Fertility and Modernity: Appendix

Enrico Spolaore
Tufts University and NBER

Romain Wacziarg
UCLA and NBER

June 2019

Abstract

This online appendix contains: 1) A theoretical appendix, 2) A discussion of the population-level dataset and the corresponding empirical results, 3) A description of the sources of the data.

1 Theoretical Appendix: Extension to the Case $\beta d(Y,Z) < 1$

In the theoretical analysis, we assumed for simplicity that the social barriers between society Y and society Z are prohibitive: $\beta d(Y,Z) \geq 1$. Here we extend the analysis to the case $\beta d(Y,Z) < 1$.[1] This extension slightly complicates the formal analysis but does not modify the substance of the model's implications about the relationship between fertility transition dynamics and distance from the innovator.

At time 0, only the innovator society has adopted the new social norms, and therefore $M_{X0} = 1$, $M_{Y0} = M_{Z0} = 0$. At time 1, the new social norms are adopted by all households in society Y for whom the social threshold μ_i is smaller or equal to the mass of households who have already adopted the innovation in society X, weighed by their social distance - that is, all households such that:

$$\mu_i \leq [1 - \beta d(X,Y)]M_{X0} = 1 - \beta d(X,Y) \tag{1}$$

Consequently, at time 1 the new social norms are adopted by the following fraction of households in society Y:

$$M_{Y1} = \min\left\{\frac{1 - \beta d(X,Y)}{\mu_H - \mu_L}, 1\right\} \tag{2}$$

In society Z two cases are possible. For $\mu_L \geq 1 - \beta d(X,Z)$ (relatively high levels of societal conformism and/or high levels of inter-societal barriers $\beta d(X,Z)$), no household in society Z adopts the new social innovation at time 1. For $\mu_L < 1 - \beta d(X,Z)$, a positive fraction of households in society Z also adopts the new social norms introduced in society X. In that case, the mass of households adopting the new social norm is given by:

$$M_{Z1} = \min\left\{\frac{1 - \beta d(X,Z)}{\mu_H - \mu_L}, 1\right\} \tag{3}$$

The number of adopters of new social norms is lower in society Z than in society Y ($M_{Z1} \leq M_{Y1}$), because of the larger relative social distance from the innovator $d(X,Z) > d(X,Y)$. The only instance when $M_{Z1} = M_{Y1}$ is in the extreme case when all households in both societies adopt the new social norms immediately, which would occur at very low levels of barriers and/or conformism, that is, for $\mu_H \leq 1 - \beta d(X,Z)$.

At time 1, the average level of fertility in society Y is:

$$f_{Y1} = M_{Y1} f_m + (1 - M_{Y1}) f_n \tag{4}$$

[1] For simplicity, we continue to assume $\mu_L < 1 - \beta d(X,Y)$.

while the average level in society Z is:

$$f_{Z1} = M_{Z1} f_m + (1 - M_{Z1}) f_n \tag{5}$$

In general, $f_{Z1} \geq f_{Y1}$, with the highest gap between f_{Z1} and f_{Y1} occuring when $f_{Z1} = f_n$, i.e. for $\mu_L \geq 1 - \beta d(X, Z)$. In contrast, there will be no gap ($f_{Z1} = f_{Y1}$) in the extreme case $M_{Y1} = M_{Z1} = 1$, i.e. for $\mu_H \leq 1 - \beta d(X, Z)$.

In the rest of this analysis, we abstract from polar cases, and focus on the intermediate range of parameters for which a positive number of households, but not all households, adopt the novel behavior in society Z at time 1 - that is, the case $\mu_L < 1 - \beta d(X, Z) < \mu_H$.

At time 2, in society Y the new social norms are adopted by all households with critical threshold μ_i such that:

$$\mu_i \leq 1 - \beta d(X, Y) + \frac{1 - \beta d(X, Y)}{\mu_H - \mu_L} + [1 - \beta d(Y, Z)] \frac{1 - \beta d(X, Z)}{\mu_H - \mu_L} \tag{6}$$

which implies the following number of modern households in society Y at time 2:

$$M_{Y2} = \min \left\{ \frac{1}{\mu_H - \mu_L} [1 - \beta d(X, Y) + \frac{1 - \beta d(X, Y)}{\mu_H - \mu_L} + [1 - \beta d(Y, Z)] \frac{1 - \beta d(X, Z)}{\mu_H - \mu_L}], 1 \right\} \tag{7}$$

By the same token, at time 2 in society Z the new social norms are adopted by all households with critical threshold μ_i such that:

$$\mu_i \leq 1 - \beta d(X, Z) + \frac{1 - \beta d(X, Z)}{\mu_H - \mu_L} + [1 - \beta d(Y, Z)] \frac{1 - \beta d(X, Y)}{\mu_H - \mu_L} \tag{8}$$

which implies the following number of modern households in society Z:

$$M_{Z2} = \min \left\{ \frac{1}{\mu_H - \mu_L} [(1 - \beta d(X, Z) + \frac{1 - \beta d(X, Z)}{\mu_H - \mu_L} + [1 - \beta d(Y, Z)] \frac{1 - \beta d(X, Y)}{\mu_H - \mu_L}], 1 \right\} \tag{9}$$

and so on for $t \geq 2$. Thus, at each time t, societies closer to the innovator have larger shares of households adopting the new behavior: $M_{Yt} \geq M_{Zt}$.

As in Section 3, let $T(f^{\#})$ denote the *earliest* time at which a given level $f^{\#} = M^{\#} f_m + (1 - M^{\#}) f_n$ is achieved. As $M_{Yt} \geq M_{Zt}$, we have that, as in Section 3, $T(f^{\#})$ occurs earlier for society Y at distance $d(X, Y)$ than for society Z at distance $d(X, Z) > d(X, Y)$:

$$T_Y(f^{\#}) < T_z(f^{\#}) \tag{10}$$

implying:

Proposition 1: *Societies at a smaller social distance from the social innovator experience an earlier transition to lower fertility.*

The relation of the dynamics of the fertility transition status with the distance from the innovator is also analogous to the one derived in Section 3, and can be illustrated with a similar numerical example. To fix ideas, assume that $\beta d(X, Y) = 2/3$, $\beta d(X, Z) = 4/5$ and $\beta d(Y, Z) = 6/7$. Societies Y and Z will experience transitions to lower fertility as detailed in the following table:

Time	M_{Yt}	f_{Yt}	M_{Zt}	f_{Zt}
1	$\frac{1}{3}$	$\frac{1}{3}f_m + \frac{2}{3}f_n$	1/5	$\frac{1}{5}f_m + \frac{4}{5}f_n$
2	$\frac{2}{3} + \frac{1}{5}\frac{1}{7}$	$(\frac{2}{3} + \frac{1}{35})f_m + (\frac{1}{3} - \frac{1}{35})f_n$	$\frac{2}{5} + \frac{1}{3}\frac{1}{7}$	$(\frac{2}{5} + \frac{1}{21})f_m + (\frac{3}{5} - \frac{1}{21})f_n$
3	1	f_m	$\frac{3}{5} + \frac{1}{21} + \frac{1}{7}(\frac{2}{3} + \frac{1}{35}) = \frac{183}{245}$	$\frac{183}{245}f_m + \frac{62}{245}f_n$
4	1	f_m	1	f_m

In this example, as in the one discussed in Section 3, society Y will achieve full modernity before society Z: at time 3 rather than at time 4. Eventually, both societies transition to the full modern equilibrium in which fertility is equal to f_m. Overall, fertility levels are inversely related to distance from the innovator in the earlier phases of the transition to lower fertility, but the relation between fertility and distance from the innovator across societies eventually fades. In sum, as in Section 3, we have:

Proposition 2: *The absolute magnitude of the negative relationship between a society's transition status and its distance from the innovator is lower in the earlier phases of the diffusion of the new fertility behavior, becomes higher over time, and falls again in the latest stages of the fertility transition.*

The levels of fertility also follow a similar pattern in relation to distance from the innovator:

Proposition 3: *In the earlier phases of the diffusion of the fertility decline, there is a strong positive relationship between fertility levels and distance from the innovator, but this relationship becomes weaker as more societies adopt modern social norms over time. Consequently, measured correlations between fertility levels and relative social distance from the innovator are high and positive during the earlier phases of the transition, and decline over time as more societies decrease their fertility levels.*

2 Empirical Appendix

2.1 Description of the Population-level Dataset

In addition to the regional dataset describes in the paper, we also constructed a more aggregated dataset of 37 populations, allowing the use of alternative measures of social distance. This population-level dataset is used for purposes of robustness with respect to the measure of social distance.

The availability of alternative measures of social distance guided the choice of the populations included in this dataset, so we begin with a discussion of these alternative measures. The first one is genetic distance, which captures the degree of genealogical relatedness between populations, and is used for such a purpose in the economics literature (see, for instance, Spolaore and Wacziarg, 2009). It is a summary measure of differences in gene frequencies between populations, constructed using neutral genes that are not subject to selective pressure (Cavalli-Sforza, Menozzi and Piazza, 1994). When two populations split apart, random genetic changes result in increasing genetic differentiation from one generation to the next. The longer the separation time, the greater the genetic distance computed from a set of neutral genes. The specific measure we use, F_{ST}, correlates strongly with historical separation times by construction. The data is available for 26 European populations, mostly overlapping with nation states and, in a few cases, with specific subnational units (the Scottish, Basque, Sardinian, and Lapp populations).[2] Figure A1 displays the philogenetic tree of these 26 European populations, giving a sense of their degree of genealogical relatedness.

The second set of measures of social distance consists of linguistic distance. As in the regional dataset, the main source of linguistic distance data consists of linguistic trees provided in Ethnologue. In contrast to the languages used in the regional data, these refer to languages still spoken contemporaneously. We have separate linguistic tree data on Lithuanian, Latvian, Belarusan, Ukrainian, Frisian, Walloon, Flemish, Czech and Slovak, Occitan, Catalan, Breton, and these languages are therefore matched to the corresponding regions in the fertility dataset (for instance the availability of linguistic data guided considering Walloon Belgium and Flemish Belgium as separate units).[3] For each language, we can count the number of nodes it shares with a reference language, say French. The measure of linguisitic distance is then the maximum number of shared

[2] To exploit the latter variation, we consider Scotland, the Basque Country, Sardinia and Lapland as separate data points of our sample of populations.

