Web appendix for "The Weight of History on European Cultural Integration: a Gravity Approach"

***American Economic Review*, Papers and Proceedings May 2011**

Empirical Model, Description of Variables and Additional Results

Pauline Grosjean*

1. Empirical Model

As usual in the trade gravity literature, the relationship between cultural distance and physical distance is assumed to be concave and the following dyadic model is estimated:

$$DC_{\ln} = \alpha_1 E + \alpha_2 \ln(DIST_{\ln}) + \alpha_3 I + \alpha_4 C + \alpha_5 DS_{\ln} + \alpha_6 DO_{\ln} + \alpha_7 DR_{\ln} + \alpha_8 DI_{\ln} + \delta_l + \delta_n + \varepsilon_{\ln}$$

where the MD variables are metrics of dissimilarity between pairs of locations *l* and *n*. Manhattan Distances, which sum over the absolute differences in shares of responses for multinomial variables (Head and Mayer 2008) are used throughout. $DC_{\ln}$ is defined as: $DC_{\ln} = \sum_{i=1}^{I} |s_{il} - s_{in}|$ where s_{il} (respectively s_{in}) is the share, in location *l* (respectively *n*) of responses allocated to each modality *i* of the *I* modalities of the following widely used question on social trust, which was replicated in the *Life in Transition Survey*: *"Generally Speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?"* Responses modalities are on a scale of 1 to 5, with 1 corresponding to "Complete Distrust" and 5 to "Complete Trust".

Vertical transmission of preferences from parents to children implies some hysteresis in the transmission of preferences, which explains why history matters (Bisin and Verdier, 2001; Tabelini, 2010). The influence of history is reflected through different dummy variables *E*, which reflect how long each locality pair was included in the same empire for (see Section 2). The minimum duration of integration is 100 years and the maximum 700. A similar past history is expected to decrease cultural distance so that α_1 is expected to be negative.

The next covariate in (1) examines the role of geographic distance between locations. The geographical distance between members of PSU pairs is computed from longitudinal and

* University of San Francisco. pagrosjean@usfca.edu

latitudinal coordinates of the geographical center of the PSU using the great circle formula. Since geographic distance is an obstacle to cultural diffusion, so that α_2 is expected to be positive. As usual in the gravity trade models, *I* is a dummy variable indicating whether the two districts are located in the same country and *C* is a dummy variable to indicate whether two locations belong to different but adjacent countries.

Controls include dissimilarities between location pairs in terms of income ($DI_{\ln}$), education ($DE_{\ln}$), religion ($DR_{\ln}$), economic occupations ($DO_{\ln}$) and the characteristics of the local economy ($DI_{\ln}$). To reflect differences in the characteristics of the local economy, I use the index developed by Grosjean and Senik (forth) that captures the development of the modern sector of the economy and the advancement towards a market economy.[1] The categories used in the construction of the Manhattan distance in occupations are the following: unemployed; pensioner; student; housewife; or employed, with this last category broken up in different subcategories of: white collar, blue collar, service worker and farmer or farm-worker. The different categories of employment were constructed from the responses about the respondent's first job, using the ISCO classification.[2] Differences between PSUs in the age and gender of respondents are also included as controls.

Finally, the regression specification includes an error term and a set of intercepts for each location (Primary Sampling Unit). Also, to account for the fact that standard errors are correlated among every pair that contains a specific observation, standard errors are adjusted for clustering on any observation that contains either member of a pair, following the multi-way clustering method by Cameron et al. (2010).

The main advantage of this econometric specification over a traditional cross-sectional specification is to take into account and correct for the dependence between observations, which consist not only in spatial dependence but, more importantly, in the reflection problem Manski (1993). Estimating a cross sectional equation in which the dependent variable would be the share

[1] This index ("*industrial index*") measures the proportion of respondents in the active labor force who are self-employed with more than five employees, or have a formal labor contract and either: work in a small enterprise, work in a medium enterprise, work in a private firm, work in a newly created enterprise (since 1989).

