

Appendix for “Publishing Nations: Technology Acquisition and Language Standardization for European Ethnic Groups”

July 18, 2017

Contents

1	List of 171 Ethnic Groups in the Data Set	A2
2	List of 47 European States in the Data Set	A7
3	Regression Tables	A9
3.1	Cox proportional hazards and logistic regression models used in paper	A9
3.2	Robustness checks I: state as the unit of analysis	A11
3.2.1	State-only sample	A11
3.2.2	Stateless ethnic groups-only sample	A13
3.3	Robustness checks II: The role of human capital	A15
3.4	Instrumental variables (IV) approach: Distance from Mainz as IV	A17

1 List of 171 Ethnic Groups in the Data Set

Table A1

	name	homeland city	printing press	dictionary
1	Abaza	Adyge-Khabl		
2	Abkhaz	Sukhumi		1986
3	Adyge	Maykop		
4	Aguls	Tpig, Dagestan		
5	Ajars	Batumi		
6	Akhvakh	Akhvakhsky district, Dagestan		
7	Alanders	Mariehamn		
8	Albanians	Tirana	1555	1980
9	Alsations	Strasbourg	1460	1899
10	Andalusians	Seville	1477	1941
11	Andis	Khasavyurt, Dagestan		
12	Andorrans	Andorra La Vella		
13	Archis	Arsha-Makhi, Dagestan		1977
14	Armenians	Yerevan	1512	1971
15	Aromanians	Trikala		1963
16	Austrians	Vienna	1461	1963
17	Avars	Kunzakh, Khunzakhsky district, Dagestan		
18	Azeris	Baku	1870	1966
19	Azoreans	Ponta Delgada		
20	Bagulals	Tsumadinsky district		
21	Balkars	Karachayevsky district		1996
22	Bashkorts	Ufa		1993
23	Basques	Vitoria	1545	1989
24	Bats	Tusheti		2003
25	Bavarians	Munich	1482	1827
26	Belarussians	Minsk	1520	1977
27	Bezhetas	Bezhta village, Tsuntinsky district, Dagestan		
28	Bosnians	Sarajevo	1866	1992
29	Botligs	Botlikh village, Botlikhsky district, Dagestan		
30	Bretons	Rennes	1485	1958
31	Budugs	Budug		
32	Bulgarians	Sofia	1566	1895
33	Burgundians	Dijon	1491	
34	Canarians	Santa Cruz de Tenerife		1995
35	Carpatho-Rusyns	Uzhgorod	1572	

continued on next page

continued from previous page

	name	homeland city	printing press	dictionary
36	Catalans	Barcelona	1475	1930
37	Chamalals	Agvali village, Tsumadinsky district, Dagestan		
38	Chavash	Cheboksary		1928
39	Chechens	Grozny		1961
40	Cherkess	Cherkessk		1956
41	Cornish	Truno	1700	
42	Corsicans	Ajaccio	1769	
43	Crimean Tatars	Bakhchysarai		
44	Croats	Zagreb	1483	1901
45	Csangos	Miercurea Ciuc	1533	1936
46	Czechs	Prague	1487	1935
47	Danes	Copenhagen	1490	1918
48	Dargins	Makhachkala, Dagestan		2005
49	Didos	Kidero village, Tsuntinsky district, Dagestan		
50	Don Cossacks	Novocherkassk		1975
51	Dutch	Amsterdam	1477	1864
52	English	London	1476	1755
53	Estonians	Tallinn	1535	1918
54	Faroese	Torshavn		1966
55	Finns	Helsinki	1542	1953
56	Flemish	Brussels	1475	1873
57	French	Paris	1470	1858
58	Frisians	Leeuwarden	1630	1874
59	Friulis	Udine	1484	1985
60	Gagauz	Comrat		
61	Galicians	Santiago de Compostela	1483	1986
62	Georgians	Tbilisi	1709	1884
63	Germans	Berlin	1540	1852
64	Gibraltarians	Gibraltar	1878	
65	Ginugs	Tsuntinsky district, Dagestan		
66	Godoberis	Godoberi village, Botlikhsky district, Dagestan		
67	Greeks	Athens	1759	1936
68	Guernseians	St. Peter Port		
69	Hungarians	Budapest	1473	1890
70	Icelanders	Reykjavik	1540	1912
71	Ingrians	Saint Petersburg	1711	1953
72	Ingush	Nazran		
73	Irish	Dublin	1550	1913
74	Istrians	Pulj	1905	1979
75	Italians	Rome	1467	1861