[3] For populations with more than one language, such as Switzerland, we matched to the language used by the

nodes (in the case of French, 10) minus the number of common nodes, i.e. the number of different linguisitic nodes between two populations.

Finally, as an alternative measure of linguistic distance, we use data from the field of lexicostatistics, or quantitative linguistics (Dyen et al., 1992). The data are based on basic meanings shared by most societies (concepts such as "mother" and "blood"). Consider a list of 200 meanings. Each language uses a word to convey each of these meanings. For a given meaning, two words are designated as "cognates" if linguists determine that they stem from a common root in a proto-language. For instance "table" in English and "tavolo" in Italian both stem from the Latin root "tabula" and are therefore cognates. For two languages, the percentage of cognate words from within the list then is a measure of the closeness between these languages. We use one minus the percentage of cognate words as a measure of linguistic distance.[4]

These data on social distance were merged with PEFP data on marital fertility levels and transition dates. From periodic observations of the I_g index, PEFP provides estimates of fertility transition dates for 19 countries or nations (Coale and Watkins, 1986, page 39, Table 2.1).[5]. We augment these with additional dates for 20 sub-national regions, also defined as the date when I_g has fallen by at least 10%. These regions were chosen as a function of the availability of any of the three measures of social distance to France for the corresponding geographic units.[6] For these regions, the transition dates were obtained by visual inspection of the PEFP map of fertility transition dates, cross-validated using the raw data on I_g. In total, we have I_g data and marital

majority of the population, in this case Swiss German. Our analysis at the regional level is not subject to this problem, since regions are sufficiently small so that, in the context of Europe, each region can only be matched to one ancestral language.

[4]For non-Indoeuropean languages, which are not covered in Dyen et al. (1992) we coded the percentage of cognate words with French as zero. This was the case for Finland, Hungary, Lapland, and the Basque Country where either Uralic languages or linguistic isolates are spoken.

[5]These countries or nations are Austria, Belgium, Denmark, England and Wales, Finland, France, Germany, Greece, Hungary, Ireland, Italy, The Netherlands, Norway, Portugal, European Russia, Scotland, Spain, Sweden and Switzerland. Figure A2 presents the path of I_g for a selected set of 8 of the 19 countries.

[6]The 20 additional regions / populations under consideration are: Latvia, Catalonia, Walloon Belgium, Flemish Belgium, England, Wales, Freisland, Bretagne, Provence, the Czech Republic, Lithuania, Ukraine, Poland, Lapland, Slovakia, Yugoslavia, Belarus, the Basque Country, Iceland and Sardinia. Combining this list of regions with the aforementioned list of countries, we end up with data on 37 distinct "populations" (England and Wales as well as Walloon Belgium and Flemish Belgium are treated as separate populations, *replacing* the corresponding countries).

fertility transition dates, along with social distance measures, for a set of 37 geographic areas, or "populations".

These populations and the corresponding fertility transition dates are displayed in Table A1. This table shows clearly that France was at the frontier of the process of fertility decline, with an estimated transition date of 1827, and regions closer to France followed before more distant regions. Moreover, there is anecdotal evidence, from Table A1, of a linguistic and cultural channel: a region linguistically and culturally very close to France such as Walloon Belgium experienced the transition to lower marital fertility about 30 years earlier than Flemish Belgium, even though the two regions are part of the same country and are geographically close to France. Sardinia, a genetic isolate, came last in 1934. Transitions also occurred in clusters that are aligned with linguistic cleavages: Countries speaking Germanic and Scandinavian languages tended to transition between 1887 and 1905, while many Romance language countries of Southern Europe featured a later transition: Italy, Portugal, Spain have transition dates between 1913 and 1920.

2.2 Empirical Results Obtained from the Population-level Dataset

Table A2 provides summary statistics for the main variables in the population-level dataset. The average date of transition is 1904 (Panel A). The correlation between the lexicostatistical measure of linguistic distance and that based on language trees is 0.939 (Panel B). On the other hand, the linguistic distance measures are more weakly correlated with genetic distance, with correlations of $0.26-0.27$.[7] This justifies the use of genetic and linguistic distance as alternative measures of social distance. We also see positive correlations between each of the three measures of social distance, and both the fertility transition date and the marital fertility rate averaged over 1911-1940. For instance, the correlation between genetic distance to France and the marital fertility transition date is 0.45.

[7] While linguistic trees and phylogenetic trees look a lot like each other (Cavalli Sforza et al., 1994), various reasons explain why the distance measures themselves are only moderately positively correlated. First, language replacement can weaken the link. For instance, the genetic admixture from the Magyar (Finno-Ugric) conquest of the current Hungarian territory in the 9^{th} and 10^{th} centuries AD only added very little to the gene pool of Hungary. The original language was completely replaced, but Hungarians remain genetically close to other Eastern European populations. Second, there can be gene replacement through migration, with the incoming population adopting the hosts' language. Finally, a split between two languages may have occurred far in the past but only once, or successive splits may have occured in close succession but more recently, weakening the link between linguistic distance based on common nodes and genealogical distance. See Spolaore and Wacziarg (2016) for a further discussion along these lines.

2.2.1 Genetic Distance and the Transition Date

We first explore the determinants of the marital fertility transition date, testing Proposition 1 in the population-level dataset using genetic distance as a measure of social distance:

$$MTD_i = \beta_0 + \beta_1 FST_i^f + X_i'\beta_2 + \varepsilon_i \tag{11}$$

where MTD_i refers to the marital transition date of population i, FST_i^f denotes its F_{ST} genetic distance from the French population and X_i is a vector of controls, mostly geographic distance measures that are potentially correlated with both genetic distance and the transition date.

The results are presented in Table A3. Across columns we successively add more controls. The univariate regression of column (1) reveals a strongly positive effect of genetic distance to France on the date of the fertility transition. In column (2), adding geodesic distance to France as a control weakens the effect only modestly.

The regressions of columns (1) and (2) are displayed graphically in Figures A3 and A4. Both reveal that Lapland and Sardinia are outliers. However, Table A4 reveals that excluding these two regions actually leads to an *increase* in the standardized effect of genetic distance in both specifications of columns (1) and (2). France, the origin of the fertility diffusion, also appears to be an outlier. Excluding it from the data again leads to a slight increase in the standardized effect of genetic distance. Excluding at once France, Lapland and Sardinia leads to a yet larger standardized effect of genetic distance. Thus, the relationship seems robust to outlier analysis.

In the specification of column (3), adding a wide range of geographic controls actually leads to a slightly larger effect than in column (2). In this specification the standardized beta coefficient on FST_i^f is 38.3%, a large effect. To put this number in context, a one standard deviation increase in genetic distance (equal to 69.48) delays the marital fertility transition by almost 8 years. Finally in column (4) we add per capita income in 1820 as an explanatory variable. If indeed the demographic and industrial transitions went hand in hand we would expect income per capita at the beginning of the period to be negatively associated with the date of the fertility transition. We lose 11 data points for which Maddison data on per capita income in 1820 are not available. The latter does not enter significantly, while genetic distance to France continues to have a large magnitude and remains significant at the 10% level despite the much smaller sample. This finding suggests the primary driving force behind the fertility transition was not elevated initial per capita income that might have come from early industrialization, but rather how distant a population happened to be from the population that first adopted the new fertility behavior - the French.

In Table A5, we run the same specification as in equation (11), augmented with genetic and geographic distance from England. The idea is to run a horserace between genetic distance from France and genetic distance from England (the correlation between the two genetic distances is 0.515) - the same rationale as the corresponding exercise in the regional dataset presented in the main text (Section 4.2.1). England was the birthplace of the Industrial Revolution and France that of the fertility transition. If the fertility transition stemmed mostly from economic modernization, we would expect the main axis of diffusion to stem from England. If instead it arose from a process of imitation and diffusion from the first country to undergo the transition, then genetic distance from France should win out in a horse race.

This is indeed what we find in Table A5. Column (1) runs a univariate regression of the fertility transition date on genetic distance from England, which comes out with a positive and significant coefficient. Adding geodesic distance from England (column 2) weakens the effect to almost zero, while adding genetic distance from France flips the sign to negative and insignificant (column 3). Genetic distance from France itself comes out highly significant and with a magnitude roughly similar to that found in Table A3 (40.3%). This picture does not change when adding our long list of geographic controls (column 4): Genetic distance from France wins in a horserace against genetic distance from England.

Overall, these results regarding the determinants of the fertility transition date of lower marital fertility in Europe are consistent with diffusion from France. Societies that are ancestrally closer to the French face lower initial barriers to the adoption of the novel behavior, even through eventually all European populations adopted the new behavior.

2.2.2 Genetic Distance and the Marital Fertility Rate (I_g)

Examining the determinants of the fertility transition date is informative but has limitations. The transition starts before the date specified (without the decline in I_g reaching 10%, a somewhat arbitrary threshold) and continues after the date is reached. An alternative is to directly examine the determinants of the level of I_g. For a cross-section of populations at a point in time corresponding to the main period of the fertility transition, we can hypothesize that I_g itself may be related to genealogical distance from the frontier, as predicted by Proposition 3. We consider the following specification:

$$Ig_i = \gamma_0 + \gamma_1 FST_i + X_i' \gamma_2 + \varepsilon_i \qquad (12)$$

The baseline time period over which we measure I_g in this equation (12) is 1911-1940. The choice of this period is motivated by the availability of I_g data for all 37 populations.[8] This is also a period when the fertility transition was still in full-swing in many countries, as seen in Figure A2.