[2] The ISCO categories corresponding to our white collar category are: 1: Legislator, Senior Official and Manager, 2: Professionals, and 3: Technicians and Associated Professionals. Our service workers category consists of: 4: Clerks and 5: Service workers and shop and market sales workers. 6: Skilled agricultural and fishery workers are in our "farmer and farm worker category" together with independent farmers. All the remaining ISCO categories are considered as blue collar workers.

of a specific cultural trait (say, the proportion of trusting people) would run into severe complications and shortfall. Indeed, because of the horizontal diffusion of cultural traits (Bisin and Verdier 2001, Tabellini 2010), it is likely not only that there is spatial dependence between observations but also, more importantly, that the prevalence of a specific cultural trait in a location depends in part on the share of the cultural trait in all adjacent locations and vice versa. This has been discussed and labeled the "reflection problem" by Manski (1993). As a result, the OLS estimates of such a cross-sectional equation would be biased. In Manski's formulation, the researcher posits that there are social effects, but does not observe them directly and instead infers them based on the attributes of the individual in question. To describe a simple analogy, the problem is similar to that of interpreting the almost simultaneous movements of a person and his reflection in a mirror. As argued already by Head and Mayer (2007), incorporating the share of cultural trait in adjacent locations in the dependent variable and including location fixed effects avoid the econometric issues associated with the reflection problem. Differentiating absorbs the reflection effect. Including a location fixed effect capture third-location effects on the pairwise differences (Head and Mayer 2008). Multi way clustering of standard errors corrects for the dependence between every pair that contains a specific observation. Finally, the spatial dependence between observations is directly dealt with by incorporating distance between locations as a dependent variable. Alternative specifications including distance (not in logarithmic form) and distance squared do not alter the results.

2. History

The main historical event of interest is the integration of localities in the former Empires of Central, Eastern and South Eastern Europe: the Ottoman Empire, the Habsburg Empire, Prussia (and the Teutonic Order) and the Russian Empire, which occupied all the localities in the sample (with the notable exception of Upper Zeta, the central region of Montenegro). The Ottoman Empire's territorial extension in South-Eastern Europe occurred mainly in the XIVth Century (Bulgaria, South Serbia, FYROM) and the XVth century (Albania, Bosnia and Herzegovina, Crimea, Moldavia, Wallachia and Montenegro). Territorial losses of the Ottoman Empire occurred chiefly in two waves: at the end of the XIXth century, after the Russian-Turkish War of 1877-1878, and after the Balkan Wars of 1911-1912. I designate by 'Habsburg Empire' what was the Kingdom of Hungary and the Austrian Empire and became the Austria-Hungarian

Empire after the 1867 Ausgleich. Successor states became independent after the Saint-Germain and Sevres treaties of 1918 and include territories that now belong to Croatia, Hungary, the Czech Republic, Poland, Romania, the Slovak Republic, Slovenia, Serbia and Ukraine. 'Prussia' designate Prussia per se (1525-1947) as well as Old Prussia (the Teutonic Order). Prussia encompassed territories that are today part of Poland, Lithuania and Czech Republic, as well as Germany, Denmark, Belgium and the Netherlands (not included in our sample). Most territories outside of today's Germany were lost at the Treaty of Versailles in 1919. The Russian Empire's territorial expansion in Europe occurred mainly under Peter the Great in the XVIIth century and Catherine the Great in the XVIIIth century. Bessarabia was gained from the Ottomans in 1812. The Russian Empire encompassed more or less the Soviet Union, with the addition of Polish territories, Turkish territories, but a much smaller Ukraine, not all territories of the Baltic states and without Kaliningrad. I use the Periodical Atlas of Europe in order to reconstruct Empire delimitations and their evolution across time, from 1300, the start of empire consolidation in Medieval Europe, to 2000. Table A2 presents some descriptive statistics on the duration and geographical delimitation of these empires.

3. Descriptive Statistics and Additional Results

A. Descriptive Statistics

Figure A1. Map of Dynastic Empires in Central and Eastern Europe

Notes to Figure A1: The figure indicates PSUs that belonged to the Russian, Prussian, Habsburg and Ottoman Empires for more than 200 years only in Central, Eastern and South Eastern Europe. Source: Periodical Atlas of Europe 1300-2000.

Table A1: List of countries included in the sample

Albania, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Former Yugoslav Republic of Macedonia, Hungary, Latvia, Lithuania, Moldova, Montenegro, Poland, Romania, Russian Federation, Serbia, Slovak Republic, Slovenia, Turkey, Ukraine.