continued on next page

continued from previous page

	name	homeland city	printing press	dictionary
76	Jerseyites	St. Helier		
77	Kabards	Nalchik		
78	Kalmyks	Elista		
79	Karachais	Mikoyan Shakhar		1996
80	Karatas	Karata village, Akhvakhsky district, Dagestan		
81	Karels	Petrozavodsk	1930	1968
82	Kashubians	Gdansk	1499	1994
83	Khinalugs	Quba		
84	Khunzuls	Nakhada village, Tsuntinsky district, Dagestan		
85	Khwarshia	Khvarshi village, Tsumadinsky district, Dagestan		
86	Kists	Kakheti		
87	Komis	Syktyvkar		1959
88	Kosovars	Pristina		
89	Kryts	Jek		
90	Kuban Cossacks	Krasnodar		
91	Kумыks	Buynaksk, Dagestan		
92	Ladins	Belluno	1475	
93	Laks	Vachi, Dagestan		
94	Latvians	Riga	1525	1972
95	Lezgins	Mamash, Dagestan		
96	Liechtensteiners	Vaduz		
97	Ligurians	Genoa	1472	1985
98	Lithuanians	Vilnius	1525	1956
99	Livonians	Ventspils	1542	1938
100	Lombards	Milan	1471	1835
101	Luxembourgers	Luxembourg City	1598	1950
102	Macedonians	Skopje	1840	1961
103	Madeirans	Funchal		
104	Maltese	Valletta	1642	1987
105	Manx	Douglas	1707	1835
106	Maris	Yoshkar-Ola	1767	1990
107	Megleno-Romanians	Kilkis	1512	1963
108	Meskhtekians	Akhalsikhe		
109	Moldovans	Chisinau	1642	1977
110	Monegasques	Monaco-Ville	1795	

continued on next page

continued from previous page

	name	homeland city	printing press	dictionary
111	Montenegrins	Podgorica	1494	
112	Moravians	Brno	1486	1982
113	Mordvins	Saransk		1973
114	Mountain Jews	Majalis, Dagestan		
115	Nenets	Naryan-Mar		
116	Nogais	Kizlyar, Dagestan		
117	Normans	Rouen	1485	
118	Northumbrians	Newcastle	1646	
119	Norwegians	Oslo	1643	1930
120	Occitans	Toulouse	1476	1875
121	Ossetians	Vladikavkaz		1958
122	Piedmontese	Turin	1474	1983
123	Poles	Warsaw	1522	1807
124	Pomaks	Edirne	1554	
125	Portuguese	Lisbon	1489	1852
126	Rhinelanders	Cologne	1464	1928
127	Romands	Geneva	1478	1926
128	Romanians	Bucharest	1708	1907
129	Romansh	Chur	1557	1939
130	Romas	Soroca		
131	Russians	Moscow	1564	1891
132	Rutuls	Rutul, Dagestan		
133	Samis	Kautokeino	1649	1939
134	Sanjakis	Novi Pazar		
135	San Marinese	San Marino		
136	Sards	Cagliari	1493	
137	Savoyards	Chambéry	1484	1969
138	Saxons	Dresden	1524	1953
139	Scanians	Malmo	1528	
140	Scots	Edinburgh	1508	1931
141	Serbs	Belgrade	1552	1818
142	Sicilians	Palermo	1478	1977
143	Silesians	Wroclaw	1475	1962
144	Slovaks	Bratislava	1500	1959
145	Slovenes	Ljubljana	1575	1976
146	Sorbs	Bautzen	1555	1978
147	Spaniards	Madrid	1566	1886
148	Swabians	Stuttgart	1486	1901
149	Swedes	Stockholm	1483	1893
150	Swiss-Germans	Zürich	1479	1881

continued on next page

continued from previous page

	name	homeland city	printing press	dictionary
151	Szeklars	Cluj-Napoca	1578	1975
152	Tabasarans	Khiv, Dagestan		
153	Talysh	Astara		
154	Tats	Shabran		
155	Tatars	Kazan	1804	1969
156	Terek Cossacks	Stavropol		
157	Tindis	Tindi village, Tsumadinsky district, Dagestan		
158	Tsakhurs	Balakan, Dagestan		
159	Turks	Ankara	1726	1963
160	Tyroleans	Innsbruck	1547	1866
161	Udis	Qabala district		
162	Udmurts	Izhhevsk		1890
163	Ukrainians	Kiev	1616	1886
164	Venetians	Venice	1469	1829
165	Veps	Sheltozero district		1972
166	Volga Germans	Saratov		1974
167	Vorarlbergers	Bregenz	1658	1955
168	Waldensians	Torre Pellice	1479	1973
169	Walloons	Namur		1927
170	Welsh	Cardiff	1546	1887
171	Zetlanders	Lerwick		