The results are presented in Table A6. Focusing on the baseline specification of column 3, with all the geographic distance controls, we find a positive effect of genetic distance to France on the level of I_g. The magnitude is large as a one standard deviation increase in genetic distance raises I_g by 57% of its standard deviation. Few of the controls are significant at conventional levels of significance. This finding still holds when controlling for per capita income in 1913, to account for the possible impact of economic modernization on fertility behavior (we find no impact of per capita income itself on marital fertility). These results provide evidence consistent with Proposition 3, in line with those found using the regional dataset.

2.2.3 Linguistic Distance and the Fertility Transition

In addition to genetic distance we now consider the effects of linguistic distance on either the transition dates or I_g. We use the same specifications as in equations (11) and (12), replacing genetic distance to France alternately with the number of different linguistic nodes with the French language ("Français") or the percentage of words not cognate with French.

Results are presented in Table A7 and are consistent with the hypothesis that genealogical distance to the frontier is positively associated with the transition date and the level of I_g. In column (1), with the transition date as the dependent variable, the number of different linguistic nodes with the French language enters significantly (at the 5% level), positively and strongly, with a standardized magnitude of 56.7%, a bit larger than the previously documented effect of genetic distance. In column (2) we consider the lexicostatistical measure of linguistic distance, finding a positive effect that is significant at the 10% level and a standardized magnitude around 45%. The last two columns explore the determinants of I_g averaged over 1911-1940, finding results in line with the previous findings. The effect of the measure based on linguistic trees is still significant

[8] Earlier periods feature fewer observations. If more than one observation on I_g was available for a given population during this time interval, all observations were averaged. On average, there were 2.25 observations on I_g for each population between 1911 and 1940.

at the 5% level, while the effect of the lexicostatistical measure is significant at the 13% level (one should not overemphasize these differences in significance levels given the paucity of observations in the population-level dataset of 37 data points).

In sum, approaching the issue with different dependent variables reflecting the fertility transition and different measures of genealogical distance, we document a positive association between ancestral distance to France and the adoption of marital fertility limitations, consistent with a model of behavioral adoption from the innovator. Moreover, distance to France trumps distance to England, the birthplace of the Industrial Revolution, suggesting that the diffusion of marital fertility limitations followed a diffusion process very different from that of industrialization.

2.3 Data Sources for the Regional and Population-level Datasets

I. Regional Dataset

Category	Variable	Description	Source
Fertility	trandate	Marital Fertility Transition Date	Coale and Watkins (1986)
	i_g	Index of Marital Fertility (Ig)	Coale and Watkins (1986)
	imr	Infant Mortality Rate	Coale and Watkins (1986)
Linguistic Distance	ling1	Mid-19th century regional language (primary)	http://www.muturzikin.com/
	ling2	Mid-19th century regional language (secondary)	http://www.muturzikin.com/
	ethnologuename	Ethnologue language name	Ethnologue (2017)
	diff_nodes_french	# of different nodes with Français	ling1 + Ethnologue (2017)
	diff_nodes_english	# of different nodes with English	ling2+ Ethnologue (2017)
Geographic Distance	landlocked	=1 if region is landlocked	Spolaore and Wacziarg (2009)
	island	=1 if region is on an island	Spolaore and Wacziarg (2009)
	Latitude	Latitude of region centroid	ACME Mapper 2.2
	Longitude	Longitude of region centroid	ACME Mapper 2.2
	distance	Geodesic distance to Paris, km	Authors' calculations using lat. and long.
	dlongitude	Absolute difference in longitudes, to Paris	Authors' calculations using longitude
	dlatitude	Absolute difference in latitudes, to Paris	Authors' calculations using latitude
	contiguitytofrance	Contiguity to France	Author's coding
	mountainrange	=1 if region is barred by a mountain range from France	Authors' coding
	common_water	=1 if area shares a sea / ocean w/ France	Authors' coding
Area and Population	areasqkm	Area sq km	http://www.populstat.info/populhome.html
	population	Population of region at year closest to 1850	http://www.populstat.info/populhome.html
	popdensity19	population density at mid-19th century	http://www.populstat.info/populhome.html
Urbanization	urbanization_1800	urbanization rate using Bairoch 1800	Bairoch, Batou and Chèvre (1988)
	urbanization_1850	urbanization rate using Bairoch 1850	Bairoch, Batou and Chèvre (1988)
Literacy	literacy1880	Literacy rate in 1880	Austria, Belgium, Czechoslovakia, Italy, Yugoslavia: Flora (1983); England and Wales: Stephens (1973); France: Squicciarini and Voigtländer (2016); Germany: Cipolla (1969); Hungary: Toth (1996); Spain: Nunez (1990)

II. Population-level Dataset

Category	Variable	Description	Source
Fertility	trandate	Marital Fertility Transition date	Coale and Watkins (1986)
	i_g	Index of Marital Fertility (Ig)	Coale and Watkins (1986)
Geography	contig	Dummy=1 for contiguity with France	Spolaore and Wacziarg (2009)
	dist	Geodesic Distance (1000s of km) to France	Spolaore and Wacziarg (2009)
	min_av_elev	Average elevation between countries to France	Spolaore and Wacziarg (2009)
	difflat	Absolute difference in latitudes, from France	Spolaore and Wacziarg (2009)
	difflong	Absolute difference in longitudes, from France	Spolaore and Wacziarg (2009)
	landlock_dummy	Dummy=1 if landlocked	Spolaore and Wacziarg (2009)
	island_dummy	Dummy=1 if an island	Spolaore and Wacziarg (2009)
	common_water_	Dummy=1 if shares at least one sea or ocean with France	Spolaore and Wacziarg (2009)
Genetic/ Linguistic Distance	gendist_fst_europe	Fst genetic distance in Europe, from France	Spolaore and Wacziarg (2009)
	language_lex	Language name, lexicostatistical classification	Dyen et al. (1992)
	pc_cognate_french	Percent cognate with French, Dyen et al. data	Dyen et al. (1992)
	lingdist_dom_formula	Linguistic Distance Index, plurality languages, from French	Ethnologue (2017)
	lingdist_weighted_formula	Linguistic Distance Index, weighted, from French	Ethnologue (2017)
	reldist_dominant_formula	Religious Distance Index, plurality religions, from France	Ethnologue (2017)
	reldist_weighted_formula	Religious Distance Index, weighted, from France	Ethnologue (2017)
Income	income1820	Per capita income, 1820	Maddison (2003)
	income1913	Per capita income, 1913	Maddison (2003)

2.4 References for Data Sources

Bairoch, P., J. Batou and P. Chèvre. 1988. *La Population des Villes Europèennes: Banque de Donnèes et Analyse Sommaire des Rèsultats, 1800-1850.* Genève: Droz.

Brinkhoff, T.. *City Population.* https://citypopulation.de

Cavalli-Sforza, Luigi Luca, Paolo Menozzi and Alberto Piazza. 1994. *The History and Geography of Human Genes.* Princeton: Princeton University Press.

Cipolla, C. M. 1969. *Literacy and Development in the West, Vol. 35.* Baltimore: Penguin Books.

Coale, A. J. and S. Cotts Watkins. 1986. *The Decline of Fertility in Europe.* Princeton, NJ: Princeton University Press.

Dyen, I., J. B. Kruskal and P. Black. 1992. "An Indoeuropean Classication: A Lexicostatistical Experiment." *Transactions of the American Philosophical Society*, 82 (5): 1-132.

European Linguistic Map: http://www.muturzikin.com/carteseurope/europe.htm.

Flora, P. 1983. *State, Economy, and Society in Western Europe 1815-1975: A Data Handbook in 2 Volumes.* Campus-Verlag.

Lahmeyer, J. *Population Statistics.* http://www.populstat.info.

ACME Mapper 2.2. http://mapper.acme.com.

Maddison, A. 2003., *The World Economy: Historical Statistics.* Paris, France: OECD Development Center.

Nunez, C. E. 1990. "Literacy and Economic Growth in Spain, 1860-1977." *Education and Economic Development since the Industrial Revolution*, ed. by G. Tortella. Valencia: Generalitat Valenciana.

Simons, G. F. and C. D. Fennig. 2017. *Ethnologue: Languages of the World* (20th edn.) Dallas, TX: SIL International. Online version: http://www.ethnologue.com

Spolaore, E. and R. Wacziarg. 2009. "The Diffusion of Development." *Quarterly Journal of Economics*, 124(2), pp. 469-529.

Squicciarini, M. and N. Voigtländer. 2015. "Human Capital and Industrialization: Evidence from the Age of Enlightenment." *Quarterly Journal of Economics*, 130 (4): 1825-1883.

Stephens, W. B. 1987. *Education, Literacy, and Society, 1830-70: The Geography of Diversity in Provincial England.* Manchester University Press.

Tòth, I. G. 1996. *Literacy and Written Culture in Early Modern Central Europe.* Budapest: Central European University Press.