Table A2: Geographical and temporal delimitations of Empires in the sample

	400 years or more	200 to 400 years	200 years or less
Ottoman Empire	Albania, Sancak of Bosnia, Bulgaria (except part which became independent in 1878), FYROM, Turkey	Rest of Bosnia, Bulgaria which became independent in 1878, Bessarabia, Crimea, Moldavia, outer Montenegro, Serbia except Vojvodina, Wallachia	Transylvania, Vojvodina, southern Hungary, eastern Croatia
Habsburg Empire	Slovakia, Hungary (except southern-ottoman Hungary)	Croatia (except Dalmatia), Czech Republic, Southern (ottoman) Hungary , Polish Silesia, Slovenia, Transylvania, Vojvodina	Dalmatia, Galicia, Habsburg Poland (Krakow, Rzeszow), Bosnia, western Ukraine (Lviv)
Prussia	Pomerania (Poland)	Estonia, Latvia, Polish Silesia, Royal Prussia	Polish Silesia, Kaliningrad, Klaipeda (Lithuania)
Russian Empire		Russia (except Kaliningrad), Ukraine (except Crimea and Kouban)	Belarus, Baltic states (except Klaipeda), Moldova, eastern-central Poland (Warsaw, Lodz), rest of Ukraine

Source: Periodical Atlas of Europe 1300-2000

Table A3: Descriptive statistics

Variable	Description	Mean	s.d..	Min	Max
Social trust	Manhattan Distance social trust	0.87	0.37	0	2
Occupation	Manhattan Distance Occupation	0.89	0.31	0.04	2.48
Same Empire	Dummy = 1 if PSU pair in same former empire (Habsburg, Ottoman, Prussia and Russian Empires) for more than 100 years	0.25	0.43	0	1
Distance in km	distance between 2 PSUs (in km)	1029	783	0	12737
Distance	Logarithm of distance in km	6.66	0.83	0	9.45
MD religion	Manhattan distance for main religions (Muslim, Christian, Jewish, other religion and Atheist)	0.79	0.71	0	2
MD social class	Manhattan distance of proportions of rich, poor and middle income	0.65	0.36	0	2
MD education	Manhattan distance education (5 education categories)	0.89	0.37	0	2
Difference in age	Absolute value of difference in average age of PSUs	8.12	6.14	0	49.46
MD labor index	Absolute value of the difference in the proportion of respondents in the active labor force who are self-employed with more than five employees, or have a formal labor contract and either: work in a small enterprise, work in a medium enterprise, work in a private firm, work in a newly created enterprise (since 1989)	0.58	0.45	0	3.61
Same country	Dummy =1 if PSU pair in same country	0.05	0.21	0	1
Contiguous	Dummy =1 if PSU pair in adjacent countries	0.17	0.38	0	1
Same Empire for 100 to 200 years	Dummy = 1 if PSU pair in same former empire (Habsburg, Ottoman, Prussia and Russian Empires) for between 100 and 200 years	0.12	0.32	0	1
Same Empire for 200 to 400 years	Dummy = 1 if PSU pair in same former empire (Habsburg, Ottoman, Prussia and Russian Empires) for between 100 and 200 years	0.03	0.16	0	1
Same Empire for more than 400 years	Dummy = 1 if PSU pair in same former empire (Habsburg, Ottoman, Prussia and Russian Empires) for more than 400 years	0.05	0.21	0	1
Habsburg Empire	Dummy = 1 if PSU pair in the Habsburg Empire for more than 100 years	0.12	0.32	0	1
Ottoman Empire	Dummy = 1 if PSU pair in the Ottoman Empire for more than 100 years	0.20	0.40	0	1
Prussia	Dummy = 1 if PSU pair in Prussia for more than 100 years	0.014	0.12	0	1
Russian Empire	Dummy = 1 if PSU pair in the Russian Empire for more than 100 years	0.12	0.32	0	1
USSR	Dummy =1 if PSU pair in former USSR	0.11	0.31	0	1
Yugoslavia	Dummy =1 if PSU pair in former Yugoslavia	0.08	0.27	0	1
EU	Dummy =1 if PSU pair in European Union (EU)	0.23	0.42	0	1
EU candidate	Dummy =1 if PSU pair in candidate countries to EU	0.01	0.12	0	1
EU-EU candidate	Dummy =1 if 1 PSU in EU, the other in candidate country	0.14	0.34	0	1

Source: Life in Transition Survey 2009; Periodical Atlas of Europe 1300-2000.