2 List of 47 European States in the Data Set

This table provides a list of 47 European states in my dataset, which describes the earliest date of the printing press, the name of the city where the first printing press was set up (whenever the name is available), the language with the largest population, and the date of the vernacular dictionary. The *Language* column denotes the official languages; for countries with multiple official languages, the language with the largest population is listed on the first row and those with fewer populations are on the second row.

The data is sorted by the date of print acquisition.

Table A2

	country	printing press	city	dictionary	language
1	Germany	1450	Mainz	1852	German (standard)
2	Austria	1461	Vienna	1963	German (regional)
3	Italy	1464	Subiaco	1861	Italian
4	Netherlands	1465	Haarlem	1864	Dutch
5	France	1470	Paris	1858	French
6	Switzerland	1470	Basel, Beromünster	1881	Swiss-German, French, Italian, Romansh
7	Belgium	1473	Bruges	1927	French (regional), Dutch
8	Hungary	1473	Buda (Budapest)	1890	Hungarian
9	Spain	1473	Valencia, Barcelona	1886	Spanish
10	Poland	1474	Kraków (Cracow)	1807	Polish
11	Czech Republic	1476	Plzeň (Pilsen)	1935	Czech
12	United Kingdom	1476	London	1755	English
13	Denmark	1482	Odense	1918	Danish
14	Croatia	1483	Zagreb	1901	Croatian
15	Sweden	1483	Stockholm	1893	Swedish
16	Portugal	1487	Faro	1852	Portuguese
17	Turkey	1493	Istanbul	1963	Turkish
18	Montenegro	1494	Podgorica		Montenegrin
19	Slovakia	1500	Bratislava	1959	Slovak
20	Romania	1508	Târgoviște	1907	Romanian (standard)
21	Armenia	1512	Yerevan	1971	Armenian
22	Greece	1512	Salonica (Thessaloniki)	1936	Greek
23	Belarus	1520		1977	Belarusian, Russian
24	Lithuania	1525	Vilnius	1956	Lithuanian
25	Iceland	1530	Hollar	1912	Icelandic

continued on next page

continued from previous page

	name	printing press	city	dictionary	language
26	Bosnia and Herzegovina	1531	Goražde	1992	Bosnian
27	Ireland	1550	Dublin	1913	Irish, English
28	Serbia	1552	Belgrade	1818	Serbian
29	Albania	1555	Tirana	1980	Albanian
30	Russia	1564	Moscow	1891	Russian
31	Ukraine	1574	Lviv (Lvov)	1886	Ukrainian
32	Slovenia	1575	Ljubljana	1976	Slovene
33	Latvia	1587	Riga	1972	Latvian
34	Luxembourg	1598	Luxembourg	1950	Luxemburgish, French, German
35	Estonia	1630	Tallinn	1918	Estonian
36	Finland	1642	Turku	1953	Finnish, Swedish
37	Malta	1642		1987	Maltese
38	Moldova	1642	Iași	1977	Romanian (regional)
39	Norway	1643	Olso	1930	Norwegian
40	Georgia	1709	Tbilisi	1884	Georgian
41	Monaco	1795			Monégasque
42	Bulgaria	1828	Samokov	1895	Bulgarian
43	Macedonia	1840		1961	Macedonian
44	Azerbaijan	1870	Baku	1966	Azeri
45	Andorra				
46	San Marino				
47	Liechtenstein				

3 Regression Tables

3.1 Cox proportional hazards and logistic regression models used in paper

Table A3 reports the complete regression estimates of Table 3 of the main body.