Table A1 - Dates of the Fertility Transition in a Sample of 37 European Populations

Country name	Date	Country name	Date
France	1827	Austria	1907
Latvia	1865	Hungary	1910
Catalonia	1875	Ukraine	1910
Walloon Belgium	1875	Finland	1912
Switzerland	1887	Poland	1912
Germany	1888	Greece	1913
England	1892	Italy	1913
Scotland	1894	Lapland	1915
Freisland	1897	Slovakia	1915
Netherlands	1897	Yugoslavia	1915
Denmark	1898	Portugal	1916
Sweden	1902	Spain	1920
Norway	1903	Ireland	1922
Bretagne	1905	Russia	1922
Czech Republic	1905	Belarus	1925
Flemish Belgium	1905	Basque Country	1930
Lithuania	1905	Iceland	1930
Wales	1905	Sardinia	1934
Provence	1906		

Source: Coale and Watkins (1986)

Table A2 – Summary Statistics for the Main Population-Level Variables

Panel A – Means and Standard Deviations

Variable	# Obs.	Mean	Std. Dev.	Min	Max
Marital Fertility Transition Date	37	1904	20.099	1827	1934
Marital Fertility Index (Ig) (1911-1940 average)	37	0.461	0.098	0.284	0.670
Genetic distance to France*	37	69.324	69.475	0.000	350
Geodesic distance to France (1000 km)	37	1.176	0.800	0.000	2.494
# of different linguistic nodes with Français	37	7.946	2.571	0.000	10
% not cognate with Français, lexicostatistical measure	37	674.378	268.810	0.000	999

(* Genetic distance was multiplied by 10,000 to facilitate readability of the regression tables)

Panel B – Simple Correlations

	Marital Fertility Transition Date	Marital Fertility Index (Ig)	Genetic distance to France	Geodesic distance to France	# of different linguistic nodes with Français
Marital Fertility Index (Ig) (1911-1940 average)	0.708	1			
Genetic distance to France	0.448	0.521	1		
Geodesic distance to France (1000 km)	0.358	0.422	0.478	1	
# of different linguistic nodes with Français	0.493	0.310	0.270	0.355	1
% not cognate with Français, lexicostatistical measure	0.376	0.231	0.261	0.362	0.939

Note: Summary statistics based on 37 observations: Austria, Basque Country, Belarus, Bretagne, Catalonia, Czech Republic, Denmark, England, Finland, Flemish Belgium, France, Freisland, Germany, Greece, Hungary, Iceland, Ireland, Italy, Lapland, Latvia, Lithuania, Netherlands, Norway, Poland, Portugal, Provence, Russia, Sardinia, Scotland, Slovakia, Spain, Sweden, Switzerland, Ukraine, Wales, Walloon Belgium, Yugoslavia.

Table A3 - Population-Level Regressions for the Transition Date
(Dependent variable: Marital Fertility Transition Date)

	Univariate	**Control for distance**	**Control for geography**	**Control for initial income 1820**
Genetic distance from France	0.130 (2.45)**	0.104 (1.93)*	0.111 (2.26)**	0.107 (2.05)*
Geodesic distance from France (1000s of km)		4.666 (0.88)	4.316 (0.40)	-12.222 (0.55)
Absolute difference in latitudes, from France			-69.611 (0.88)	-52.858 (0.46)
Absolute difference in longitudes, from France			4.782 (0.21)	124.772 (0.54)
1 for contiguity with France			-11.320 (1.09)	-13.818 (1.57)
Dummy=1 if an island			1.167 (0.10)	2.738 (0.20)
Dummy=1 if shares at least one sea or ocean with France			7.862 (1.00)	12.035 (0.57)
Average elevation between countries to France			28.236 (0.70)	45.242 (0.94)
=1 if landlocked			-1.599 (0.22)	-13.797 (0.53)
Per capita income, 1820, from Maddison				-0.007 (0.39)
Constant	1,895.115 (361.65)***	1,891.406 (256.60)***	1,885.426 (131.40)***	1,889.543 (35.65)***
R^2	0.20	0.23	0.30	0.36
N	37	37	37	26
Standardized Beta on genetic distance from France(%)	44.842	35.969	38.298	41.187

Notes:

- t-statistics in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$
- This table presents regressions of the marital fertility transition date on genetic distance from France, geographic distance measures and per capita income in 1820.
- The sample pertains to 37 European populations for which distinct genetic and / or linguistic distance data are available.

Table A4 - Population-Level Regressions for the Transition Date - Outlier Analysis
(Dependent variable: Marital Fertility Transition Date)

	Remove Sardinia & Lapland	**Remove Sardinia & Lapland**	**Remove France**	**Remove France**	**Remove Fr, Sardinia & Lapland**	**Remove Fr, Sardinia & Lapland**
Genetic distance from France	0.296 (3.06)***	0.279 (2.63)**	0.101 (2.71)**	0.093 (2.03)*	0.209 (4.28)***	0.218 (2.56)**
Geodesic distance from France (1000s of km)		1.200 (0.21)		1.502 (0.34)		-0.645 (0.12)
Constant	1,886.620 (253.03)***	1,886.182 (231.88)***	1,899.055 (544.10)***	1,897.792 (436.57)***	1,893.283 (467.30)***	1,893.545 (444.85)***
R^2	0.27	0.28	0.20	0.21	0.23	0.23
N	35	35	36	36	34	34
Standardized Beta (%)	52.404	49.415	45.184	41.711	47.709	49.697

Notes:

- t-statistics in parentheses; * $p<0.1$; ** $p<0.05$; *** $p<0.01$.
- This table presents regressions of the marital fertility transition date on genetic distance from France and geodesic distance. It examines sensitivity to the removal of outliers (Sardinia and Lapland) as well as removal of the frontier country (France).
- The starting sample pertains to 37 European populations for which distinct genetic and / or linguistic distance data are available.

Table A5 – Population-Level Regressions, Horserace with Distance to England
(Dependent variable: Marital Fertility Transition Date)

	Univariate	**Control for distance**	**Horserace, simple**	**Horserace, with geographic controls**
Genetic distance from England	0.152 (2.67)**	0.062 (0.89)	-0.036 (0.67)	-0.125 (1.35)
Geodesic Distance from England, (1000s of km)		6.520 (0.93)	9.776 (0.90)	76.939 (3.08)***
Genetic distance from France			0.117 (2.54)**	0.160 (3.49)***
Geodesic distance from France (1000s of km)			-3.472 (0.30)	-59.688 (1.96)*
Constant	1,895.918 (377.11)***	1,892.873 (279.48)***	1,890.223 (275.00)***	1,893.256 (140.95)***
R^2	0.10	0.13	0.24	0.40
N	37	37	37	37
Standardized Beta - England (%)	32.147	13.079	-7.708	-26.406
Standardized Beta - France (%)			40.302	55.217

Notes:

- Robust t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01.
- This table presents regressions of the marital fertility transition date on genetic distance from France, genetic distance from England, and other controls. All regressions are based on a sample of 37 populations.
- Additional geographic controls in column 4 (estimates not reported) include all those in column 3 of Table A2, i.e. absolute difference in latitudes, absolute difference in longitudes, contiguity dummy, island dummy, landlocked dummy, shared sea/ocean dummy, average elevation along the path to France / England, entered both relative to France and relative to England where applicable.

Table A6 - Population-Level Regressions for Ig, 1911-1941 Period
(Dependent variable: Index of Marital Fertility, Ig)

	Univariate	**Distance control**	**All Geography Controls**	**All Geography Controls plus Income**
Genetic distance from France	0.733 (3.55)***	0.582 (2.39)**	0.802 (3.66)***	0.961 (3.24)***
Geodesic distance from France (1000s of km)		27.360 (1.49)	-17.484 (0.33)	-65.110 (0.43)
Absolute difference in latitudes, from France			-832.471 (1.89)*	-413.682 (0.64)
Absolute difference in longitudes, from France			135.769 (1.33)	-373.766 (0.32)
1 for contiguity with France			-86.143 (1.83)*	-109.189 (1.96)*
Dummy=1 if an island			61.349 (1.34)	98.192 (0.92)
Dummy=1 if shares at least one sea or ocean with France			28.761 (0.57)	12.274 (0.13)
Average elevation between countries to France			225.541 (2.05)*	221.951 (1.76)*
Dummy=1 if landlocked			-133.433 (1.93)*	-93.008 (0.81)
Per capita income, 1913, from Maddison				-0.040 (1.27)
Constant	410.528 (20.13)***	388.782 (15.46)***	412.678 (5.95)***	603.407 (3.71)***
R^2	0.27	0.31	0.51	0.55
N	37	37	37	29
Standardized Beta (%)	52.141	41.429	57.066	72.114

Notes:

- Robust t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01.
- This table presents regressions of the level of marital fertility (Ig) on genetic distance from France, geographic distance controls, and income per capita. The data on marital fertility (Ig) pertains to the 1911-1941 period: if more than one observation was available on Ig for a given population in that period, the available observations were averaged.
- The marital fertility index was multiplied by 1,000 for readability.

Table A7 - Population-Level Regressions Using Linguistic Distance
(Dependent variable: As in the second row)

	Transition Date	Transition Date	Ig 1911-1940	Ig 1911-1940
# of different linguistic nodes with Français	4.432 (2.43)**		13.739 (2.16)**	
% not cognate with French (lexicostatistical measure)		0.034 (1.81)*		0.113 (1.58)
Geodesic distance from France (1000s of km)	22.318 (2.39)**	22.365 (2.10)**	91.735 (1.94)*	93.127 (1.77)*
Absolute difference in latitudes, from France	-139.423 (1.82)*	-146.493 (1.67)	-1,040.387 (2.10)**	-1,081.450 (2.00)*
Absolute difference in longitudes, from France	-22.115 (1.28)	-19.797 (1.12)	-57.001 (0.73)	-49.367 (0.63)
1 for contiguity with France	3.961 (0.55)	1.754 (0.27)	-20.579 (0.41)	-25.314 (0.53)
=1 if an island	-3.678 (0.29)	-2.289 (0.18)	30.500 (0.58)	34.897 (0.68)
=1 if shares at least one sea or ocean with France	9.775 (1.11)	10.800 (1.04)	15.612 (0.27)	20.622 (0.34)
Average elevation between countries to France	6.577 (0.25)	10.842 (0.35)	124.769 (1.31)	135.742 (1.45)
=1 if landlocked	-3.933 (0.52)	-3.304 (0.41)	-147.410 (2.04)*	-145.477 (2.00)*
Constant	1,849.404 (79.45)***	1,860.408 (73.14)***	311.517 (3.57)***	338.945 (3.87)***
R^2	0.43	0.34	0.44	0.41
N	37	37	37	37
Standardized Beta on linguistic distance (%)	56.684	45.080	36.174	31.020

Notes:

- Robust t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01)
- This table presents regressions of the fertility transition date (columns 1-2) or the level of marital fertility (columns 3-4) on measures of linguistic (rather than genetic) distance.
- Ig was multiplied by 1,000 to make the numbers more readable (columns 3 and 4). All regressions are based on a sample of 37 populations. Results do not change materially with the addition of per capita income in 1820 to columns (1) and (2) or the addition of per capita income in 1913 to columns (3) or (4).

