(0.000600)
Observations	370	303	237	193	169	74
Adjusted R-squared	0.993	0.994	0.994	0.993	0.992	0.995
Mean Gini	0.341	0.339	0.338	0.338	0.339	0.350
Distance to Frontier	*200km*	*150km*	*100km*	*75km*	*50km*	*0 km*
State FE	✓	✓	✓	✓	✓	✓
Binned Controls (by decile)	✓	✓	✓	✓	✓	✓

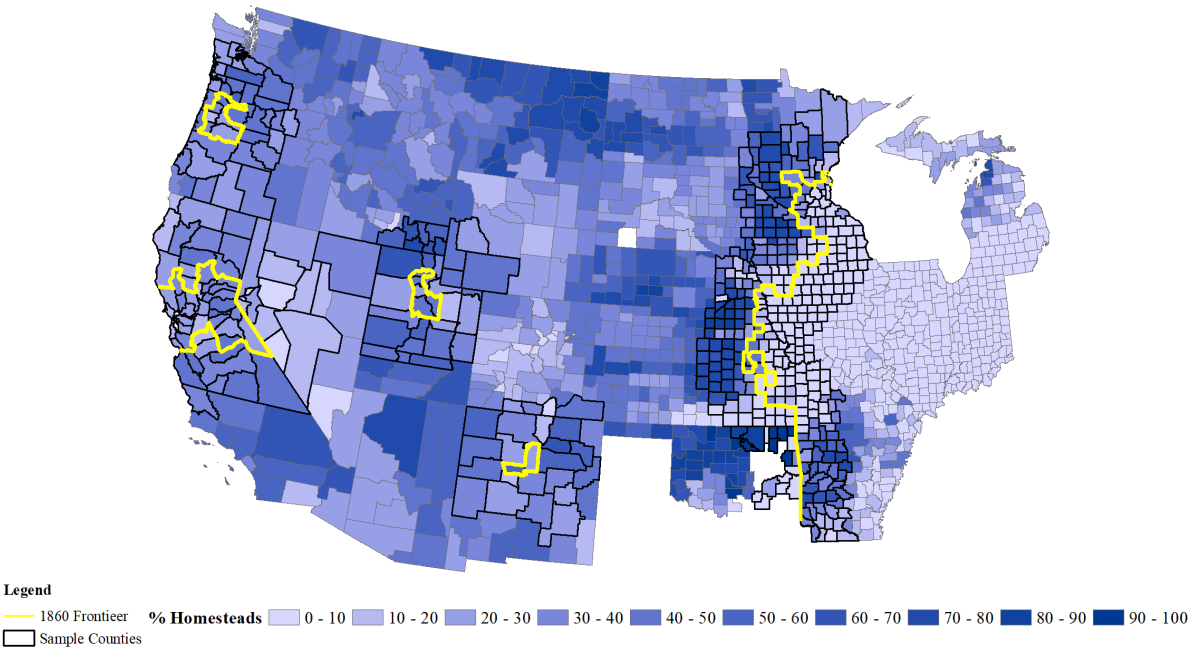
Notes: This table presents the estimates of the relationship between the percentage of land patents in a county that were issued as homesteads and historic land endowment inequality in progressively smaller subsets of counties near the frontier of settlement in 1860 (based on Bazzi et al. 2020 and Turner). This table omits counties that overlap with Native American Reservations and counties along smaller frontiers in Western states, focusing on the main frontier of settlement in the Midwest depicted in Figures A2 to A7. All columns include state fixed effects fixed effects for each decile of elevation, ruggedness, precipitation, temperature, soil quality, stream density, and rail density, effectively comparing counties with very similar resource endowments and initial conditions. Moving from left to right, each column includes progressively smaller samples of counties that are closer to the 1860 frontier, beginning with a 200-kilometer cutoff in column 1 and ending with directly adjacent counties in column 6. We omit decade fixed effects from these models because—by construction—the samples consistent of counties that were settled just before vs. just after 1860 only. Panel A uses the natural log of the 1900 land endowment Gini as the dependent variable and panel B uses the natural log of the 1900 farm size Gini as the dependent variable. Spatial HAC standard errors following Conley (2008) and Hsiang (2010) are reported in parentheses. We use a uniform kernel density and a 150 km cutoff when estimating the spatially correlated standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure A1: Counties within 200 Kilometers of 1860 Frontier



