CAGE

Railways and the European Fertility Transition

CAGE working paper no. 686

October 2023

Carlo Ciccarelli,
James Fenske,
and Jordi Martí Henneberg

WARWICK
THE UNIVERSITY OF WARWICK

UKRI Economic and Social Research Council

RAILWAYS AND THE EUROPEAN FERTILITY TRANSITION

CARLO CICCARELLI[†], JAMES FENSKE[‡], AND JORDI MARTÍ HENNEBERG[§]

ABSTRACT. We show that the spread of the railway network slowed the decline of fertility in Europe during the late nineteenth and early twentieth centuries. We construct novel data on market access across sub-national regions in Europe and use both a panel fixed effects approach and an instrumental variables strategy that leverages variation in market access stemming from access to distant markets. We find that greater market access predicts higher fertility, with a standardized magnitude of 0.14. Consistent with an interpretation that market access increased fertility by raising incomes relative to the returns to child quality and the opportunity cost of childbearing, we show that our results are driven by locations that achieved higher levels of income per capita despite lagging in human capital and female labor force participation.

1. INTRODUCTION

Did railways affect fertility in Europe during the continent's fertility transition? Had Europe's fertility not declined during the late nineteenth and early twentieth centuries – for example, crude birth rates fell in the UK by 33% between 1880 and 1910 – the rising output that came with industrialization would have been absorbed by rising populations, rather than allowing for rising living standards (Crafts and Mills, 2020; Galor and Weil, 1999). Over the same period, the transportation network in Europe transformed dramatically; between 1870 and 1910 alone, Europe's railway network more than doubled in length (Martí-Henneberg, 2013). Railways could have facilitated the fertility decline, for example by increasing urbanization or by spurring economic changes that raised the return to investments in child quality over child quantity. Or they could have raised incomes or the returns to child labor, working to slow the overall decline in fertility. To assess the relationship between railways and fertility in Europe over the course of

[†]UNIVERSITY OF ROME TOR VERGATA
[‡]UNIVERSITY OF WARWICK
[§]UNIVERSITAT DE LLEIDA
E-mail addresses: carlo.ciccarelli@uniroma2.it, J.Fenske@warwick.ac.uk, Marti.Henneberg@geosoc.udl.es.
Date: October 19, 2023.
We are grateful to seminar audiences at the Economic History Society Annual Conference, Gran Sasso Science Institute, Sapienza University of Rome, the University of Exeter, the University of Rome Tor Vergata, the University of Warwick, and Humboldt University for their comments. We are also grateful to Jörg Baten, Kerstin Enflo, Herman de Jong, Emanuele Felice, Jane Humphries, Claudia Olivetti, Joan Roses, Paul Sharp, and Nikolaus Wolf, for generously sharing data with us.

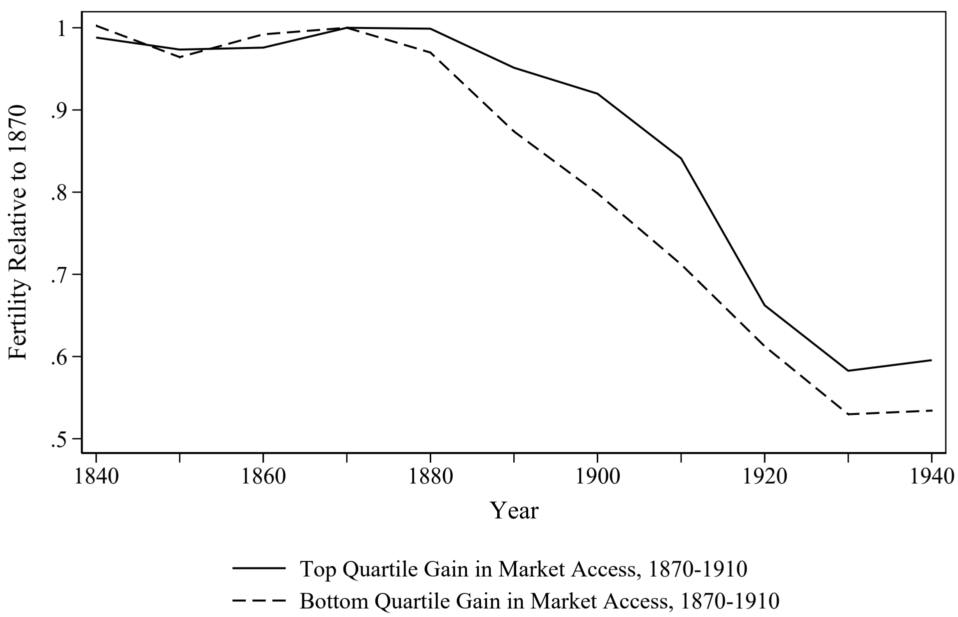
Europe's fertility transition would require panel data on fertility in Europe, measures of changing exposure to railways over time, and plausibly exogenous variation in railway exposure.

In this paper, we use panel data on European fertility, novel measures of market access due to the growing railway network, and variation in market access stemming from access to distant markets in order to assess the relationship between railways and fertility over the course of Europe's fertility transition. We begin with data from the Princeton European Fertility Project. These data cover fertility in an unbalanced panel of locations approximating second-level administrative regions. We combine these data with decadal shapefiles of Europe's railway network between 1840 and 1940 from Martí-Henneberg (2013). We use these shapefiles to compute a measure of market access for each location in each decade and show that, conditional on fixed effects for locations and decades, market access predicts a slowing of the fertility decline. The magnitude of this conditional correlation is such that a one standard deviation increase in market access predicts that fertility is greater by 0.14 standard deviations. To address issues of potential omitted variables and reverse causation not accounted for by our fixed effects, and to justify a causal interpretation of the results, we implement an instrumental variables strategy that exploits variation in market access due to markets more than 500 kilometers away. This strategy confirms our fixed effects results.

To provide intuition for these results, we provide descriptive evidence in Figure 1. In this figure, we show trends over time in the Princeton European Fertility Project's primary measure of fertility between 1840 and 1940 for two subsamples of the data. The first subsample is the locations that were in the top quartile of gains in market access between 1870 and 1910. The second subsample is the locations in the bottom quartile of the increase in market access over the same period. The fertility index is normalized to 1 in both groups in 1870. Both groups start with similar, and flat, fertility trends before 1870. From 1870, however, both groups begin to experience a decline in fertility. In the group that experiences greater gains in market access, however, this decline is less pronounced, trailing the decline in the second group by roughly a decade.

Our results are consistent with the interpretation that children are a normal good and that, if the return to child quality is low, increases in income will raise fertility. As a proximate mechanism, we show that market access predicts greater nuptiality of women aged 20-24. In support of our interpretation, we show that the effects we estimate of railways are greatest in parts of Europe that were ultimately (by c. 1914) more developed, but also those where human capital and female labor force participation lagged. Further, we show that income per capita rises in locations that gain market access. By contrast, we present evidence against alternative interpretations based on the diffusion

FIGURE 1. Trends in fertility by market access



of fertility norms, the direct presence of railways, sectoral change, urbanization, human capital, and political connections.

We subject these results to several robustness checks. Results are robust to tests of the parallel trends assumption, to alternative measures of market access based on alternative constructions of historic populations and of the transportation network, to tests of the exclusion restriction, to alternative functional form and specification assumptions, to exploiting within-country variation for identification, to controlling for possible confounding effects of coal deposits, and to alternative sample constructions, among other checks.

1.1. **Contribution.** We contribute primarily to two broad literatures. The first literature considers fertility transitions broadly, including the fertility transition in Europe (Aaronson, Lange and Mazumder, 2014; Bleakley and Lange, 2009). The broad question of why fertility fell in advanced economies remains an open one. As Guinnane (2011) put it:

> Despite at least one hundred years of academic and official interest in the decline of fertility, this question is not one for which economists have a clear, empirically well-founded explanation.

Much of the debate over Europe's fertility transition revolves around the importance of economic motives such as the quality-quantity trade-off (Becker and Lewis, 1973; Fernihough, 2017; Friedlander, Schellekens and Ben-Moshe, 1991), relative to the importance of cultural factors (Basso and Cuberes, 2017; Beach and Hanlon, 2023; Blanc, 2021;

Blanc and Wacziarg, 2020; Spolaore and Wacziarg, 2022). Alternatively, both broad sets of factors may have acted in tandem (Brown and Guinnane, 2002) – or not at all. More broadly, empirical work has shown that several factors help explain historical fertility declines in advanced countries, including lower infant mortality (Ager, Worm Hansen and Sandholt Jensen, 2018), work opportunities for women (Crafts, 1989; Schultz, 1985), and rising education (Becker and Lewis, 2010; De La Croix and Perrin, 2018; Hansen, Jensen and Lønstrup, 2018; Murphy, 2015; Murtin, 2013). Because railways connected locations in Europe not only to other markets but to other sources of cultural diffusion, or focus on railways allows us to disentangle several economic and cultural channels that could help explain Europe's fertility transition.