Table A3: Regression outputs of the printing press' impact on language standardization.

model	bivariate		demand side		supply side		fully specified	
	Cox (1)	logit (2)	Cox (3)	logit (4)	Cox (5)	logit (6)	Cox (7)	logit (8)
Printing press	2.159*** (0.269)	2.933*** (0.342)	1.600*** (0.356)	2.178*** (0.456)	1.455*** (0.354)	1.932*** (0.503)	1.450*** (0.339)	2.083*** (0.556)
University			0.0004*** (0.0001)	0.001*** (0.0002)	0.0004*** (0.0001)	0.001*** (0.0003)	0.0004*** (0.0001)	0.001*** (0.0002)
Bishop			-0.000002 (0.0001)	-0.0001 (0.0002)	0.00003 (0.0001)	0.00004 (0.0002)	-0.00002 (0.0001)	-0.00002 (0.0003)
Log urban potential			0.114 (0.226)	0.098 (0.365)	-0.002 (0.607)	-0.307 (0.793)	0.186 (0.636)	-0.276 (0.821)
Log distance to Wittenberg/Zürich					0.0001 (0.0003)	-0.0001 (0.001)	0.0004 (0.0003)	0.0002 (0.001)
War frequency					-0.031** (0.014)	-0.043*** (0.016)	-0.030** (0.013)	-0.048*** (0.018)
Elevation					-0.0001 (0.0004)	-0.00002 (0.001)	0.0001 (0.0004)	0.0005 (0.001)
Terrain ruggedness					-0.003** (0.001)	-0.004** (0.002)	-0.002** (0.001)	-0.004** (0.002)
Hub Roman road					0.156 (0.296)	0.494 (0.634)	0.129 (0.337)	0.431 (0.7)
Roman road					-0.009 (0.247)	-0.023 (0.467)	0.081 (0.304)	0.151 (0.603)
Oceanic port					-0.415** (0.212)	-0.712* (0.411)	-0.477* (0.26)	-0.721 (0.539)
Island					-0.057 (0.475)	0.185 (0.85)	-0.038 (0.52)	0.2 (0.965)
Ottoman Empire							-0.621 (0.812)	-0.822 (0.959)
Russian Empire							0.243 (0.237)	0.354 (0.473)
Habsburg Empire							-0.075 (0.33)	-0.252 (0.704)
Orthodox							0.279 (0.284)	0.725 (0.614)
Protestant							0.383 (0.235)	0.613 (0.474)
Islam							-0.939** (0.446)	-1.072 (0.694)

continued on next page

continued from previous page

model	bivariate		demand side		supply side		fully specified	
	Cox (1)	logit (2)	Cox (3)	logit (4)	Cox (5)	logit (6)	Cox (7)	logit (8)
Northern Europe							-0.246 (0.35)	-0.371 (0.719)
Southern Europe							0.017 (0.317)	-0.264 (0.695)
Eastern Europe							-0.194 (0.363)	-0.39 (0.768)
Western Asia							-0.621 (0.595)	-1.207 (1.043)
t		0.024 (0.017)		0.021 (0.017)		0.04 (0.029)		0.04 (0.029)
t^2		0.00001 (0.00002)		0.00001 (0.00002)		-0.00001 (0.00004)		-0.00001 (0.00004)
Intercept		-16.151*** (3.697)		-16.159*** (3.767)		-19.244*** (6.554)		-19.811*** (6.452)
AIC	1664.72	516.44	1650.75	494.78	1645.65	484.26	1645.01	480.83
Observations	2,052	2,052	2,052	2,052	2,052	2,052	2,052	2,052

Note: Robust standard errors clustered by ethnic groups for all models. *Western Europe* is used as the reference category for region fixed effects and thus omitted. *Catholic* is used as the reference category for religion fixed effects and thus omitted. *** denote $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

3.2 Robustness checks I: state as the unit of analysis

3.2.1 State-only sample

Table A4 reports the complete output of the abridged Cox proportional hazards models shown in Table 4 of the main body. In this series of robustness checks, I use states as the unit of analysis. “List” refers to a set of alternative dates of dictionary publications for countries with multiple official languages in the dependent variable.