Table A8 - Probit for Fertility Transition, with Controls for Literacy, Urbanization, and Population Density, Regional Dataset (Dependent variable: fertility transition status indicator)

	1841	1861	1881	1901	1921
# of different nodes with Français	-0.002 (0.52)	-0.001 (0.36)	-0.074 (6.03)***	-0.030 (1.89)*	-0.000 (0.32)
Geodesic distance to Paris, 1000 km	-0.020 (0.66)	-0.007 (0.50)	0.968 (2.59)***	0.351 (0.63)	0.003 (0.33)
Absolute difference in longitudes to Paris	-1.521 (0.40)	-0.452 (0.31)	-65.929 (2.96)***	-77.814 (2.27)**	-0.161 (0.33)
Absolute difference in latitudes to Paris	0.897 (0.33)	0.523 (0.55)	-112.765 (3.64)***	-79.514 (1.85)*	-0.278 (0.33)
Literacy rate, 1880	0.000 (0.11)	0.000 (0.32)	0.002 (1.11)	0.007 (2.91)***	0.000 (0.33)
Urbanization rate, 1850	-0.010 (0.47)	-0.004 (0.35)	-0.004 (0.03)	0.581 (2.58)***	0.000 (0.25)
Population density in mid-19th century	0.000 (0.31)	0.000 (0.26)	-0.000 (0.11)	0.000 (0.56)	0.000 (0.37)
Pseudo-R^2	0.64	0.70	0.51	0.40	0.52
Standardized effect of linguistic distance from Français (%)	-3.970	-1.024	-98.470	-17.571	-0.005

Notes:

- t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01
- The dependent variable for year t is defined as 1 is a region has undergone the fertility transition by year t (defined as having attained a 10% decline in I_g by date t, as in Coale and Watkins, 1986), zero otherwise.
- The table reports probit marginal effect. The standardized effect is equal to the probit marginal effect multiplied by the standard deviation of linguistic distance to Français, divided by the mean of the dependent variable.
- Regressions are based on a balanced sample of 298 regions from 8 countries: Austria, Czechoslovakia, England & Wales, France, Germany, Hungary, Italy, Spain.
- Unlike in Table 6, there are not enough observations to estimate the relationship for 1941.

Table A9 - Regional Regressions for Ig, with Country Fixed-Effects, Controlling for Minimum Distance to Any Area that Already Transitioned (Dependent variable: Index of Marital Fertility, Ig)

	Period 3 (1851-1880)	**Period 5 (1871-1900)**	**Period 7 (1891-1920)**	**Period 9 (1911-1940)**	**Period 11 (1931-1960)**
# of different nodes with Français	23.001 (11.43)***	20.336 (10.45)***	19.701 (9.98)***	12.750 (6.69)***	7.446 (4.73)***
Minimum linguistic distances to areas that made fertility transition	8.293 (1.60)	13.817 (2.10)**	0.920 (0.09)	160.093 (2.09)**	238.942 (1.10)
Geodesic distance to Paris, Km	0.081 (1.22)	-0.016 (0.27)	-0.097 (1.58)	-0.014 (0.40)	-0.023 (0.84)
Minimum geodesic distances to areas that made fertility transition	-0.018 (0.92)	0.065 (3.13)***	0.247 (7.97)***	0.174 (3.13)***	1.474 (4.70)***
Constant	493.525 (11.96)***	479.853 (11.92)***	416.324 (10.03)***	70.483 (1.32)	186.282 (4.56)***
R^2	0.69	0.63	0.63	0.65	0.66
Number of regions	531	659	675	766	748
Standardized Beta on linguistic distance from France (%)	54.055	45.746	42.274	26.209	17.981
Standardized Beta on minimum linguistic distance (%)	9.734	7.340	0.290	6.986	6.000

Notes:

- t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01
- All columns include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy =1 if region is barred from France by a mountain range, dummy for contiguity to France, dummy if region shares at least one sea or ocean with France, dummy for landlocked region, dummy for region being on an island.
- Country fixed-effects are period-specific due to changing borders.
- Ig was multiplied by 1,000 for readability of the estimates.
- Not enough data to estimate the relationship for Period 1.

Table A10 - Regional Regressions for Ig, with Country Fixed-Effects, Controlling for Infant Mortality Rate, Various Periods (Dependent variable: Index of Marital Fertility, Ig)

	Period 3 (1851-1880)	**Period 5 (1871-1900)**	**Period 7 (1891-1920)**	**Period 9 (1911-1940)**	**Period 11 (1931-1960)**
# of different nodes with Français	19.100 (5.63)***	25.223 (8.78)***	22.044 (7.62)***	8.302 (2.72)***	6.232 (2.43)**
Infant Mortality Rate	215.522 (1.39)	355.760 (2.16)**	780.315 (3.95)***	1,796.989 (5.43)***	1,856.716 (4.65)***
Geodesic distance to Paris, km	0.224 (1.50)	0.113 (0.72)	-0.059 (0.38)	0.011 (0.09)	0.003 (0.03)
Constant	365.045 (5.80)***	357.240 (4.17)***	242.213 (2.78)***	222.280 (3.57)***	190.713 (2.42)**
R^2	0.65	0.61	0.65	0.50	0.45
Number of regions	245	285	291	205	183
Standardized Beta on linguistic distance to Français (%)	43.855	57.520	44.669	26.025	23.840
Standardized Beta on Infant Mortality Rate (%)	6.968	11.298	20.876	42.667	45.430

Notes:

- t-statistics in parentheses: * $p<0.1$; ** $p<0.05$; *** $p<0.01$
- All columns include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy =1 if region is barred from France by a mountain range, dummy for contiguity to France, dummy if region shares at least one sea or ocean with France, dummy for landlocked region, dummy for region being on an island.
- Country fixed-effects are period-specific due to changing borders.
- Ig was multiplied by 1,000 for readability of the estimates.
- Not enough data to estimate the relationship for Period 1,

Table A11 - Cross-regional Regressions for Ig, with Country Fixed-Effects, Controlling for Population Density, Various Periods (Dependent variable: Index of Marital Fertility, Ig)

	Period 1 (1831-1860)	**Period 3 (1851-1880)**	**Period 5 (1871-1900)**	**Period 7 (1891-1920)**	**Period 9 (1911-1940)**	**Period 11 (1931-1960)**
# of different nodes	15.974	19.236	15.741	13.956	8.989	6.902
with Français	(3.94)***	(7.50)***	(6.31)***	(5.21)***	(3.82)***	(3.43)***
Population density,	-0.009	-0.011	-0.015	-0.019	-0.016	-0.012
mid-19th century	(1.46)	(2.74)***	(3.60)***	(4.19)***	(4.01)***	(3.67)***
Geodesic distance to	0.089	0.064	-0.037	-0.055	-0.072	-0.037
Paris, km	(0.32)	(0.82)	(0.58)	(0.81)	(1.79)*	(1.08)
Constant	639.583	546.166	546.746	466.258	147.182	205.418
	(13.56)***	(10.45)***	(10.58)***	(8.42)***	(2.37)**	(3.89)***
R^2	0.70	0.67	0.64	0.62	0.72	0.71
Number of regions	159	431	519	519	552	539
Standardized Beta on linguistic distance to Français (%)	37.479	44.891	35.601	30.242	18.282	16.881

Notes:

- t-statistics in parentheses: * $p<0.1$; ** $p<0.05$; *** $p<0.01$
- All columns include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy =1 if region is barred from France by a mountain range, dummy for contiguity to France, dummy if region shares at least one sea or ocean with France, dummy for landlocked region, dummy for region being on an island.
- Country fixed-effects are period-specific due to changing borders.
- Ig was multiplied by 1,000 for readability of the estimates.

Table A12 - Cross-regional Regressions for Ig, with Country Fixed-Effects, Controlling for Urbanization, Various Periods (Dependent variable: Index of Marital Fertility, Ig)

	Period 1 (1831-1860)	**Period 3 (1851-1880)**	**Period 5 (1871-1900)**	**Period 7 (1891-1920)**	**Period 9 (1911-1940)**	**Period 11 (1931-1960)**
# of different nodes with Français	13.881 (3.19)***	18.766 (6.49)***	14.580 (5.55)***	11.717 (4.31)***	7.759 (3.35)***	5.750 (2.73)***
Urbanization rate in 1850 (Bairoch)	-68.622 (1.55)	-47.950 (1.96)*	-101.998 (4.60)***	-151.058 (6.59)***	-132.705 (7.43)***	-87.188 (5.38)***
Geodesic distance to Paris, km	0.233 (0.76)	0.089 (0.85)	-0.003 (0.03)	-0.054 (0.66)	-0.115 (2.80)***	-0.062 (1.69)*
Constant	655.427 (13.44)***	554.253 (8.53)***	550.971 (8.88)***	488.239 (7.61)***	197.107 (3.26)***	227.973 (4.17)***
R^2	0.66	0.67	0.66	0.67	0.77	0.70
Number of regions	145	326	403	404	437	429
Standardized Beta on linguistic distance from Français (%)	32.924	44.166	33.024	25.471	16.154	14.617

Notes:

- t-statistics in parentheses: * $p<0.1$; ** $p<0.05$; *** $p<0.01$
- All columns include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy =1 if region is barred from France by a mountain range, dummy for contiguity to France, dummy if region shares at least one sea or ocean with France, dummy for landlocked region, dummy for region being on an island.
- Country fixed-effects are period-specific due to changing borders.
- Ig was multiplied by 1,000 for readability of the estimates.