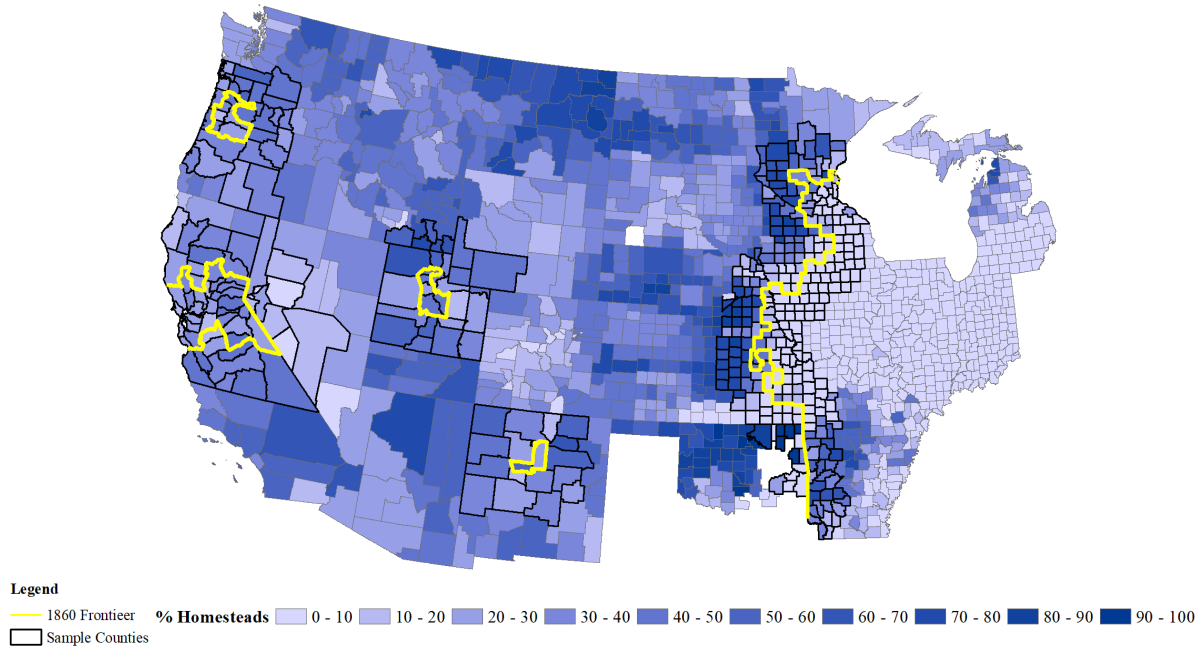
Notes: This figure depicts Bazzi et al.'s (2020) estimate of the frontier of settlement in 1860 in yellow. Beyond this line, population density fell below two people per square mile. Counties within 200 kilometers of the 1860 frontier are outlined in bold. All counties are shaded according to the percentage of patents issues as homesteads, as in panel (a) of Figure 1 (darker shading indicates a higher percentage of homesteading).