Table A4: Cox proportional hazards model outputs of the printing press’ impact on language standardization.

model	List 1 (1)	List 2 (2)	List 3 (3)	List 4 (4)
Printing press	14.711*** (0.98)	14.984*** (0.92)	14.725*** (1.1)	14.814*** (0.97)
First year of printing press	-0.002 (0.001)	-0.001 (0.001)	-0.002* (0.001)	-0.002 (0.001)
University	0.0002 (0.0002)	0.0003 (0.0002)	0.0001 (0.0002)	0.0002 (0.0002)
Bishop	-0.0003** (0.0002)	-0.0002 (0.0002)	-0.0001 (0.0002)	-0.0002 (0.0002)
Log urban potential	0.932 (0.595)	1.335** (0.532)	1.254** (0.58)	0.937 (0.62)
Log distance to Wittenberg/Zürich	0.0007* (0.0004)	0.001* (0.0003)	0.001** (0.0004)	0.001* (0.0004)
War frequency	-0.003 (0.013)	-0.001 (0.012)	-0.002 (0.013)	-0.006 (0.013)
Elevation	-0.002*** (0.0005)	-0.001 (0.0006)	-0.002*** (0.0005)	-0.002*** (0.0005)
Terrain ruggedness	-0.005* (0.003)	-0.006** (0.003)	-0.005* (0.003)	-0.005* (0.003)
Hub Roman road	0.336 (0.364)	0.468 (0.399)	0.547 (0.427)	0.462 (0.377)
Roman road	0.402 (0.324)	-0.028 (0.338)	0.234 (0.38)	0.378 (0.324)
Oceanic port	-0.8** (0.337)	-0.59* (0.351)	-0.886*** (0.344)	-0.805*** (0.336)
Island	0.115 (0.258)	0.268 (0.271)	0.160 (0.262)	0.118 (0.25)

continued on next page

continued from previous page

model	List 1 (1)	List 2 (2)	List 3 (3)	List 4 (4)
Ottoman Empire	-1.822* (1.1)	-2.127** (1.06)	-1.949* (1.09)	-1.702 (1.1)
Russian Empire	-0.244 (0.289)	-0.054 (0.251)	-0.225 (0.295)	-0.216 (0.298)
Habsburg Empire	-0.642 (0.441)	-0.925** (0.455)	-0.863* (0.503)	-0.641 (0.45)
Orthodox	-0.095 (0.462)	0.0002 (0.453)	-0.164 (0.535)	-0.092 (0.472)
Protestant	1.250*** (0.303)	0.268 (0.31)	0.885*** (0.316)	1.136*** (0.317)
Islam	-0.371 (1.163)	-0.267 (0.948)	0.020 (1.051)	-0.389 (1.173)
Northern Europe	-0.378 (0.303)	0.110 (0.311)	-0.343 (0.312)	-0.259 (0.307)
Southern Europe	0.685* (0.359)	0.537 (0.395)	0.629 (0.4)	0.780** (0.374)
Eastern Europe	1.105** (0.5)	0.640 (0.526)	0.758 (0.593)	1.074** (0.513)
Western Asia	1.592* (0.963)	0.757 (0.965)	1.076 (0.958)	1.646* (0.94)
AIC	538.29	596.87	558.7	533.87
Observations	564	564	564	564

Note: Robust standard errors clustered by country for all models. *Western Europe* and *Catholic* are used as the reference category, respectively, for geographic region fixed effects and for religion fixed effects so they are omitted. *** denote $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

3.2.2 Stateless ethnic groups-only sample

Table A5 documents Cox regression estimates using stateless ethnic groups (n = 124). This is reported in Column 5 of Table 4 in the main body.

Table A5: Regression outputs of the printing press' impact on language standardization.

model	Cox (1)
Printing press	4.915** (2.293)
Print year	-0.002* (0.001)
University	0.0002 (0.0002)
Bishop	-0.00005 (0.0002)
Log urban potential	0.700 (0.967)
Log distance to Wittenberg/Zürich	0.001 (0.0005)
War frequency	-0.032 (0.02)
Elevation	0.0005 (0.0006)
Terrain ruggedness	-0.004** (0.002)
Hub Roman road	-0.254 (0.396)
Roman road	0.021 (0.369)
Oceanic port	-0.434 (0.395)
Island	-0.438 (0.966)
Ottoman Empire	-15.591*** (0.875)
Russian Empire	0.979*** (0.297)
Habsburg Empire	0.543 (0.571)

continued on next page

continued from previous page

model	Cox (1)
Orthodox	0.202 (0.403)
Protestant	0.469 (0.297)
Islam	-1.295** (0.571)
Northern Europe	-0.413 (0.537)
Southern Europe	0.136 (0.48)
Eastern Europe	-1.079*** (0.412)
Western Asia	-2.881** (1.328)
AIC	815.41
Observations	1,488

Note: Robust standard errors clustered by country for all models. *Western Europe* and *Catholic* are used as the reference category, respectively, for geographic region fixed effects and for religion fixed effects and are omitted. *** denote $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

3.3 Robustness checks II: The role of human capital

Table A6 documents the full regression estimates of Table 5 of the main body.