B. Additional Results

Table B1: Results of Full Specification

	1	2	3	4
	Manhattan Distance		Social Trust	
Distance	0.030***	0.031***	0.028***	0.030***
	[0.003]	[0.004]	[0.003]	[0.004]
Same Empire more than 100 years	-0.046***	-0.047***	-0.041***	-0.041***
	[0.008]	[0.008]	[0.007]	[0.007]
Same country		-0.002		0.006**
		[0.009]		[0.009]
Contiguous		0.008		0.008***
		[0.006]		[0.006]
Presence of mountain range				0.016**
				[0.007]
Differences in:				
Religious affiliation			0.019***	0.019***
			[0.005]	[0.005]
Income			0.0001	0.0001
			[0.001]	[0.001]
Education			0.018**	0.018**
			[0.006]	[0.006]
Age			-0.001***	-0.001***
			[0.0002]	[0.0002]
Industrial index			0.004	0.004
			[0.004]	[0.004]
Gender			0.003	0.003
			[0.005]	[0.005]
Occupations			0.026***	0.026***
			[0.007]	[0.007]
Observations	547581	547581	535095	535095
R-squared	0.452	0.452	0.447	0.447

Table B2: Influence of Different Empires, Current Democratization and EU Integration

	1	2	3
	Manhattan Distance		Social Trust
Same Empire more 100 years		-0.040***	-0.041***
		[0.008]	[0.007]
Habsburg more 100 years	0.009		
	[0.009]		
Ottoman more than 100 years	-0.091***		
	[0.001]		
Prussia more than 100 years	-0.041**		
	[0.017]		
Russian Empire more than 100 years	0.048**		
	[0.010]		
Distance	0.029***	0.029***	0.027***
	[0.004]	[0.004]	[0.004]
Democracy difference		0.003*	
		[0.001]	
EU			-0.004
			[0.009]
EU-EU candidates			-0.036
			[0.029]
EU candidates			0.050***
			[0.010]
Socio-demo and eco controls	yes	yes	yes
Same country	yes	yes	yes
Contiguous	yes	yes	yes
Observations	535095	535095	535095
R-squared	0.448	0.447	0.449

Notes to Table B1 and B2: All dependent variables are Manhattan Distance measures of dissimilarity between PSU pairs. "*Distance*" is expressed as the logarithm of the physical distance between PSUs in km. The *Same Empire…* dummies takes value 1 if both PSUs have belonged to the same Empire for the number of years indicated. *Same Country* and *Contiguity* are dummy variables that take value 1 if members of the pair belong to, respectively, the same country or adjacent countries. *Socio-demo and eco controls* include dissimilarity between pairs of PSUs in terms of religious affiliation, social class composition, education, age, industrial index and occupations. *Democracy difference* is the difference in Polity IV scores between the respective countries where the PUSs are located. It is 0 if both members are in the same country.

EU takes value one if both member of the PSU pair belongs to countries that are EU members. *EU candidates* takes value one if both member of the PSU pair belongs to countries that are EU. *EU-candidate EU* takes value one if one PSU belongs to a member states and the other to a candidate country. All regressions include a constant and control for location fixed effects. Robust standard errors are adjusted for clustering on any observation which contains either member of a pair (multi-way clustering by each member of the pair following the Cameron et al. (2010) method). *** indicates significance at 1%, ** at 5%, * at 10%.

Source: Life in Transition Survey 2009; Periodical Atlas of Europe 1300-2000.

4. Additional References

Bisin, Alberto and Thierry Verdier 2001. "The Economics of Cultural Transmission and the Dynamics of Preferences." Journal of Economic Theory, 97(2): 298-319.

Head, Thierry and Keith Head, 2008. "Detection of local interactions from the spatial pattern of names in France." *Journal of Regional Science*, 48: 67-95.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem," The Review of Economic Studies, 60(3), 531–542

Tabellini, Guido 2010 **"**Culture and Institutions: Economic Development in the Regions of Europe." *Journal of the European Economic Association*, 8(4): 677-716.