Table A13 - Cross-Regional Regressions for Ig, with Country Fixed-Effects, Controlling for the Literacy Rate, Various Periods (Dependent variable: Index of Marital Fertility, Ig)

	Period 1 (1831-1860)	**Period 3 (1851-1880)**	**Period 5 (1871-1900)**	**Period 7 (1891-1920)**	**Period 9 (1911-1940)**	**Period 11 (1931-1960)**
# of different nodes with Français	13.717 (2.85)***	26.992 (11.85)***	27.289 (12.36)***	24.796 (10.53)***	15.207 (7.21)***	7.598 (4.28)***
Literacy rate, 1880	-0.663 (1.14)	-0.086 (0.25)	-0.728 (2.31)**	-1.103 (3.29)***	-0.636 (1.87)*	-0.461 (1.56)
Geodesic distance to Paris, km	0.283 (0.96)	-0.010 (0.10)	-0.110 (1.42)	-0.147 (1.78)*	-0.092 (1.16)	-0.055 (0.82)
Constant	533.776 (5.31)***	364.634 (3.86)***	378.741 (3.98)***	289.177 (2.85)***	-33.664 (0.33)	104.208 (1.24)
R^2	0.61	0.72	0.67	0.62	0.62	0.57
Number of regions	128	351	408	408	407	386
Standardized Beta on linguistic distance from Français (%)	33.611	63.098	60.697	54.554	37.304	20.848

Notes:

- t-statistics in parentheses: * $p<0.1$; ** $p<0.05$; *** $p<0.01$
- All columns include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy =1 if region is barred from France by a mountain range, dummy for contiguity to France, dummy if region shares at least one sea or ocean with France, dummy for landlocked region, dummy for region being on an island.
- Country fixed-effects are period-specific due to changing borders.
- Ig was multiplied by 1,000 for readability of the estimates.

Table A14 - Cross-Regional Regressions for Ig, with Country Fixed-Effects and Literacy Rate - Restricted to Countries with Complete Regional Literacy Data (Dependent variable: Index of Marital Fertility, Ig)

	Period 1 (1831-1860)	**Period 3 (1851-1880)**	**Period 5 (1871-1900)**	**Period 7 (1891-1920)**	**Period 9 (1911-1940)**	**Period 11 (1931-1960)**
# of different nodes with Français	13.717 (2.85)***	22.240 (5.91)***	17.820 (5.61)***	14.849 (5.06)***	9.645 (3.68)***	5.949 (2.39)**
Literacy rate in 1880	-0.663 (1.14)	0.399 (0.95)	-0.240 (0.62)	-0.517 (1.45)	-0.155 (0.41)	-0.200 (0.56)
Geodesic distance to Paris, km	0.283 (0.96)	-0.220 (1.30)	-0.197 (1.44)	-0.260 (2.07)**	-0.602 (4.64)***	-0.415 (3.38)***
Constant	533.776 (5.31)***	360.216 (4.57)***	429.920 (5.76)***	402.466 (5.83)***	233.739 (2.86)***	291.444 (3.78)***
R^2	0.61	0.68	0.70	0.76	0.79	0.68
Number of regions	128	172	220	220	209	209
Standardized Beta on linguistic distance from Français (%)	33.611	54.673	38.620	31.141	20.607	13.756

Notes:

- t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01
- All columns include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy =1 if region is barred from France by a mountain range, dummy for contiguity to France, dummy if region shares at least one sea or ocean with France, dummy for landlocked region, dummy for region being on an island.
- Country fixed-effects are period-specific due to changing borders.
- Ig was multiplied by 1,000 for readability of the estimates.
- The sample is restricted to regions from Austria, England & Wales, France, Italy, Spain, Hungary, for which data on literacy is available with better regional coverage.

Table A15 - Cross-Regional Regressions for Ig, with Country Fixed-Effects, controlling for Literacy, Density and Urbanization Together, Various Periods (Dependent variable: Index of Marital Fertility, Ig)

	Period 3 (1851-1880)	**Period 5 (1871-1900)**	**Period 7 (1891-1920)**	**Period 9 (1911-1940)**	**Period 11 (1931-1960)**
# of different nodes with Français	21.930 (6.01)***	19.883 (6.16)***	16.232 (4.92)***	9.985 (3.52)***	5.252 (2.02)**
Population density, mid-19th century	-0.008 (1.68)*	-0.009 (1.86)*	-0.010 (2.03)**	-0.008 (1.80)*	-0.007 (1.78)*
Urbanization rate in 1850 (Bairoch)	-22.770 (0.63)	-64.724 (2.04)**	-96.750 (2.98)***	-99.447 (3.69)***	-74.337 (3.02)***
Literacy rate, 1880	-0.327 (0.73)	-0.526 (1.32)	-0.668 (1.64)	-0.236 (0.62)	-0.376 (1.06)
Geodesic distance to Paris, km	0.024 (0.21)	-0.072 (0.79)	-0.143 (1.52)	-0.117 (1.33)	-0.073 (0.89)
Constant	511.384 (6.94)***	520.362 (7.85)***	494.435 (7.30)***	131.021 (1.70)*	223.566 (3.13)***
R^2	0.66	0.66	0.66	0.70	0.61
Number of regions	246	297	297	294	286
Standardized Beta on linguistic distance from Français (%)	52.098	44.842	36.093	23.727	13.645

Notes:

- t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01
- All columns include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy =1 if region is barred from France by a mountain range, dummy for contiguity to France, dummy if region shares at least one sea or ocean with France, dummy for landlocked region, dummy for region being on an island.
- Country fixed-effects are period-specific due to changing borders.
- Ig was multiplied by 1,000 for readability of the estimates.
- Not enough data to estimate the relationship for Period 1.

Table A16 - Cross-Regional Regressions for Ig, with Country Fixed-Effects, controlling for Infant Mortality Rate, Literacy, Density and Urbanization Together, Various Periods (Dependent variable: Index of Marital Fertility, Ig)

	Period 3 (1851-1880)	**Period 5 (1871-1900)**	**Period 7 (1891-1920)**	**Period 9 (1911-1940)**	**Period 11 (1931-1960)**
# of different nodes with Français	27.480 (5.54)***	26.135 (4.58)***	17.752 (3.52)***	9.891 (3.02)***	4.770 (1.58)
Infant Mortality Rate	501.505 (2.49)**	473.437 (2.10)**	1,078.987 (4.28)***	1,567.811 (5.37)***	1,819.834 (4.95)***
Population density, mid-19th century	-0.006 (0.99)	-0.006 (0.89)	-0.003 (0.54)	-0.001 (0.32)	-0.002 (0.45)
Urbanization rate in 1850 (Bairoch)	-77.594 (1.45)	-130.911 (2.28)**	-213.463 (4.32)***	-208.276 (5.10)***	-187.854 (4.98)***
Literacy rate, 1880	-0.800 (1.20)	-1.020 (1.40)	-1.126 (1.76)*	-0.503 (1.07)	-0.516 (1.20)
Geodesic distance to Paris, km	-0.047 (0.25)	-0.080 (0.42)	-0.371 (2.25)**	-0.237 (2.16)**	-0.167 (1.68)*
Constant	425.461 (3.76)***	475.055 (5.28)***	383.935 (4.89)***	242.899 (3.90)***	206.300 (3.50)***
R^2	0.70	0.66	0.75	0.45	0.42
Number of regions	178	178	181	143	143
Standardized Beta on linguistic distance from Français (%)	60.931	59.574	36.326	44.247	26.471

Notes:

- t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01
- All columns include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy =1 if region is barred from France by a mountain range, dummy for contiguity to France, dummy if region shares at least one sea or ocean with France, dummy for landlocked region, dummy for region being on an island.
- Country fixed-effects are period-specific due to changing borders.
- Ig was multiplied by 1,000 for readability of the estimates.
- Not enough data to estimate the relationship for Period 1.
- Due to limited availability of data, samples in these regressions include regions from France, Germany and England & Wales.

Table A17 - Cross-Regional Regressions, German-French Horserace, with country fixed-effects (Dependent Variable: Fertility Transition Date)

	Only distance	Control for all distances	Control for micro-geography	Control for literacy, density, urbanization
# of different nodes with Français	2.143 (3.50)***	2.052 (3.26)***	2.128 (3.36)***	3.311 (4.17)***
# of different nodes with German	-0.193 (0.39)	-0.435 (0.77)	-0.371 (0.65)	-1.142 (1.45)
Geodesic distance to Berlin, km	0.002 (0.63)	0.020 (2.77)***	0.015 (1.86)*	0.058 (2.11)**
Geodesic distance to Paris, km	0.009 (2.69)***	-0.010 (1.36)	-0.006 (0.81)	-0.030 (1.39)
Literacy rate, 1880				0.023 (0.26)
Population density, mid-19th century				0.001 (0.02)
Urbanization rate, 1850				-12.607 (2.02)**
Constant	1,882.652 (258.37)***	1,882.092 (255.43)***	1,876.150 (245.94)***	1,857.814 (138.19)***
R^2	0.71	0.72	0.72	0.77
N	771	771	771	298
Standardized Beta, German (%)	-1.817	-4.089	-3.487	-9.833
Standardized Beta, French (%)	24.289	23.254	24.118	34.267

Notes:

- Robust t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01.
- Country fixed-effects are defined as per 1846 borders.
- Column (2) includes controls for: absolute difference in longitudes to London, absolute difference in latitudes to London, absolute difference in longitudes to Paris, absolute difference in latitudes to Paris.
- Columns (3) and (4) include all the controls in column (2) plus: dummy for contiguity to England, dummy for regions that share at least one sea or ocean with England, dummy for contiguity to France, dummy for regions that share at least one sea or ocean with France, dummy for regions barred by a mountain range to France, dummy for landlocked region, dummy for regions located on an island.
- The broadest sample of 771 regions pertains to the regions of the following 25 countries: Austria, Belgium, Bulgaria, Czechoslovakia, Denmark, England and Wales, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Luxemburg, Netherlands, Norway, Poland, Portugal, Romania, Russia, Scotland, Spain, Sweden, Switzerland, Yugoslavia.