The second broad literature to which we contribute is the one considering the effects of railways in economic history (Donaldson, 2018; Fogel, 1964; Hornbeck and Rotemberg, 2021). While there is an extensive literature on how railways have shaped variables such as economic geography and structural transformation, the literature on how railways shaped human capital outcomes such as literacy, health, and fertility is much smaller (Andersson, Berger and Prawitz, 2023; Chaudhary and Fenske, 2023; Tang, 2017; Zimran, 2020). Multi-country studies with meso-level data like ours are even more rare. The closest paper to ours, Guldi and Rahman (2022), finds that market access reduced fertility via specialization in the United States. In comparison to that paper, we focus on Europe, introduce novel market access measures and an instrumental variables approach, and find different results through other mechanisms. Katz (2018) also considers the possible effect of railroad expansion on fertility in the United States. He considers the effect of distance from a county to the nearest railroad, and, following Atack, Bateman and Margo (2010), uses distance to the nearest straight connecting line as an instrument. He finds a positive effect of railroads on human capital and a negative effect on fertility. Forero, Gallego and Tapia (2021) follow, using panel data and a similar identification strategy, and find positive effect of railroads on urban and rural population, but no effects on human capital or fertility in Chile between 1860 and 1920. Overall, the literature on the effect of railroad diffusion on fertility is still in its infant stage and, unsurprisingly, is characterized by a lack of consensus. We add evidence from the European context and consider differences in the impacts of market access rather than direct railway access on fertility.

The rest of this paper proceeds as follows. In Section 2, we provide historical background and outline our empirical strategy. In Section 3, we describe our sources of data. We present our main results in Section 4. We outline our conceptual framework, present evidence in its favor, and rule out alternative explanations of our results in section 5. We present robustness checks in section 6. Section 7 concludes.

2. Historical background and empirical strategy

2.1. **Historical background.**

2.1.1. *The fertility transition in Europe.* During the nineteenth and twentieth centuries, European countries registered, with different timings and intensities, a substantial shift from high to low birth rates. The number of children that a woman would have during her lifetime dropped from five or more at the beginning of the 19th century to less than two at the beginning of the 21st century. This long-term process of decline of 50% or more in the number of children the average woman bears is known as the fertility transition (Guinnane, 2011).

The reference source used to quantify the origins of the fertility transition in Europe is the Princeton European Fertility Project (PEFP), a large-scale research project carried out in the 1960s and 1970s by a team of international researchers and coordinated by Ansley Coale and Susan Watkins (Watkins and Coale, 1986). The PEFP produced a battery of indices on overall fertility, marital and non-marital fertility, and nuptiality for European regions covering a period from the late eighteenth century to the mid twentieth century (see Brown and Guinnane (2007) and Guinnane, Okun and Trussell (1994) for critical reviews). The PEFP scholars stressed the importance of parity-specific limitations in determining the fertility transition. That is, couples change their fertility behavior to limit additional births once the optimal number of children has been reached. This behavior differs from other forms of birth control such as delayed marriage or celibacy existing already in pre-modern demographic regimes. A methodological point that emerged clearly after the PEFP is that regionally disaggregated data are needed to investigate properly the European fertility transition.[1] On the eve of the First World War, when the data allow us to have a rather complete picture at the European level, fertility was relatively low in France, where the transition had started already in the late 18th century, and was also low in England and Wales and in other northern countries. In Southern and Eastern Europe instead, fertility was still relatively high. Considerable within-country heterogeneity was also evident, with Italy representing perhaps the leading example.

With the fertility transition, "for the first time in human history, technological progress led to an elevation in living standard in the long run" (Galor, 2022, p. 85). On the one hand, given the relevance of the topic, the literature on the matter is sizeable (see Beach and Hanlon, 2023; Spolaore and Wacziarg, 2022, and literature therein). On the other

[1] A set of country-specific monographs were published as a by-product of the PEFP. The one on Italy, Livi Bacci (1977), has data at the provincial and district level, as the Italian sources were rich enough to further disaggregate the data at the sub-regional level.

hand, alternative explanations for the fertility transition have been proposed in the literature. Overall fertility is a combination of reproductive behaviour both within and outside marriage, and therefore involves optimal family size and marriage behavior, including the age at marriage. Given this complexity, it is unsurprising that there is no single accepted explanation.

One branch of the literature is rooted in the work of Gary Becker, notably Becker (1960) and Becker and Lewis (1973). This set of studies focuses mainly on the change in economic incentives for having children. Technological progress allowed for more efficient methods of production, increased parental income, relaxed families' budget constraints, and so raised fertility via an income effect. At the same time however, by increasing the demand for education and parental investment in children's education, given the amount of resources available to families, it reduced fertility via a substitution effect (Fernihough, 2017; Galor, 2011; Galor and Weil, 1999; Klemp and Weisdorf, 2019). In the long run, also thanks to extended life expectancy, reduced child mortality, and restrictions to child labor, all of which are factors increasing the net present value of investments in education, the substitution effect prevailed.

Considering the heterogeneous economic conditions both between and within countries in 19th century Europe, there are reasons to suspect that in the short and medium terms, the income and substitution effects may have been roughly balanced, with a possible prevalence of the income effect in some contexts. Livi Bacci (1986) documents, for example, an *increase* in fertility in north-east Italy at the turn of the 20th century that was in contrast with prevailing trends. Livi Bacci shows that this increase in fertility was a consequence of the economic improvement of the rural population during the *Belle Époque* and the consequent improvement in nutritional standards. This better nutrition reduced the incidence of the *pellagra* diseases tied to the poor dietary regime of the rural population.

The effect of technological progress tied to the second industrial revolution affected different segments of the labor force differentially depending on their sectors of specialization, education, and gender. Boserup (1970) argues that the changing composition of the economic structure associated with the transition to modern economic growth often reduced women's participation. Goldin (1995) shows, with a special focus on the United States, that as industrialization increased household incomes, women first retreated from the labor force. Only later, with the closing of the educational gap against men, did they experience a re-engagement in market work. This pattern is often referred to as the "feminization U" (Humpries and Sarasúa, 2012).[2] Several related studies consider the effect overall education and of female education on fertility (Becker, Cinnirella and Woessmann, 2013; Becker and Lewis, 2010; Diebolt, Menard and Perrin, 2017).

Another branch of the literature points instead to the influence of cultural traits and religious factors, and the transmission of new attitudes and norms. Indeed, already the scholars of the PEFP were skeptical about the centrality of economic factors in explaining the fertility transition. More recently, Beach and Hanlon (2023) consider the sharp change in fertility in Britain in 1877 and document the declines in fertility among culturally British households relative to non-British households. They exploit the publicity surrounding the Bradlaugh-Besant trial of 1877.[3] For the first time in England, the positive effects of limiting the size of families reached the people, with a consequent generalized increase in the demand for information about contraception. Spolaore and Wacziarg (2022) use a new index of linguistic distance between European regions to show that the fertility decline resulted from the gradual diffusion of new fertility behaviour from French-speaking regions to the rest of Europe.