Table A6: Regression outputs of the printing press' impact on language standardization.

model	Cox (1)	logit (2)	Cox (3)	logit (4)
Printing press			1.394*** (0.342)	2.023*** (0.558)
Vernacular Bible	0.531** (0.251)	0.745** (0.356)	0.461* (0.257)	0.676* (0.365)
University	0.0005*** (0.0001)	0.001*** (0.0002)	0.0003** (0.0001)	0.001*** (0.0003)
Bishop	0.00006 (0.0001)	0.0001 (0.0003)	-0.00004 (0.0001)	-0.0001 (0.0003)
Log urban potential	-0.157 (0.613)	-0.554 (0.805)	0.168 (0.616)	-0.259 (0.818)
Log distance to Wittenberg/Zürich	-0.0001 (0.0003)	-0.0004 (0.001)	0.0004 (0.0003)	0.0002 (0.001)
War frequency	-0.026** (0.013)	-0.038** (0.016)	-0.03** (0.013)	-0.048*** (0.018)
Elevation	0.0003 (0.0004)	0.001 (0.001)	0.0002 (0.0004)	0.001 (0.001)
Terrain ruggedness	-0.003*** (0.001)	-0.005** (0.002)	-0.003** (0.001)	-0.004** (0.002)
Hub Roman road	0.217 (0.341)	0.531 (0.693)	0.054 (0.336)	0.267 (0.712)
Roman road	0.007 (0.29)	0.082 (0.582)	0.107 (0.31)	0.252 (0.62)
Oceanic port	-0.213 (0.252)	-0.317 (0.487)	-0.436* (0.263)	-0.658 (0.541)
Island	-0.36 (0.507)	-0.374 (0.888)	-0.115 (0.511)	0.063 (0.949)
Ottoman Empire	-0.666 (0.792)	-0.914 (0.907)	-0.589 (0.792)	-0.834 (0.892)
Russian Empire	0.018 (0.252)	-0.046 (0.491)	0.173 (0.235)	0.244 (0.472)
Habsburg Empire	0.075 (0.341)	0.007 (0.646)	-0.088 (0.337)	-0.265 (0.7)

continued on next page

continued from previous page

model	Cox (1)	logit (2)	Cox (3)	logit (4)
Orthodox	0.467 (0.291)	0.963 (0.588)	0.286 (0.286)	0.786 (0.602)
Protestant	0.337 (0.248)	0.561 (0.494)	0.345 (0.236)	0.527 (0.482)
Islam	-0.913** (0.454)	-1.031 (0.661)	-0.870* (0.448)	-1.016 (0.679)
Northern Europe	-0.179 (0.346)	-0.259 (0.699)	-0.188 (0.347)	-0.304 (0.717)
Southern Europe	0.006 (0.314)	-0.271 (0.700)	0.001 (0.313)	-0.363 (0.72)
Eastern Europe	-0.190 (0.376)	-0.343 (0.747)	-0.102 (0.371)	-0.281 (0.753)
Western Asia	-0.374 (0.658)	-0.763 (1.025)	-0.650 (0.594)	-1.296 (1.039)
t		0.044 (0.03)		0.041 (0.029)
t^2		-0.00001 (0.00004)		-0.00001 (0.00004)
Intercept		-19.325*** (6.777)		-20.064*** (6.476)
AIC	1660.38	495.57	1642.95	478.43
Observations	2,052	2,052	2,052	2,052

Note: Robust standard errors clustered by ethnic groups for all models. *Western Europe* is used as the reference category for region fixed effects and thus omitted. *Catholic* is used as the reference category for religion fixed effects and thus omitted. *** denote $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

3.4 Instrumental variables (IV) approach: Distance from Mainz as IV

Table A7 reports the full regression estimates of Table 6 of the paper.