Table A18 - Regional regressions for Ig, horserace with German, with country fixed effects
(Dependent variable: Index of Marital Fertility, Ig)

	(1)	(2)	(3)	(4)	(5)	(6)
	Period 1[a] (1831-1860)	**Period 3[b] (1851-1880)**	**Period 5[c] (1871-1900)**	**Period 7[d] (1891-1920)**	**Period 9[e] (1911-1940)**	**Period 11[f] (1931-1960)**
# of different nodes with Français	18.025 (1.91)*	25.841 (8.83)***	25.078 (8.28)***	26.665 (8.17)***	19.012 (6.32)***	15.403 (6.05)***
# of different nodes With German	2.565 (0.20)	3.424 (0.98)	4.590 (1.25)	10.506 (2.65)***	8.855 (2.49)**	11.362 (3.76)***
Geodesic distance to Paris, km	0.123 (0.46)	0.021 (0.30)	-0.037 (0.61)	-0.005 (0.08)	-0.055 (1.19)	-0.073 (1.89)*
Geodesic distance to Berlin, km	0.065 (0.31)	0.126 (1.80)*	0.193 (2.81)***	0.165 (2.20)**	0.060 (1.15)	0.095 (2.16)**
Constant	563.827 (3.62)***	449.754 (9.42)***	437.218 (8.94)***	310.232 (5.90)***	-8.465 (0.14)	108.168 (2.18)**
R2_B	0.69	0.70	0.62	0.60	0.66	0.66
N_G	184	531	659	675	766	748
Standardized Beta, French (%)	45.424	60.728	56.413	57.217	39.083	37.194
Standardized Beta, German (%)	1.781	6.606	8.736	19.470	15.461	22.105

<u>Notes:</u> t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01
All regressions include additional controls for: Absolute difference in longitudes to Paris, absolute difference in longitudes to Berlin, absolute difference in latitudes to Paris, absolute difference in latitudes to Berlin, dummy for contiguity to France, dummy=1 if region is landlocked, dummy=1 if region is on an island, dummy=1 if region is barred by a mountain range from France.
Country fixed-effects are period-specific due to changing borders.
Ig was multiplied by 1000 for readability of the estimates.
In terms of their 1946 borders, countries to which regions belong are as follows:
(a): 5 countries as follows: Denmark, England and Wales, France, Netherlands, Switzerland.
(b): 20 countries as follows: as in (a) plus: Austria, Belgium, Finland, Germany, Ireland, Italy, Norway, Poland, Russia, Scotland, Sweden, Czechoslovakia, Hungary, Romania, Yugoslavia.
(c): 24 countries as follows: as in (b) plus Greece, Luxemburg, Portugal and Spain.
(d): 25 countries as follows: as in (c) plus Bulgaria.
(e): 25 countries as follows: as in (d).
(f): 24 countries as follows: as in (e) minus Czechoslovakia.

Table A19 - Regional regressions for Ig, with country fixed-effects, with Log Distance Instead of Distance (Dependent variable: Index of Marital Fertility, Ig)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Period 1 (1831-1860)	**Period 3 (1851-1880)**	**Period 5 (1871-1900)**	**Period 7 (1891-1920)**	**Period 9 (1911-1940)**	**Period 11 (1931-1960)**
# of different nodes with Français	**17.004 (5.14)*** **	**23.092 (12.51)*** **	**21.826 (11.37)*** **	**19.921 (9.55)*** **	**12.710 (6.57)*** **	**7.386 (4.58)*** **
Log of Geodesic Distance to Paris	11.363 (1.95)*	12.249 (2.48)**	9.211 (1.75)*	5.806 (1.02)	2.566 (0.50)	2.768 (0.65)
Constant	541.813 (5.24)***	438.704 (9.21)***	417.593 (8.52)***	345.645 (6.57)***	39.579 (0.64)	171.290 (3.53)***
R2_B	0.69	0.69	0.62	0.59	0.65	0.64
# of regions	184	531	659	675	766	748
(# of countries, 1946 borders)	(5)	(20)	(24)	(25)	(25)	(24)
Standardized Beta on linguistic distance from Français (%)	42.852	54.268	49.098	42.746	26.128	17.835

Notes:

- t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01
- All regressions include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy =1 if region is barred from France by a mountain range, dummy for contiguity to France, dummy if region shares at least one sea or ocean with France, dummy for landlocked region, dummy for region being on an island. Country fixed-effects are period-specific due to changing borders.
- Ig was multiplied by 1,000 for readability of the estimates.
- In terms of their 1946 borders, countries to which regions in the sample belong are as follows:
    - Column (1): Denmark, England and Wales, France, Netherlands, Switzerland.
    - Column (2): as in column (1) plus: Austria, Belgium, Finland, Germany, Ireland, Italy, Norway, Poland, Russia, Scotland, Sweden, Czechoslovakia, Hungary, Romania, Yugoslavia.
    - Column (3): as in column (2) plus Greece, Luxemburg, Portugal and Spain.
    - Columns (4) and (5): as in columns (3) plus Bulgaria.
    - Column (6): as in columns (4) and (5) minus Czechoslovakia.

Table A20 - Ig Regressions with Country Fixed-Effects and Additional Controls, Period 5 (1871-1900), with Log Distance Instead of Distance. (Dependent variable: Index of Marital Fertility, Ig)

	(1)	(2)	(3)	(4)	(5)	(6)
	Infant Mortality	Population Density	Urbanization Rate	Literacy	All but IMR	All additional controls
# of different nodes with Français	25.266 (9.00)***	15.676 (6.23)***	14.060 (5.35)***	26.522 (12.03)***	19.524 (6.00)***	25.596 (4.68)***
Log of Geodesic Distance to Paris	11.659 (1.71)*	-0.844 (0.13)	8.622 (1.56)	11.011 (1.99)**	2.823 (0.39)	-2.759 (0.31)
Infant Mortality Rate	389.332 (2.47)**					430.576 (2.19)**
Population density, mid-19th century		-0.015 (2.87)***			-0.007 (1.22)	-0.007 (0.92)
Urbanization rate 1850 (Bairoch)			-93.780 (4.13)***		-65.381 (2.06)**	-125.950 (2.25)**
Literacy rate in 1880				-0.630 (2.03)**	-0.471 (1.20)	-0.979 (1.36)
Constant	294.247 (4.29)***	546.793 (8.73)***	505.671 (7.53)***	308.235 (3.12)***	491.951 (6.68)***	462.323 (4.76)***
R^2	0.61	0.64	0.66	0.67	0.66	0.66
Number of regions (Number of countries, 1946 borders)	285 (6)	519 (18)	403 (18)	408 (8)	297 (7)	178 (3)
Standardized Beta on linguistic distance from Français (%)	57.619	35.455	31.844	58.990	44.033	58.345

Notes:
- t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01
- All regressions include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy=1 if region is barred by a mountain range from France, dummy for contiguity to France, dummy=1 if area shares at least one sea or ocean with France, dummy=1 if region is landlocked, dummy=1 if region is on an island.
- Country fixed effects are defined as per 1886 political borders.
- In terms of their 1946 borders, countries to which regions in the sample belong are as follows:
  - Column (1): Belgium, Denmark, England and Wales, France, Germany, Switzerland.
  - Columns (2) and (3): Austria, England and Wales, Finland, France, Germany, Hungary, Ireland, Italy, Luxemburg, Netherlands, Norway, Poland, Portugal, Romania, Scotland, Spain, Sweden, Switzerland.
  - Column (4): Austria, Belgium, England and Wales, France, Germany, Hungary, Italy, Spain.
  - Column (5): Austria, England and Wales, France, Germany, Hungary, Italy, Spain
  - Column (6): England and Wales, France, Germany.

Table A21 - Regional regressions for Ig, without country fixed-effects
(Dependent variable: Index of Marital Fertility, Ig)

	(1)	(2)	(3)	(4)	(5)	(6)
	Period 1 (1831-1860)	**Period 3 (1851-1880)**	**Period 5 (1871-1900)**	**Period 7 (1891-1920)**	**Period 9 (1911-1940)**	**Period 11 (1931-1960)**
# of different nodes with Français	**21.126 (5.74)*** **	**22.537 (11.62)*** **	**19.041 (10.99)*** **	**16.187 (8.86)*** **	**-0.757 (0.39)**	**-2.788 (1.64)**
Geodesic distance to Paris, km	0.096 (0.43)	-0.198 (4.32)***	-0.167 (4.04)***	-0.136 (3.15)***	-0.008 (0.23)	-0.089 (2.93)***
Constant	501.180 (8.82)***	601.173 (18.95)***	573.934 (18.94)***	429.025 (14.07)***	316.893 (11.04)***	272.614 (10.87)***
R^2	0.63	0.55	0.46	0.45	0.36	0.28
# of regions (# of countries, 1946 borders)	184 (5)	531 (20)	659 (24)	675 (25)	766 (25)	748 (24)
Standardized Beta on linguistic distance from Français (%)	53.239	52.964	42.832	34.733	-1.556	-6.732

Notes:

- t-statistics in parentheses: * p<0.1; ** p<0.05; *** p<0.01
- All regressions include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy =1 if region is barred from France by a mountain range, dummy for contiguity to France, dummy if region shares at least one sea or ocean with France, dummy for landlocked region, dummy for region being on an island.
- The specification does not include country fixed-effects.
- Ig was multiplied by 1,000 for readability of the estimates.
- In terms of their 1946 borders, countries to which regions in the sample belong are as follows:
    - Column (1): Denmark, England and Wales, France, Netherlands, Switzerland.
    - Column (2): as in column (1) plus: Austria, Belgium, Finland, Germany, Ireland, Italy, Norway, Poland, Russia, Scotland, Sweden, Czechoslovakia, Hungary, Romania, Yugoslavia.
    - Column (3): as in column (2) plus Greece, Luxemburg, Portugal, Spain.
    - Columns (4) and (5): as in columns (3) plus Bulgaria.
    - Column (6): as in columns (4) and (5) minus Czechoslovakia.