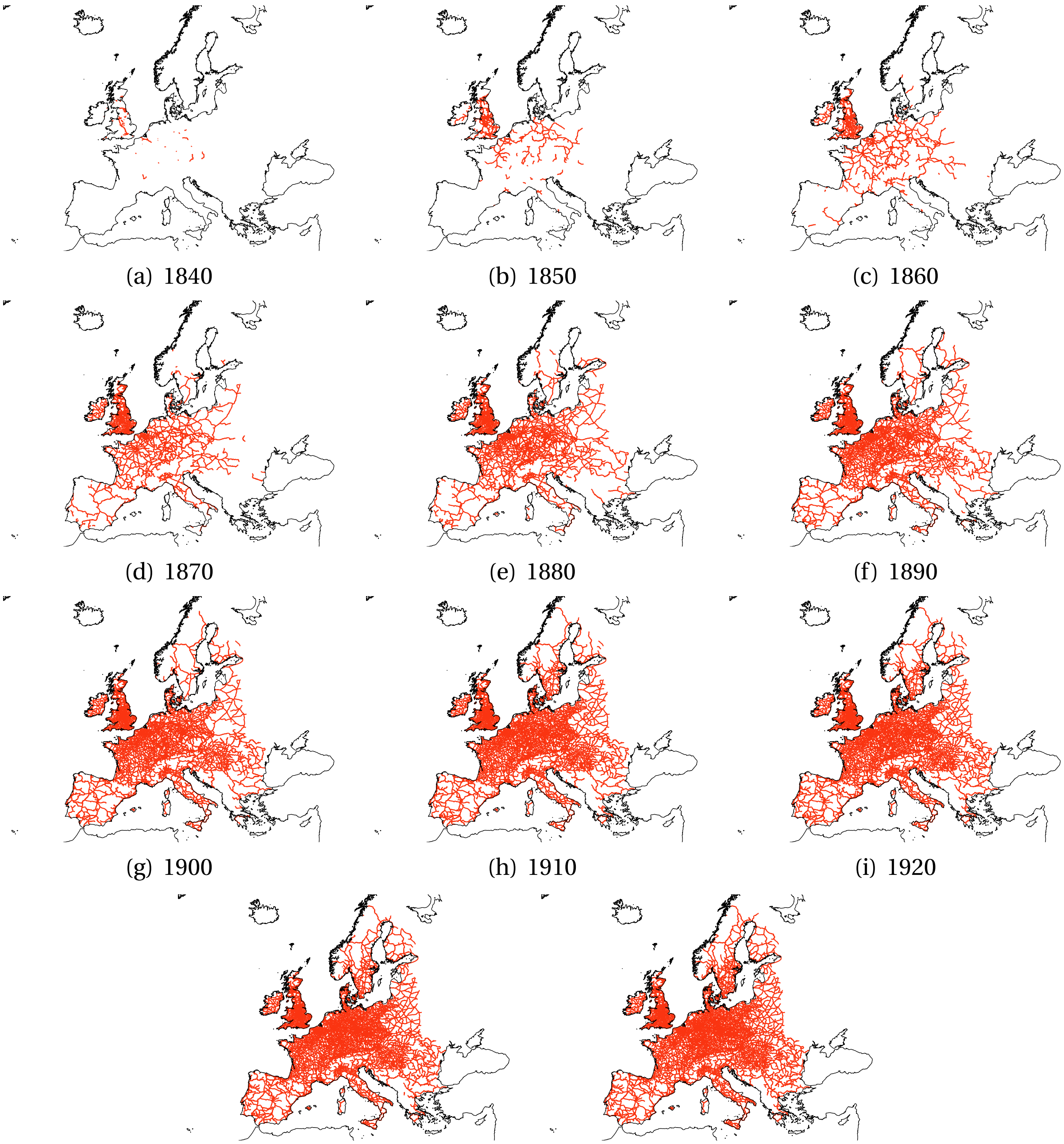
2.1.2. *Railways in Europe.* The beginning of the railway era in Europe is usually dated to 1825, with the opening of the Stockton-Darlington line in England. The first inter-city railway in Europe was the Liverpool and Manchester Railway, which opened in 1830. As Figure 2 illustrates below, between 1830 and 1850, railways spread mostly across Central and Western Europe, with lines opening in Belgium, France, Germany, and The Netherlands.[4] Between 1850 and 1870, railways began spreading into Central Europe and parts of Eastern Europe, crossing regions of the Habsburg Empire. During the 1870s, railways diffused further into Southern Europe (Italy and Spain, but also Portugal) and into Northern Countries (Sweden and Finland, and later on also in Norway). In 1880, the skeleton of the European network, with some 130 thousand kilometers of extension, was completed. In the last decades of the nineteenth century, with the construction of secondary lines and minor trunks filling the gaps, the network became increasingly dense. In 1900, 1920, and 1940, the extension of railways in Europe reached, respectively, some 185, 210, and 230 thousand kilometers. This last figure is very close to the peak reached in the 1960s (Martí-Henneberg, 2013; O'Brien, 1983).

[2]Even during the first industrial revolution, technological progress in the textile sector in England led to job displacement among female spinners as factory owners preferred to employ fewer workers (often men) to operate the machines (Galor, 2022).

[3]Annie Besant and Charles Bradlaugh were prosecuted for diffusing a book explaining various methods of birth control.

[4]We discuss the data sources underlying these maps in Section 3.2.

FIGURE 2. Railways over time



(a) 1840 (b) 1850 (c) 1860 (d) 1870 (e) 1880 (f) 1890 (g) 1900 (h) 1910 (i) 1920 (j) 1930 (k) 1940

European railways were built for both economic and political reasons. In England, the diffusion of railways mostly followed the country's industrialization. Elsewhere in Europe, railways were one of the factors fostering industrial development and ultimately promoting economic change. Political and military reasons, including nation-building as in the Italian and German cases of 1861 and 1870, also lay behind railway construction (Alesina, Giuliano and Reich, 2021). Political and military concerns often led to a break of gauge at national borders, occurring when a line of one track gauge in one nation met a line of a different gauge in a neighboring nation.[5] This change in gauge represented a de facto border effect, in terms of extra costs for passengers and freight that might need to change trains or to other modes of transport. Most European nations eventually adopted a standard gauge of 1,435 millimeters. However, Finland, part of the Russian empire up to 1917, adopted the broad 1,524 mm gauge prevailing in Russia. Portugal and Spain adopted mostly the so called Iberian gauge of 1,668 millimeters. Ireland also adopted a broader gauge of 1,600 millimeters (Martí-Henneberg, 2021).

Geography was a crucial factor influencing the construction of the European network, due to its influence on optimal routes, topographical challenges, and on facilitating economic integration across European regions. When compared to railways in the United States and Soviet Union, the operation of railways in Europe was difficult and expensive in the middle of the twentieth century (Robert, 1964). Cities and stations were close together, necessitating frequent stops, raising fixed costs, and making unmanned stations less feasible. Lines were heavily used, and users' needs were constantly changing.

Several geographical features were crucial. First, while Figure 2 shows that in the early 1880s the skeleton of the European network was essentially built, the rugged terrain of the Swiss Alps presented an obstacle limiting trade and transportation between northern and southern Europe. This constraint was relaxed by Alpine tunnels such as the Fréjus tunnel (opened in 1871, connecting Italy and France), the Gotthard tunnel (opened in 1882, connecting Italy to Germany and passing through Switzerland), and the Simplon Tunnel (opened in 1906 and connecting Italy and Switzerland). Alpine tunnels represent central achievements of the engineers of the time. They were, at the same time, examples of rising international cooperation among sovereign states.

Second, and more generally, crossing mountains required particularly powerful locomotives or, more specifically, locomotives with an adequate power-to-weight ratio (Ciccarelli and Nuvolari, 2015). Technologically backward Southern European countries like Italy, Spain and Portugal were particularly affected by geographical challenges that required engineering solutions. These technical requirements were achieved only gradually. The introduction of electric traction, often adopted together with reduced gauge lines in mountain areas, from the end of the nineteenth century contributed considerably to this goal.

Third, the geography of river valleys such as the Rhine and Danube influenced the placement of railways along these natural transportation corridors, enabling efficient

[5]The gauge is the distance between the rails, measured in millimeters.

movement of goods and people through key European regions. Indeed, steam locomotives, with some 40 years of serving life, made abundant use of water and so proximity of railway lines to rivers was important for that reason too.

Fourth, steam locomotives made abundant use of coal, so that proximity to coal mines was also a factor influencing operating costs. In our empirical work, we will account for the geographic variables that could influence both railway placement by using location-specific fixed effects and by allowing for flexible non-linear time trends predicted by a rich set of geographic controls.

The literature has identified several mechanisms through which the diffusion of railways had an effect on the economy. First, railways expanded markets for goods by connecting previously remote areas to urban centers and ports (Donaldson and Hornbeck, 2016). The reduction in transport costs facilitated the exchange of agricultural and industrial products over long distances, encouraging specialization and trade. Railways contributed to increase cultivable land and improved the transport of agricultural products to distant markets, reducing spoilage and increasing farmers' access to broader customer bases (Atack and Margo, 2011). These effects helped modernize agriculture and made it increasingly market oriented.

Second, the growth of railways encouraged urbanization. People migrated to cities, attracted by job opportunities in expanding industries and locations with improved connectivity. This shift in population distribution contributed to the development of modern urban centers and reshaped urban hierarchies established decades and centuries prior, with individuals from the countryside exposed (often for the first time) to the stimulus stemming from the urban way of living (Atack, Margo and Rhode, 2022; Bogart et al., 2022; Hornung, 2015). It is possible that railways also fueled the "demons of density" (Glaeser, 2011) by, for instance, diffusing epidemic diseases. These negative effects have been explored less in the empirical literature.

Third, railways influenced economically relevant aspects of culture and social capital. Melander (2021) uses a market access approach to investigate the rise of social movements induced by railways in nineteenth century Sweden. Daudin, Franck and Rapoport (2019) use the exogenous variation in transportation costs resulting from the construction of railways to study the convergence of fertility induced by internal migrations and the related diffusion of cultural and economic information in French departments during 1861-1911.