Table A7: IV probit regression outputs of the printing press' impact on language standardization.

dependent variable N cumulative dictionary	First stage	Second stage			
	printing press (1)	language standardization			
		by 1850 n = 7 (2)	by 1900 n = 27 (3)	by 1950 n = 48 (4)	by 2000 n = 104 (5)
Log distance to Mainz	-0.159*** (0.060)				
Printing press		11.822 (37.038)	4.639 (3.051)	4.741** (2.284)	4.018** (1.828)
Print adoption year	0.0003*** (0.00004)	0.001 (0.006)	-0.001 (0.001)	-0.001 (0.001)	-0.0004 (0.001)
University	0.0001*** (0.00003)	0.0001 (0.005)	0.0002 (0.0003)	0.0001 (0.0003)	0.0002 (0.0002)
Bishop	0.00001 (0.00003)	-0.001 (0.004)	-0.0001 (0.0002)	-0.0002 (0.0002)	-0.00003 (0.0001)
Log urban potential	-0.004 (0.073)	7.111** (3.010)	2.221*** (0.217)	1.280*** (0.099)	1.275*** (0.074)
Log distance to Wittenberg/Zürich		-0.005 (0.005)	0.001** (0.001)	0.001*** (0.0003)	0.001*** (0.0003)
War frequency	-0.002 (0.003)	0.007 (0.092)	-0.011 (0.022)	-0.022* (0.013)	-0.025*** (0.009)
Elevation	0.0001** (0.00003)	0.005 (0.011)	-0.0002 (0.001)	-0.0004 (0.001)	-0.0001 (0.0004)
Terrain ruggedness	-0.0002** (0.0001)	-0.044 (0.070)	-0.002 (0.003)	-0.001 (0.002)	-0.001 (0.001)
Hub Roman road	0.131** (0.052)	-6.338 (4.421)	-0.458 (0.596)	-0.480 (0.456)	-0.315 (0.390)
Roman road	-0.008 (0.040)	1.358 (3.645)	0.473 (0.576)	0.268 (0.449)	0.141 (0.349)
Oceanic port	0.041 (0.043)	0.444 (3.262)	-0.380 (0.459)	-0.527 (0.366)	-0.538* (0.286)
Island	-0.005 (0.053)	6.790*** (2.019)	1.158** (0.502)	0.452 (0.429)	0.392 (0.417)

continued on next page

continued from previous page

dependent variable N cumulative dictionary	First stage	Second stage			
	printing press	language standardization			
		by 1850 n = 7	by 1900 n = 27	by 1950 n = 48	by 2000 n = 104
	(1)	(2)	(3)	(4)	(5)
Ottoman Empire	-0.049 (0.045)	4.178** (1.643)	-0.220 (0.611)	-0.151 (0.480)	-0.322 (0.467)
Russian Empire	0.020 (0.026)	-2.738*** (0.766)	0.253 (0.526)	0.262 (0.351)	0.320 (0.275)
Habsburg Empire	0.118** (0.046)	-3.027** (1.300)	-0.673 (0.777)	-0.668 (0.596)	-0.590 (0.445)
Orthodox	-0.144** (0.060)	2.110 (3.658)	0.865 (0.594)	0.544 (0.452)	0.526 (0.352)
Catholic/Protestant	-0.213*** (0.043)	-0.390 (5.030)	-0.390 (0.697)	-0.454 (0.468)	-0.269 (0.372)
Islam	-0.124** (0.056)	0.158 (4.829)	-4.899*** (0.984)	-5.652*** (0.609)	-0.680* (0.386)
Northern Europe	-0.001 (0.062)	6.131*** (1.878)	0.020 (0.480)	0.270 (0.400)	0.109 (0.387)
Southern Europe	0.156** (0.067)	5.603*** (2.136)	0.102 (0.549)	-0.297 (0.395)	-0.058 (0.315)
Eastern Europe	0.060 (0.070)	9.491** (4.449)	0.614 (0.846)	0.044 (0.591)	0.017 (0.444)
Western Asia	0.071 (0.078)	11.974* (6.995)	1.308 (1.532)	0.198 (1.097)	-0.566 (0.622)
Intercept	1.28*** (0.413)	-28.929 (18.163)	-9.522*** (1.138)	-6.967*** (0.749)	-6.992*** (0.598)
Observations	2,052	1,539	1,710	1,881	2,052
p-value, Wald exogeneity test		0.75	0.12	0.038*	0.028*
F-statistic on weak instrument	24.99				

Note: Robust standard errors clustered by ethnic groups. IV probit estimation: first stage is OLS, second stage is probit, regressed on predicted values from the first stage. In the first stage, *log urban potential* and *war frequency* are for the fourteenth centuries. *Western Europe* is used as the reference category for region fixed effects and thus omitted. In the second stage, *Catholic* is used as the reference category for religion fixed effects and thus omitted. *** denote $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.