Table A22 - Ig regressions with additional controls, without country fixed effects (Dependent variable: Index of Marital Fertility, Ig)

	(1) Infant Mortality	(2) Population Density	(3) Urbanization Rate	(4) Literacy	(5) All but IMR	(6) All additional controls
# of different nodes with Français	24.917 (9.17)***	19.143 (9.42)***	20.773 (9.29)***	27.403 (13.22)***	24.119 (9.18)***	28.850 (6.31)***
Geodesic distance to Paris, Km	-0.010 (0.06)	-0.216 (4.99)***	-0.214 (3.89)***	-0.366 (5.18)***	-0.233 (2.95)***	-0.092 (0.49)
Infant Mortality Rate	330.254 (1.93)*					478.539 (2.15)**
Population density, mid-19th century		-0.015 (3.23)***			-0.007 (1.43)	-0.006 (0.89)
Urbanization rate 1850 (Bairoch)			-99.025 (4.29)***		-103.082 (3.19)***	-134.576 (2.36)**
Literacy rate in 1880				-0.736 (2.86)***	0.427 (1.18)	-0.784 (1.17)
Constant	455.870 (7.99)***	534.424 (13.83)***	543.042 (11.62)***	537.720 (5.64)***	497.431 (10.88)***	440.841 (5.21)***
R^2	0.57	0.53	0.56	0.59	0.62	0.66
Number of regions (Number of countries, 1946 borders)	285 (6)	519 (18)	403 (18)	408 (8)	297 (7)	178 (3)
Standardized Beta on linguistic distance from Français (%)	56.823	43.294	47.049	60.950	54.395	65.761

Notes:
- t-statistics in parentheses: * $p<0.1$; ** $p<0.05$; *** $p<0.01$
- All regressions include additional controls for: Absolute difference in longitudes to Paris, absolute difference in latitudes to Paris, dummy=1 if region is barred by a mountain range from France, dummy for contiguity to France, dummy=1 if area shares at least one sea or ocean with France, dummy=1 if region is landlocked, dummy=1 if region is on an island.
- The specification does not include country fixed effects.
- In terms of their 1946 borders, countries to which regions in the sample belong are as follows:
  - Column (1): Belgium, Denmark, England and Wales, France, Germany, Switzerland.
  - Columns (2) and (3): Austria, England and Wales, Finland, France, Germany, Hungary, Ireland, Italy, Luxemburg, Netherlands, Norway, Poland, Portugal, Romania, Scotland, Spain, Sweden, Switzerland.
  - Column (4): Austria, Belgium, England and Wales, France, Germany, Hungary, Italy, Spain.
  - Column (5): Austria, England and Wales, France, Germany, Hungary, Italy, Spain
  - Column (6): England and Wales, France, Germany.

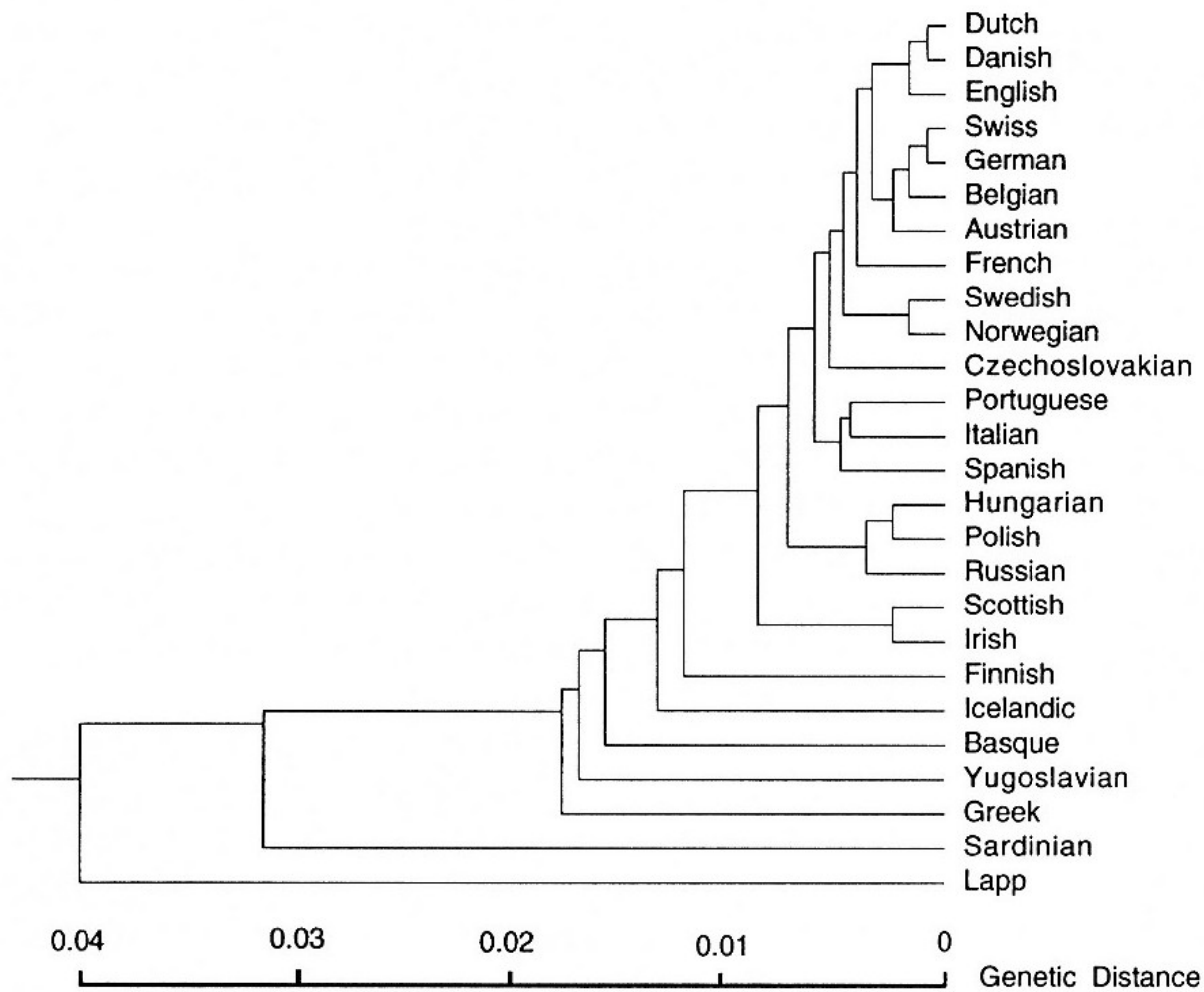

Figure A1 – Phylogenetic Tree of 26 European Populations (Source: Cavalli Sforza et al., 1994)

Figure A2 - The Fertility Transition in Selected Countries

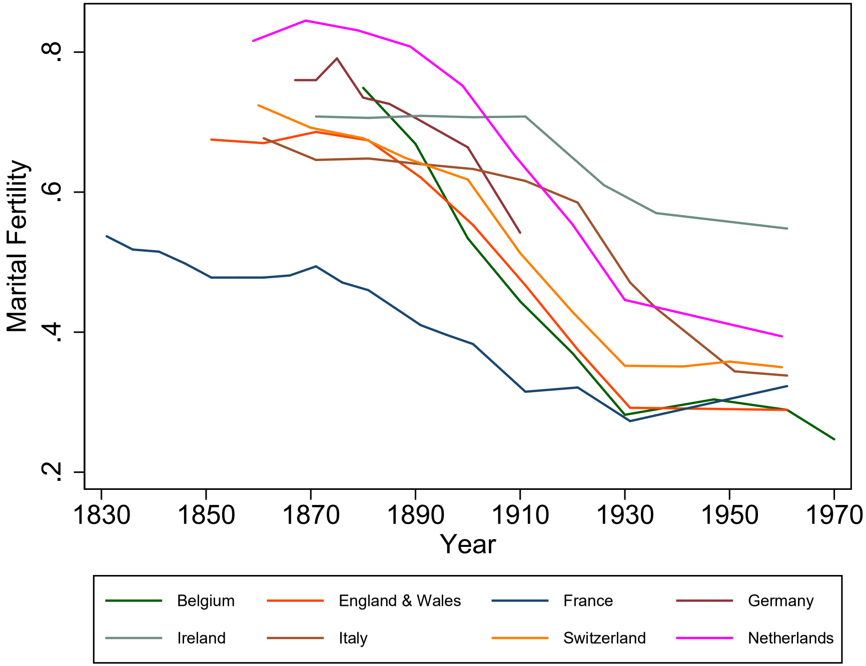

Figure A3 - Genetic Distance from France and the Fertility Transition

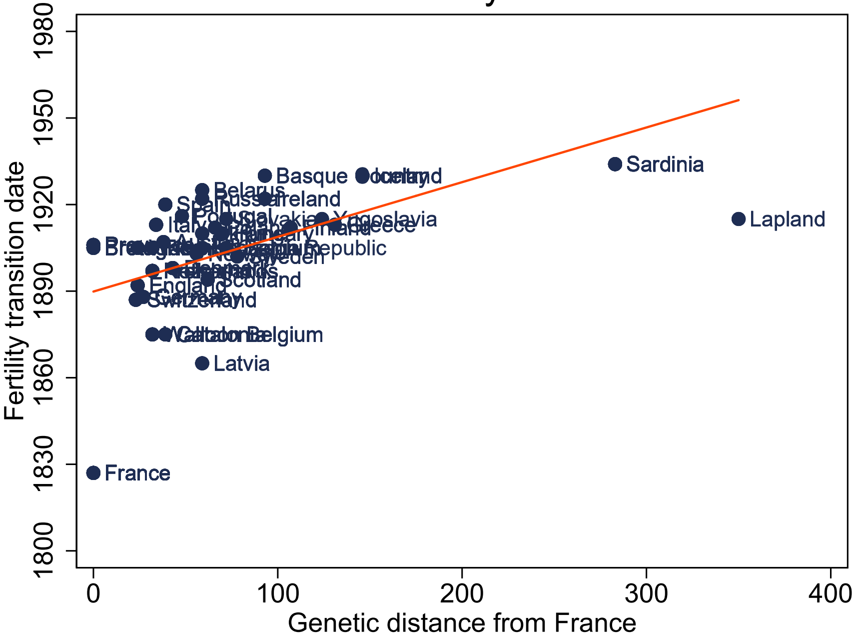

Figure A4 - Genetic Distance from France and the Fertility Transition, controlling for geodesic distance