Figure A2: Counties within 150 Kilometers of 1860 Frontier



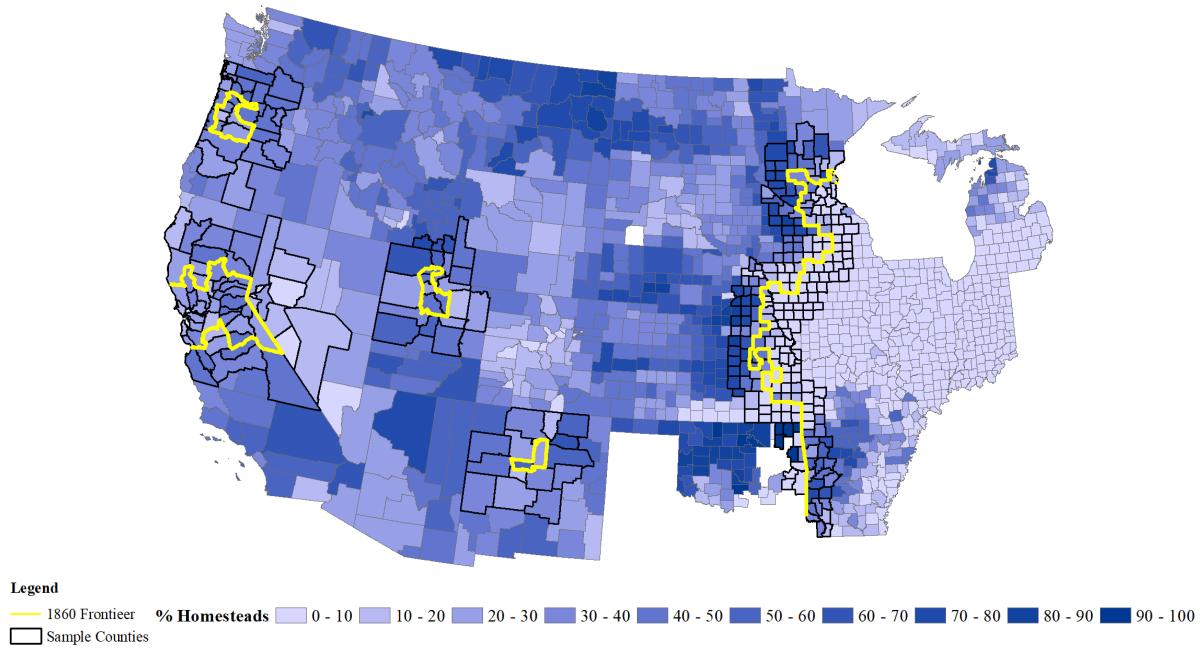
Notes: This figure depicts Bazzi et al.'s (2020) estimate of the frontier of settlement in 1860 in yellow. Beyond this line, population density fell below two people per square mile. Counties within 150 kilometers of the 1860 frontier are outlined in bold. All counties are shaded according to the percentage of patents issues as homesteads, as in panel (a) of Figure 1 (darker shading indicates a higher percentage of homesteading).

Figure A3: Counties within 100 Kilometers of 1860 Frontier



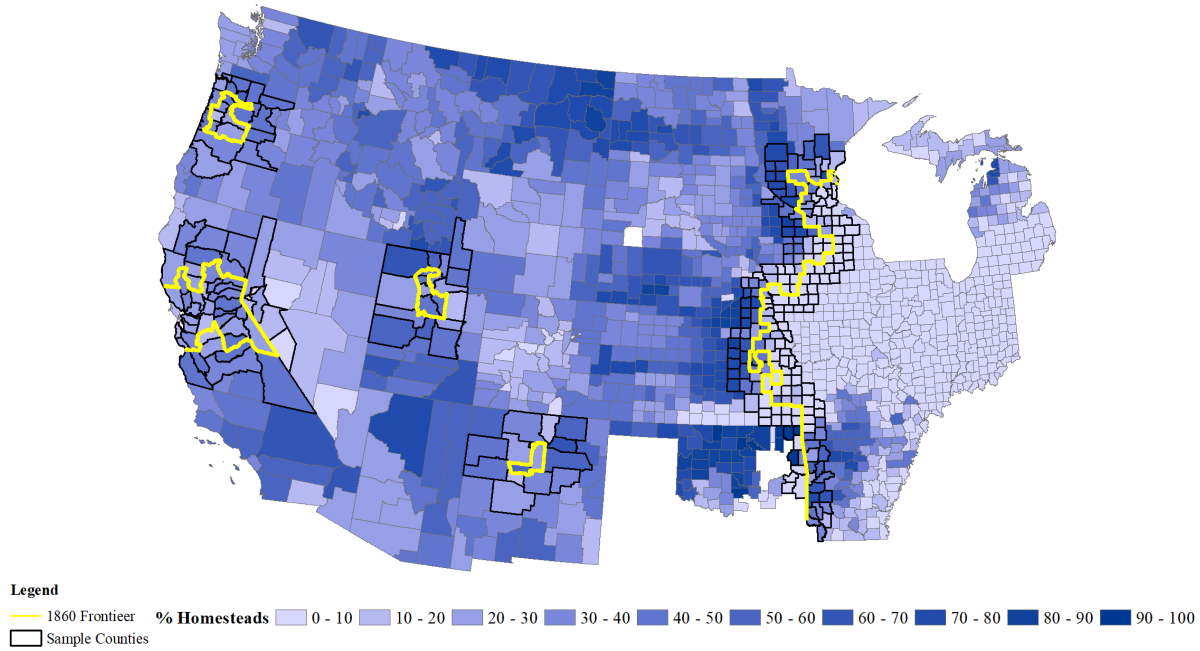
Notes: This figure depicts Bazzi et al.'s (2020) estimate of the frontier of settlement in 1860 in yellow. Beyond this line, population density fell below two people per square mile. Counties within 100 kilometers of the 1860 frontier are outlined in bold. All counties are shaded according to the percentage of patents issues as homesteads, as in panel (a) of Figure 1 (darker shading indicates a higher percentage of homesteading).