Fourth, and related to the previous point, the construction of railways often led to the development of new towns and industries along their routes, especially in previously isolated or less-developed regions. This process created jobs, increased land values, and spurred regional economic activity. Andersson, Berger and Prawitz (2023) show that inventors residing in connected areas began to develop ideas with applications outside the local economy and sold to firms also connected to the network. Further, the railway industry itself drove technological advancements in areas such as engineering (with increasingly sophisticated locomotives and trains incorporating, for example, compound steam engines and superheating), metallurgy (with steal rails gradually replacing iron rails), and telecommunications (with railway signal traffic increasingly used to prevent accidents). These developments had ripple effects on other manufacturing industries and the broader economy through backward and forward linkages (Ciccarelli and Nuvolari, 2015). From a practical point of view, estimating the impact of railway diffusion on the European economy at the regional level is a difficult task, given the data limitations. However, in our empirical work, we provide evidence of a positive effect of an increase in market access induced by railways on population and per capita GDP.

2.2. **Empirical strategy.** Our data will consist of an unbalanced panel of locations in Europe observed at a decadal frequency between 1840 and 1940. These units are similar to second-level units in Eurostat's Nomenclature of Territorial Units for Statistics system – so-called "NUTS 2" units. Examples in our data include Lazio in Italy, Warwickshire in England and Wales, and Le Rida in Spain. Our baseline estimating equation will be a two-way fixed effects specification that we estimate using ordinary least squares (OLS):

$$Fertility_{ld} = \alpha + \beta \ln(MarketAccess)_{ld} + x_l' \eta_d + \delta_l + \eta_d + \epsilon_{ld}. \tag{1}$$

In equation (1), $Fertility_{ld}$ is a measure of fertility in location l in decade d, where $d \in 1840, ..., 1940$. Generally, this will be "Total Fertility," or I_f, an index we elaborate on below. This index is a ratio between 0 and 100 of fertility in that location and decade to fertility in a reference population that has one of the highest fertility rates ever recorded. α is a constant. Our primary independent variable, $\ln(MarketAccess)_{ld}$, is also described below, and is a measure of how connected location l is to all other locations in the data in decade d. In our baseline estimation, we will construct $\ln(MarketAccess)_{ld}$ so that it only varies over time for a given location due to the expansion of the railway network. β is our coefficient of interest.

To account for confounding factors that may correlate with both market access and with fertility, we employ a number of approaches. First, we employ a two-way fixed effects approach in which we control for both fixed effects for location (δ_l) and decade (η_d).[6] Fixed effects for location will absorb all variation in observable and unobservable

[6]Brown and Guinnane (2007) stress the importance of these fixed effects in drawing inference from the same data source that we use on fertility.

variables that are constant across time for a given location, such as the sincerity of initial religious attitudes towards fertility. Fixed effects for decade, similarly, will absorb all variation in observable and unobservable variables that affect all locations in a given decade evenly, such as the invention of new ideas towards fertility. Second, we will account for several time-invariant controls in x_l. As discussed below in Section 3.3, these controls will mostly be geographic variables and, because they are time-invariant, we will interact them with decade fixed effects η_d. This procedure is equivalent to controlling for arbitrarily flexible time trends predicted by these observable confounders.

A causal interpretation of equation (1) depends, then, on the assumption that there are no time-varying confounders in the error term ϵ_{ld} that are correlated with $Fertility_{ld}$ and $\ln(MarketAccess)_{ld}$ once the flexible time trends predicted by x_l have been removed from the data. If equation (1) is to be interpreted in terms of a potential outcomes framework, rather than structurally, a causal interpretation would also depend on the assumption that, for all dosages of $\ln(MarketAccess)_{ld}$, the average change in $Fertility_{ld}$ over time for all units – had they been assigned that dose – would be the same as the average change in $\ln(MarketAccess)_{ld}$ over time for all units with that same dosage of $\ln(MarketAccess)_{ld}$. Callaway, Goodman-Bacon and Sant'Anna (2021) call this assumption "strong parallel trends," and it is stronger than the corresponding parallel trends assumption in a conventional difference in difference with binary treatment. We will validate these assumptions by showing below that our estimates are not sensitive to the inclusion of x_l as covariates, and that future market access, conditional on current market access, does not predict current fertility. We further show estimates in long differences.

It remains possible, however, that time-varying unobserved variables could confound the relationship between $\ln(MarketAccess)_{ld}$ and $Fertility_{ld}$. For example, a local rise in manufacturing productivity might attract railway development and raise the returns to child labor. To further support a causal interpretation of our results, then, we implement an instrumental variables (IV) strategy. In particular, we compute our instrument $\ln(DistantMarketAccess)_{ld}$, a measure of market access based only on access to markets at a distance of 500 kilometers or more from location l. We elaborate on this variable below, and use it as an instrument for $\ln(MarketAccess)_{ld}$. This IV approach is similar to methods used by both Chan (2023) and Donaldson and Hornbeck (2016), and isolates variation in market access due to markets far away from a location of interest, thereby reducing the scope for local unobservables to confound the relationship between market access and fertility.

In our baseline analysis, we account for possible serial correlation by clustering standard errors by location. For robustness, we report standard errors adjusted for spatial dependence.

3. Data

In this section, we describe our data sources. In section 3.1, we outline our sources and measures of fertility. In section 3.2, we present our sources of railway data and how we use them to compute measures of market access for the locations in our data. In section 3.3 we explain our how we measure our control variables. We describe other data that we use in our mechanisms analysis and robustness checks later in the paper, as these sources are introduced.

3.1. **Fertility.**

3.1.1. *Fertility.* We obtain data on fertility from the Princeton project on the decline of fertility in Europe – otherwise known as the Princeton European Fertility Project (Watkins and Coale, 1986). A collaborative endeavor during the 1960s and 1970s between historical demographers from several countries, this project resulted in eight books published by Princeton University Press, as well as several other articles and chapters. These data have been used previously in economics (e.g. Spolaore and Wacziarg (2022)), and cover several hundred European "provinces and smaller districts" between 1787 and 1970.[7]

The main fertility measure we use is the ratio of births observed in a given population in a given year to the fertility of a high-fertility reference population – the Hutterites during the years 1921-1930. The Hutterites are an Anabaptist sect found predominantly in the American Great Plains and the Canadian prairies and, while their fertility is not the highest possible biologically, no observation in our data exceeds 70% of their index rate.

In particular, the fertility index I_f that we refer to throughout as "Total Fertility" is computed by Watkins and Coale (1986) as:

$$I_f = \frac{B_f}{\sum_a f_a h_a} \tag{2}$$

In equation (2), B_f is all births in a population. a is an age bin (e.g. 25-29). f_a is the number of women in age bin a. h_a is the Hutterite fertility rate in age bin a. That is, I_f is a fertility index that adjusts for the age structure of the population, and so is an improvement over the Crude Birth Rate as a measure of fertility. The data also report analogous indices of marital fertility (I_g) and non-marital fertility (I_h). In order to make

[7]Data were downloaded from `https://opr.princeton.edu/archive/pefp/`

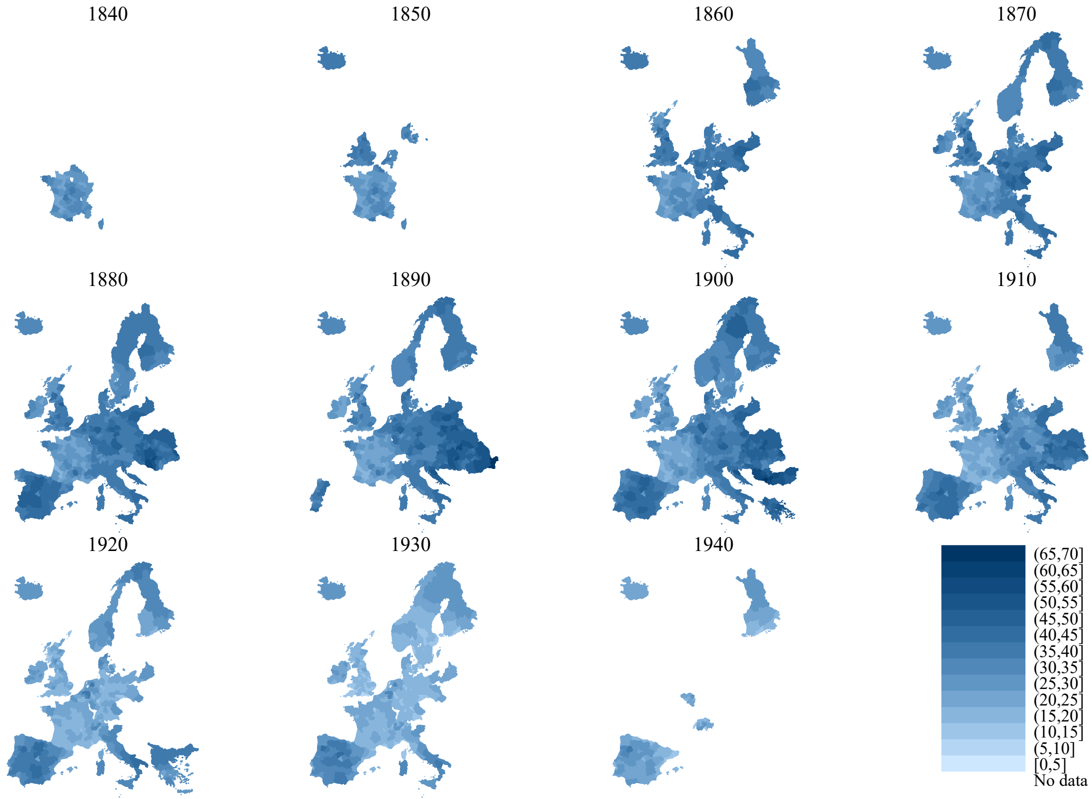
our regression coefficients legible, we multiply the original I_f index and any other fertility measures by 100, so that their feasible range is from 0 to 100, rather than from 0 to 1.

The original data are available as an extremely unbalanced panel that we collapse to a still unbalanced decadal panel. In Figure B.1, we show the availability of the data for each of the countries covered by the Princeton European Fertility Project. There are differences in the range and frequency of dates covered by the project. As examples, France has data at a more than decadal frequency and Spain has data from prior to the nineteenth century, while Albania only appears once after 1850. Further, the data are reported in three sometimes overlapping series. Series 0, Series 1, and Series 2 report fertility data for different sets of administrative boundaries, with Series 0 corresponding roughly with administrative divisions in 1900 and the other series corresponding with later years. To harmonize the data and to avoid double-counting observations, we use only Series 0 observations. These merge almost one for one into other data sources used in this project, and are the most widely available of all three series. Because we do not have railway data on Russia, we exclude it from the sample.

In order to merge the fertility data with other sources, we then map the locations in Series 0 into the locations from the Max Planck Institute for Demographic Research's polygon shapefile for 1900 (MPIDR, 2013). It is these polygon locations that, each decade, we use as units of observation in our regressions. While the fertility and shapefile sources can mostly be merged using a simple a one-to-one correspondence, there are exceptions. First, there are cases where multiple locations in the map correspond to one location in the fertility data. For example, the North Peloponnese region of Greece in the map is split into three (Achaia, Corinthia and Argolis, and Ilia), while it is undivided in the fertility data. Second, there are cases where multiple locations in the fertility data correspond to one location in the polygon map – the Danish region of Frederiksborg in the map, for example, encompasses both Frederiksborg and Hovedstaden in the fertility data.

To collapse the fertility data into a decadal panel at the location level, we average over observations within a decade for each polygon location in order to compute a value for that location in that decade. For example, we average over values from 1871 and 1875 to construct values of I_f for each location in Germany for the decade of the 1870s, while for Ireland we use the only available year from that decade – 1871. For cases in which several map polygons correspond to one unit in the fertility data, this implies assigning the same fertility values to all polygons in a decade. Where one polygon corresponds to several observations in the fertility data, it implies averaging over these observations to construct a single measure for the polygon location.

FIGURE 3. Fertility over time



In Figure 3, we present maps of I_f in our decadal panel of administrative units as mapped by MPIDR (2013). Several patterns are apparent. First, in the early years of out sample, the data are quite incomplete – in 1840, for example, only France appears. Over time, however, the sample comes to encompass the bulk of Western, Central and Southern Europe, particularly during the critical years from 1870 to 1910. Second, the decline in fertility over time is visible in the sequence of maps. France begins with lower fertility rates than other portions of Western Europe, but by 1910 fertility has declined markedly in several countries relative to its value in 1870.

3.2. **Railways.** In order to construct a measure of market access for each location in each decade, we begin with a polyline shapefile of railways from Martí-Henneberg (2013). He constructs a polyline shapefile of active railways in Europe, excluding the territory of the Soviet Union and Ottoman Empire, every decade from 1840 to 2010. We present maps of the expansion of this railway network between 1840 and 1940 in Figure 2.

While there are many ways to measure how a location is exposed to the railway network, many simple measures – such as a dummy for connectedness or the density of railways per unit area – will not capture the sum of the supply and demand forces brought by railway connection. For that reason, we use a conventional measure of the importance of transportation infrastructure: market access. Market access measures the connectedness of firms and consumers in one location to the firms and consumers in all other locations, scaled down by the costs of reaching these other locations. In several models of spatial equilibrium, this variable is a sufficient statistic for the importance of the transportation network (Redding and Sturm, 2008; Redding and Rossi-Hansberg, 2017).

Our specific approach to computing market access draws heavily on both Donaldson and Hornbeck (2016) and on Jedwab and Storeygard (2022). In particular, we compute market access ($MarketAccess_{ld}$) of location l to other markets l' in decade d as:

$$MarketAccess_{ld} = \sum_{l'} \frac{P_{l'd}}{\tau_{ll'd}^{\theta}} \tag{3}$$

In equation (3), $P_{l'd}$ is the population of location l' in decade d. θ is the trade elasticity. $\tau_{ll'd}$ is the travel cost from the centroid of l to the centroid of l' in decade d. Our instrument, distant market access, excludes l' within 500km of l when making this calculation.

In order to compute $P_{l'd}$, we obtain historical population from the History Database of the Global Environment (Klein Goldewijk, Beusen and Janssen, 2010). This source originally records population counts on a raster grid. We sum over raster points within a location polygon to estimate the population of that polygon. In order to avoid problems of endogeneity, we use population in 1830 to compute market access in all years. This approach implies that we capture changes over time due only to changes in travel costs, which we will allow to vary only due to changes in the railway network. For robustness, we will use contemporaneous values of $P_{l'd}$ that change over time and alternative sources of data on population.

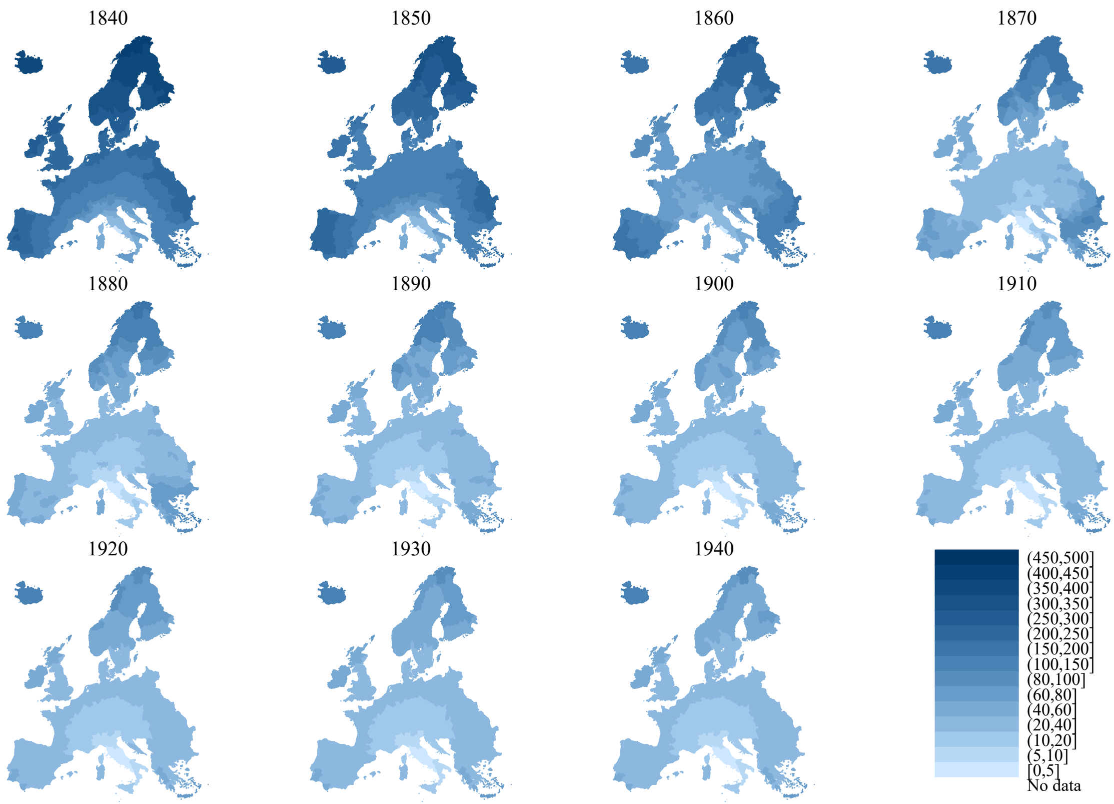
We select a value of 8.22 for θ in our baseline, following Donaldson and Hornbeck (2016). We show robustness, however, to a range of values from 1 to 12.86. It is not conventional in studies that employ market access to allow θ to change over time.