Figure A4: Counties within 75 Kilometers of 1860 Frontier



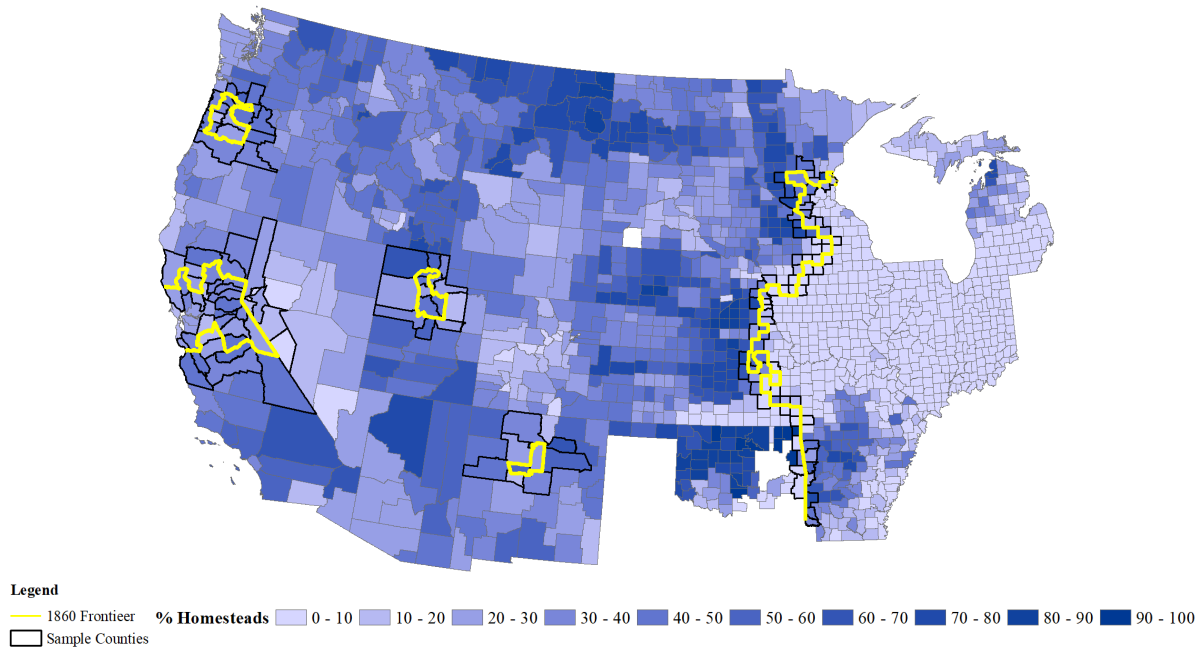
Notes: This figure depicts Bazzi et al.'s (2020) estimate of the frontier of settlement in 1860 in yellow. Beyond this line, population density fell below two people per square mile. Counties within 75 kilometers of the 1860 frontier are outlined in bold. All counties are shaded according to the percentage of patents issues as homesteads, as in panel (a) of Figure 1 (darker shading indicates a higher percentage of homesteading).