To compute τ, we estimate the time needed to travel between locations l and l' in decade d by the fastest available path. We proceed in several steps. First, we construct a set of grid cells $0.1°$ in latitude by $0.1°$ in longitude that covers both land and water in the region around Europe.[8] Second, to compute the time needed to cross a grid cell by means other than rail, we use the Özak (2010, 2018) Human Mobility Index. This index is based on walking speeds across different types of terrain and on traditional methods of seafaring. Third, we set the speed of crossing a grid cell by rail as 60 kilometers per

[8]For computational feasibility, this grid does not cover the whole globe, but is much wider than Europe. In particular, the grid spans from 33 to 83 degrees of latitude, and -25 to 75 degrees of longitude.

hour for cells that are intersected by a rail line in a given decade.[9] Treating travel times as travel costs, we then use Dijkstra's algorithm to compute the fastest route between the centroid of l to the centroid of l' and its associated travel time. We show robustness below to alternative parameterizations and the inclusion of alternative means of travel, such as roads and steamships. To give intuition for τ, we show in Figure 4 how the travel time to Rome – or more properly, Lazio – changes over the sample period. As is clear from the figure, travel times fall over the nineteenth and early twentieth centuries, often by an order of magnitude.

FIGURE 4. Hours to Rome over time

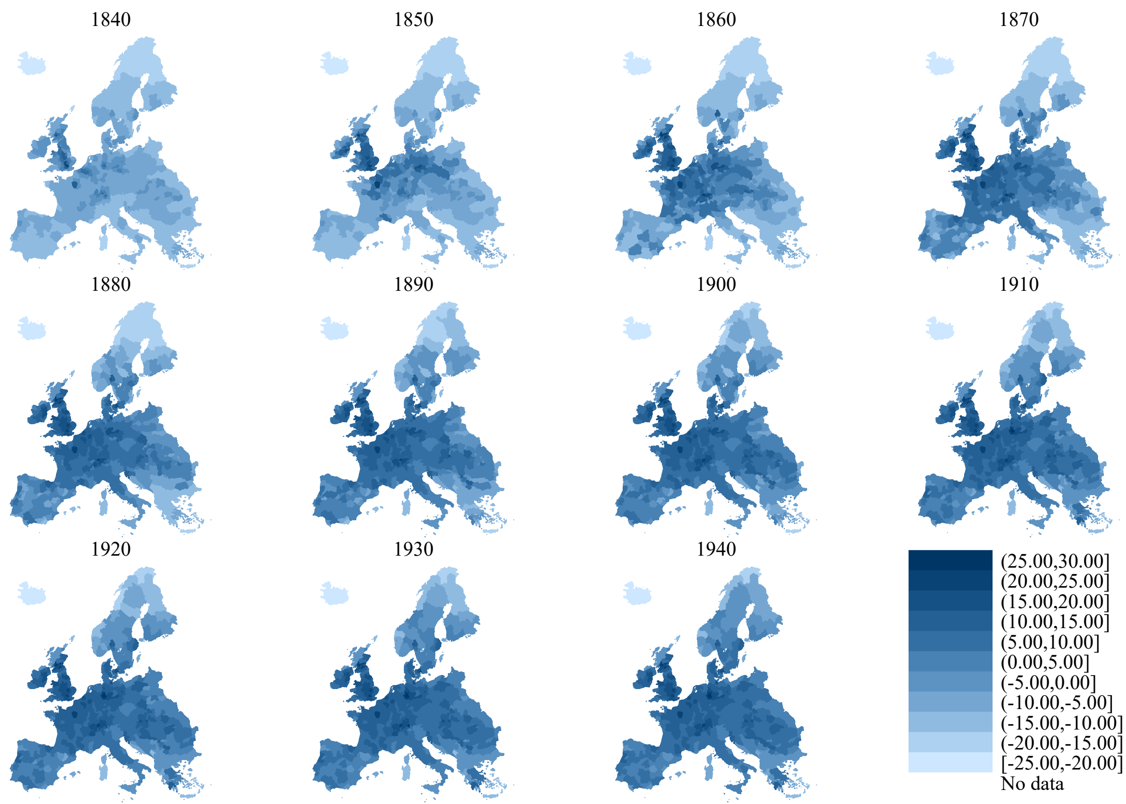


In Figure 5, similarly, we show the evolution of the natural logarithm of market access over time for the locations in our sample. Because we fix $P_{l'd}$ equal to population in 1830 in our baseline, changes over time are due solely to the expansion of the railway network. Market access in 1840 is largely a function of proximity to large population centers, though the importance of England's railway network for the market access of

[9]The speed of 60 kilometers per hour is indicative of the speed reached by trains by the mid-nineteenth century (see, for example, Nicolas (1973)). This speed is at the midpoint of several disparate estimates in the literature (Lardner, 1850; O'Brien, 1983; Schivelbusch, 1978). We show robustness to alternative railway speeds below.

locations between London and Liverpool is apparent. As the railway expands, market access rises first in Northwestern Europe before rising more generally. Regions that increase most in market access are not simply those that gained access to rail, but rather those that gained greater access to large population centers, both directly and indirectly.

FIGURE 5. ln Market access over time



3.3. **Controls.** In our baseline estimation, we include several geographic controls that we interact with decade fixed effects. Latitude and longitude we compute ourselves using the centroids of the polygons in MPIDR (2013). We compute coast distance as the minimum distance from each polygon to the polyline shapefile of coastlines from Natural Earth Data.[10] We also compute a dummy for the presence of a river by intersecting the polyline shapefile of rivers from Natural Earth Data with the polygons in MPIDR (2013). We take altitude from the World Digital Elevation Model, averaging over raster cells within a polygon. For population density in 1830, we use Klein Goldewijk, Beusen and Janssen (2010) as a source, averaging over raster cells in a polygon. We compute area directly using the MPIDR (2013) shapefile. As an overall measure of land quality, we obtain data on the Galor and Özak (2015, 2016) caloric suitability index, averaging

[10] `https://www.naturalearthdata.com/`

over raster cells within a polygon once again. For specific crops, we perform a similar procedure to measure the specific suitability of each location for barley, maize, rye, oats, and wheat using data from the from the FAO-GAEZ project.[11] For average precipitation, we use raster data from WorldClim,[12] for which the underlying source is the Climatic Research Unit at the University of East Anglia. We obtain raster data on ruggedness from Amatulli et al. (2018). We provide maps of these variables in Figures G.1 through G.15 in the appendix. We discuss other control variables included in robustness exercises as we introduce them.

Summary statistics for our principal outcomes, independent variables, and controls are presented in Table B.1.

4. Results

TABLE 1. Results: fixed effects

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.158***	0.125***
	(0.032)	(0.030)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

In Table 1, we present our OLS estimates of the two-way fixed effects equation 1. Column (1) excludes the time-invariant controls interacted with the time fixed effects. Column (2) includes these controls. Our estimates of β are significant and positive in both columns. The coefficient magnitudes indicate that a one standard deviation increase in market access predicts greater fertility by approximately 0.14 standard deviations.[13] We present a visual representation of column (1) as a binned scatterplot in Figure C.1.

As an alternative way to think about magnitudes, consider the counterfactual fertility in a given location l were market access in decade d to remain as it was in 1830. Call this

[11]https://gaez.fao.org/
[12]https://www.worldclim.org/
[13]$\frac{7.91\times 0.158}{8.93}$

counterfactual I_{ld}^A. Denoting the observed fertility index I_f for location l in decade d as I_{ld}, we compute counterfactual fertility as:

$$I_{ld}^A = I_{ld} - \hat{\beta}(\ln MarketAccess_{ld} - \ln MarketAccess_{l,1830}) \tag{4}$$

In Figure C.2, we plot the mean in each decade of counterfactual fertility I_{ld}^A as a fraction of actual fertility I_{ld} in our data, averaged over locations l. This counterfactual suggests that, on the eve of the First World War, fertility would have been 8% lower in Europe had market access remained at its 1830 level.