Figure A5: Counties within 50 Kilometers of 1860 Frontier



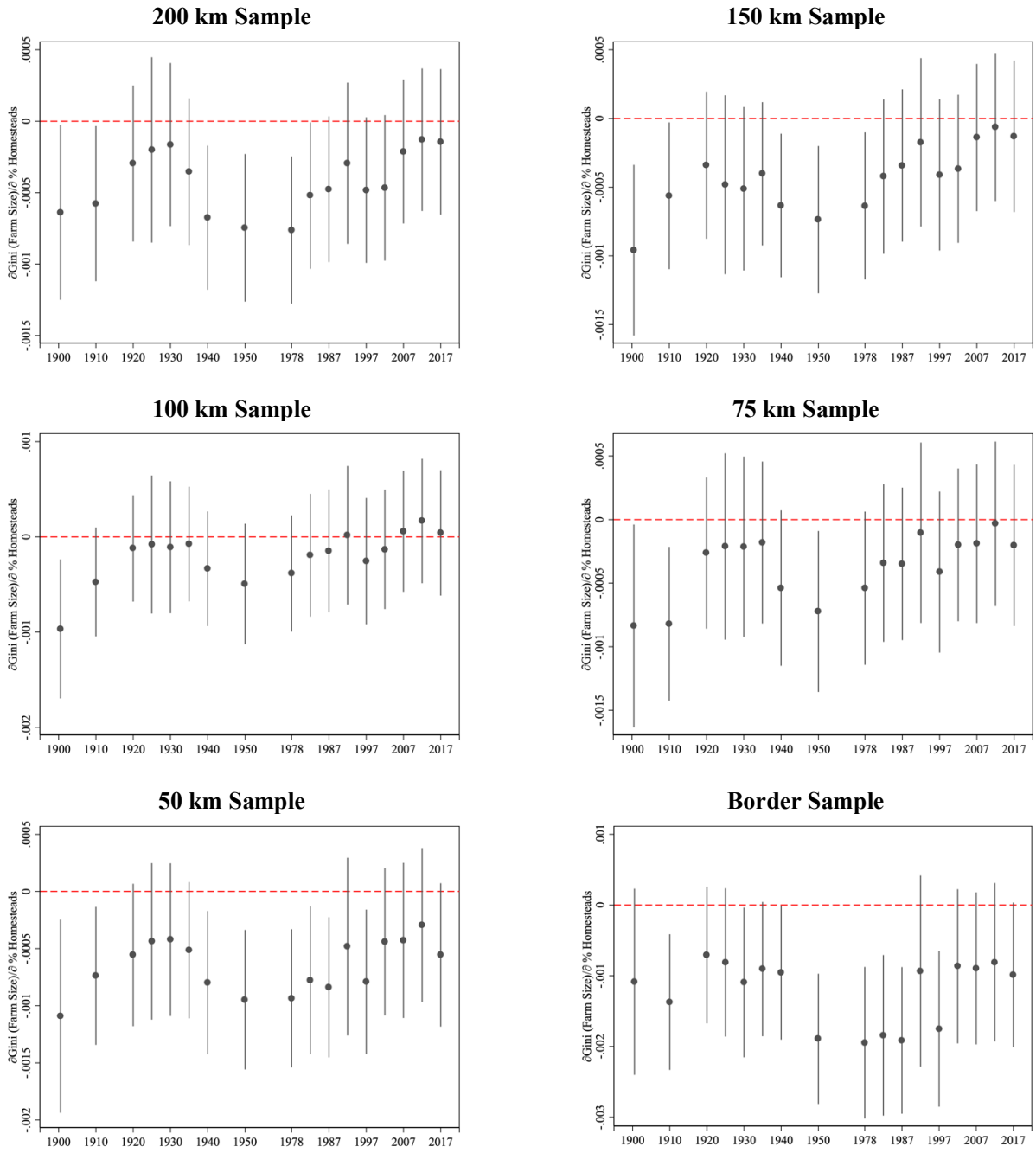
Notes: This figure depicts Bazzi et al.'s (2020) estimate of the frontier of settlement in 1860 in yellow. Beyond this line, population density fell below two people per square mile. Counties within 50 kilometers of the 1860 frontier are outlined in bold. All counties are shaded according to the percentage of patents issues as homesteads, as in panel (a) of Figure 1 (darker shading indicates a higher percentage of homesteading).

Figure A6: Counties Adjacent to 1860 Frontier



Notes: This figure depicts Bazzi et al.'s (2020) estimate of the frontier of settlement in 1860 in yellow. Beyond this line, population density fell below two people per square mile. Counties adjacent to the 1860 frontier are outlined in bold. All counties are shaded according to the percentage of patents issues as homesteads, as in panel (a) of Figure 1 (darker shading indicates a higher percentage of homesteading).

Figure A7: Long-Run Effect of Homesteading in Frontier Samples



Notes: This figure depicts alternative versions of Figure 2 the utilize the restricted, frontier-adjacent samples in Columns 1-6 of Table 4 (depicted in Figures A2 to A7) to estimate the effect of homesteading on farm size inequality over time, utilizing the same specification from Table 6.