In Table 2, we present our IV estimates of equation (1). Column (1) again excludes the time-invariant controls interacted with the time fixed effects, while Column (2) again includes these controls. We show equivalent first-stage results in Table C.1. Our estimates of β remain positive and significant across columns, and are now approximately 30% larger than our fixed effects estimates. This coefficient inflation is consistent with the attenuation bias of the OLS estimates that would be expected with measurement error in market access. The Kleibergen and Paap (2006) F statistics exceed 500; not only are weak instruments unlikely to create bias in this setting, the strong predictive power of the instrument is evidence that market access in Europe during this period was not merely a reflection of connection to nearby markets in the local vicinity, but also of access to more distant markets through the entirety of the railway network.

TABLE 2. Results: instrumental variables

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.206***	0.180***
	(0.044)	(0.044)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes
KPF	539.7	535.4

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

5. Mechanisms

In this section, we outline a conceptual framework that accounts for our main results and discuss the predictions of this framework in data. We then present empirical results consistent with this framework and discuss alternative explanations of our results for which we find contrary evidence or for which the evidence is too sparse to draw conclusions.

5.1. **Conceptual framework.**

We begin with the simple observation that, if children are a normal good, an increase in income would be expected to increase fertility (Ashraf and Galor, 2011; Black et al., 2013). Relatedly, gains to trade that stem from the expansion of the railway network can increase fertility if they encourage specialization in goods intensive in unskilled labor (Galor and Mountford, 2008). One proximate mechanism for this increase is that higher incomes allow couples to marry earlier (Weir, 1984), though if higher incomes increase fecundity or lead already-married couples to reconsider their fertility choices, both marital and non-marital fertility may rise.

If an income channel were to underpin our results, we should also be able to identify an increase in incomes per capita in response to railway-driven gains in market access. We should further find that our results are driven by locations in Europe that achieved relatively high levels of development by the eve of the First World War.

Income is not, however, the only channel through which railways can be expected to affect fertility. And indeed, there are mechanisms through which rising incomes could reduce fertility. One such channel is increases in the returns to child quality. If railways led to income gains and structural changes that increase the returns to skill, these effects would provide an incentive for parents to substitute away from investments in child quantity and towards investments in child quality (Becker and Lewis, 1973; Galor, 2022). While it is possible for this channel to be present in our data, our main results suggest that, on average, it was less quantitatively important than other channels. We should find, then, that our results should are strongest in locations where literacy and numeracy lagged despite rising market access.

A second such countervailing channel is increases in female labor participation. If market access created new opportunities for women's work, for example in factories, it would increase the opportunity cost of children (Galor and Weil, 1996; Jensen, 2012). Such processes are often thought to have been an important factor in the fertility decline in developed countries (Guinnane, 2011; Voigtländer and Voth, 2013; Wanamaker, 2012). Alternative views, however, stress that women may indeed withdraw from the labor force in response to rising incomes (Mammen and Paxson, 2000). While either effect may be present in our data, we should find that our results are strongest in areas where female labor force participation remained low in the early twentieth century.

5.2. **Evidence.** In this section, we present evidence for our interpretation of our findings.

TABLE 3. Nuptiality rises for women aged 20-24

	(1) Age at Marriage	(2) Age at Marriage	(3) Pct. Married 20-24	(4) Pct. Married 20-24
ln Market Access: (P=1830, θ=8.22)	1.050	-1.069*	0.286***	0.121**
	(0.770)	(0.614)	(0.059)	(0.050)
N	1,188	1,188	1,532	1,532
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

5.2.1. *Proximate mechanisms.* We begin by documenting evidence of the proximate mechanisms underpinning our results. In Table 3, we show that higher marriage rates are indeed a channel linking market access to greater fertility. We consider two outcomes – the average age at marriage in the population, and the nuptiality of women aged 20-24. Both variables are directly available in the Princeton European Fertility Project data, though they are available only for a subset of the observations. While we find no robust impact on the mean age at marriage in the population, we do find evidence that a greater marriage rate among women in their early 20s is a probable proximate mechanism. We document other proximate mechanisms in the appendix. In Table D.1, we show that both marital and non-marital fertility rise in response to market access, though the predicted increase in marital fertility is much larger. We measure these outcomes using the indices of marital fertility (I_g) and non-marital fertility (I_h) present in the Princeton European Fertility Project data. We also consider several other variables in the Princeton European Fertility Project that measure other possible proximate mechanisms related to urban fertility (Tables D.2 and D.3), rural fertility (Tables D.4 and D.5), and other intermediate outcomes (Tables D.6 and D.7). Generally, however, these variables are missing too often from the data to provide meaningful guidance.

5.2.2. *Higher income.* The primary channel through which we expect market access to lead to rising fertility is greater income. To this end, we use data on population and on Gross Domestic Product (GDP) per capita from Rosés and Wolf (2020). To merge variables from their data, at the NUTS 2 level, to the polygons in MPIDR (2013), we take area-weighted averages. That is, if multiple NUTS 2 polygons intersect one polygon in MPIDR (2013), we average over the values for the NUTS 2 polygons using as weights the area shares of the MPIDR (2013) polygon that they represent. Further, Rosés and Wolf (2020) do not report population and GDP uniformly at the start of each decade. Rather, they report values for 1900, 1910, 1925, 1938, and 1950. We use their 1925 values for 1920, their 1938 values for 1930, and their 1950 values for 1940. While, ideally, we would consider earlier years, we are constrained by data availability.

TABLE 4. Rising incomes, rising populations

	(1) ln RW GDP per capita	(2) ln RW GDP per capita	(3) ln RW Population	(4) ln RW Population
ln Market Access: (P=1830, θ=8.22)	0.018***	0.008**	0.007***	0.005**
	(0.006)	(0.004)	(0.003)	(0.002)
N	1,743	1,743	1,743	1,743
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

In Table 4, we show estimates of equation (1) with these measures of population and (the natural logarithm of) GDP per capita as outcomes. We label these "ln RW Population" and "ln RW GDP per capita," respectively. We find that market access predicts greater populations; this may reflect in-migration due to higher incomes, lower mortality, or may be downstream from our fertility outcome. Critically, we find evidence that greater market access predicts greater per capita income, though data are only available for the latter half of our sample.

Similarly, we show in Table 5 that our results are driven by locations that had achieved relatively high levels of income by 1900. In particular, we split the sample at the median by GDP per capita in 1900 according to the Rosés and Wolf (2020) data. The results in the table suggest that the positive link between market access and fertility is driven by the above-median sample. While the data do not allow us to split the sample by GDP per capita in earlier years, we provide suggestive evidence to this effect in Table D.8.

TABLE 5. Results by GDP per capita in 1900

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.245***	0.205***	0.031	-0.006
	(0.039)	(0.057)	(0.047)	(0.041)
N	1,768	1,768	1,738	1,738
Sample	ln RW GDP per capita Above Median in 1900	ln RW GDP per capita Above Median in 1900	ln RW GDP per capita Below Median in 1900	ln RW GDP per capita Below Median in 1900
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

In this table, we estimate GDP per capita in 1880, split the sample at the median, and reconfirm the results of Table 5.[14]

We report the results of several corroborating sample splits in the appendix. In Table D.9, we split the sample by median population in 1900, showing that the results are driven by locations that had become more populated by 1900. To estimate populations for each location, we sum over raster points from Klein Goldewijk, Beusen and Janssen (2010) within a polygon in MPIDR (2013). We split the sample at the median by the percentage of the labor force in agriculture in 1900 in Table D.10, industry in Table D.11, and services in Table D.12. We take these three variables from Rosés and Wolf (2020)

[14]To estimate GDP per capita in 1880, we proceed in two steps. First, we obtain data on the ratio of regional to national GDP at the subnational level in years prior to 1880 for a small number of countries for which it is available: Denmark, Finland, Italy, the Netherlands, and Sweden. The Danish data are from Janisse et al. (2018), the Dutch data are from correspondence with Nikolaus Wolf, the Finnish data are from Enflo et al. (2014), the Italian data are from Felice (2009) and the Swedish data are from Nilsson et al. (2023). We then use country-level GDP estimates for the corresponding years from Bolt and van Zanden (2020) to convert these into GDP per capita estimates. If GDP per capita is thus available in 1880 or 1881, we use this estimate. Otherwise, we interpolate geometrically between available years to produce an estimate for 1880. For all other countries in the sample, we use the Rosés and Wolf (2020) estimate for 1900 to generate the ratio of regional to national GDP in 1900. We then use country-level GDP estimates for 1880 from Bolt and van Zanden (2020) to convert these ratios into GDP per capita estimates for 1880, assuming the same ratios apply in this earlier year.

and, as above, map them into the locations in our data using area weights. Our results are driven by locations that had, by 1900, achieved greater structural change out of agriculture and into industry and services. Across sample splits, these patterns reconfirm our finding that our results are driven by locations that had achieved higher levels of development. The industry results in D.11 suggest further that the returns to child labor in industry may have played a mediating role as well, much like the returns to child labor that Guldi and Rahman (2022) argue mattered in the United States.[15]

TABLE 6. Results by schooling in 1900

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.014	0.024	0.259***	0.099**
	(0.037)	(0.036)	(0.055)	(0.046)
N	2,027	2,027	2,017	2,017
Sample	Above Median Years of Education in 1900	Above Median Years of Education in 1900	Below Median Years of Education in 1900	Below Median Years of Education in 1900
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

5.2.3. *Human capital and female labor force participation.* Our conceptual framework suggests that our results should be driven by locations that lagged in human capital, despite market access. In Table 6, we provide evidence that areas with lower human capital drive our results. We use years of schooling data at the country level from the Clio Infra database to split the sample at the median. We find that our results are larger for countries that lagged in schooling in 1900. In the appendix, we provide three pieces of additional supporting evidence. First, in Table D.13, we split the sample by median years of schooling in 1870. Second, in Table D.14, we make a median split by numeracy

[15]We are unable to corroborate our results using wage data, as the data on wages are too disparate. Merging data on wages from Allen (2001) to our data would permit us to run regressions with fewer than 50 observations – less than 5% of our original sample.

in 1880. Third, in Table D.15, we make the equivalent split by numeracy in 1900. For all three variables, we again use country-level data from the Clio Infra database. All three exercises confirm that it is the sample that achieved lower levels of human capital that drive our results.

TABLE 7. Results by female labor force participation (FLFP) c. 1900

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.033	0.032	0.289***	0.273***
	(0.050)	(0.000)	(0.043)	(0.047)
N	1,623	1,623	1,843	1,843
Sample	Above Median FLFP	Above Median FLFP	Below Median FLFP	Below Median FLFP
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

Our conceptual framework also suggests that, for an income channel to dominate, our results should be present for locations where new opportunities for women were not sufficiently widespread that a countervailing opportunity cost channel could prevail. We provide evidence of this prediction in Table 7. We split the sample by the median rate of female labor force participation circa 1900, showing that it is the low-participation sample that drives the results. To obtain data on female labor force participation, we use the earliest year for which a country-level estimate is available in Table A.1. of Olivetti (2013).

5.3. **Alternative mechanisms.** We consider several alternative mechanisms that could account for a link between railways and fertility. We provide evidence in the appendix that these cannot explain our results, and briefly discuss this evidence here. The alternatives we consider are: diffusion of fertility norms; mortality; the direct effects of having a railway itself; sectoral change and urbanization; human capital, and; political connections.

5.3.1. *Diffusion of fertility norms.* The diffusion of fertility norms may have played a role in spreading lower fertility from France to the rest of Europe (Daudin, Franck and Rapoport, 2019). But it does not explain the connection we find between market access and fertility in our data. We take two approaches here. First, we construct a measure of weighted fertility access that measures the initial fertility of the other locations to which a given location gains access over time. In particular, we compute:

$$WeightedFertilityAccess_{ld} = \sum_{l'} w_{l'd} F_{l'0}, \tag{5}$$

where:

$$w_{l'd} = \frac{\tau_{ll'd}^{-\theta}}{\sum_{l'} \tau_{ll'd}^{-\theta}}. \tag{6}$$

In equation (5), $F_{l'0}$ is the initial fertility of location l', measured in the first year it appears in the Princeton European Fertility Project data, while $w_{l'd}$ are travel cost weights that sum to 1 for a given observation ld. Other variables are defined as in equation (3), which defined market access. The intuition for this measure is that it weights more highly locations l' to which location l is more connected in terms of having lower travel times. In Table E.1, we re-estimate equation (1) replacing our measure of market access with the natural logarithm of $WeightedFertilityAccess_{ld}$ as the principal independent variable. This measure is only significant at the 10% level in one specification, and the standardized coefficient is an order of magnitude smaller than the coefficient we estimate for market access.[16]

As an alternative approach to considering the diffusion of fertility norms, we consider gravity-type regressions. We begin by reshaping our data into a panel of dyads such that each observation is a pair of locations l and l' observed in decade d. Next, we estimate the following specification:

$$FertilityDistance_{ll'd} = \alpha + \beta TimeDistance_{ll'd} + \delta_{ll'} + \eta_d + \epsilon_{ll'd}. \tag{7}$$

In equation (7), $FertilityDistance_{ll'd}$ is one of two measures of the difference between fertility in the two locations in a given decade. The first measure, "AD. Total Fertility" is the absolute difference in Total Fertility I_f between the two locations in period d. The second measure, "ln Fertility Difference," is the natural logarithm of this absolute difference. α is a constant. $TimeDistance_{ll'd}$ is one of two measures of the time costs of reaching l' from l in decade d. The first measure is the $\tau_{ll'd}$ that we estimate as part of our

[16]We compute the standardized coefficient as $\frac{0.287 \times 0.309}{8.93} \approx 0.001$.

market access calculation in Section 3.2. The second is the natural logarithm of $\tau_{ll'd}$. β is the coefficient of interest, with positive values indicating fertility convergence across location pairs for which time costs decline. $\delta_{ll'}$ is fixed effects for the location pair. η_d is decade fixed effects. We use two-way clustering (Cameron, Gelbach and Miller, 2011), and cluster by both l and l'. To avoid double counting, we keep only the lower triangular part of the matrix of ll' pairs, and exclude observations where $l = l'$. The results, in Table E.2, show no evidence that locations that saw their travel times decrease saw their fertility differences converge.

5.3.2. *Mortality.* The next alternative channel we consider is mortality. If railways connected locations to sources of disease, or if market access raised morality (for example by encouraging denser population), fertility might have increased to compensate (Angeles, 2010). We approach this possibility in two ways. First, we construct a time-varying measure of Weighted Mortality Access. This measure is analogous to the Weighted Fertility Access measure that we construct in equations (5) and (6), except that we replace initial fertility $F_{l'0}$ with initial mortality, i.e. with infant mortality for location l' the first year it appears in the Princeton European Fertility Project data. In Table E.3, we re-estimate equation (1) replacing our measure of market access with the natural logarithm of Weighted Mortality Access as the principal independent variable. The estimated relationship between fertility and access to higher infant mortality via the railway is not robust to controls, and has an absolute standardized magnitude less than 0.1, again smaller than what we estimate for market access.

We corroborate this evidence by showing it is indeed locations with lower mortality that drive our results. We obtain data from de la Escosura (2022) on life expectancy at the country level in 1870. In Table E.4, we split the sample at the median, showing results are driven by countries with greater life expectancy in 1870.

5.3.3. *Direct effects of a railway.* Next, we show that our results are driven by the market access brought by railways, and not by the direct presence of railways. First, 89% of location-decade observations in our data are intersected by at least one active rail line – this is not the source of variation in our independent variable. Second, we show in Table E.5 that controlling for railway density (the length of track per unit area) does little to diminish the coefficient on market access. Indeed, the coefficient on railway density has the opposite sign.

5.3.4. *Sectoral change and urbanization.* While the evidence on sectoral shares presented above in Tables D.10, D.11, and D.12 is consistent with the interpretation that our results are driven by locations that had achieved higher levels of development by 1900, we do not find that sectoral change *per se* accounts for the relationship between market access and fertility. In Table E.6 we treat the shares of the labor force in agriculture, industry, and services as outcome variables. Here, we again take data on sectoral shares from Rosés and Wolf (2020). We find little evidence of a direct impact of market access on sectoral shares, though we caveat this finding by noting these data, like the other Rosés and Wolf (2020) data, are only available in 1900 and afterwards. Similarly, we show in Table E.7 that controlling for the urbanization rate does little to change our estimates of β. We measure urbanization here as the share of a polygon's population that is urban according to the raster data in Klein Goldewijk, Beusen and Janssen (2010).

5.3.5. *Human capital.* While the evidence on literacy and numeracy presented above in Tables D.13, 6, D.14, and D.15 suggests that our results are driven by areas that lagged in human capital, we do not find that human capital *per se* accounts for the relationship between market access and fertility. We show in Tables E.8 and E.9 that controlling for years of schooling or numeracy at the country level as recorded in the Clio Infra data does little to change our estimates of β.

5.3.6. *Political connections.* Finally, we use Table E.10 to show that political connections do not account for our results, since our findings are largely unchanged discarding locations containing country capitals from the data.

6. ROBUSTNESS

In this section, we outline robustness checks that are mostly presented in the appendix. We detail the plausibility of the parallel trends assumption in our data, the robustness of our results to alternative calculations of market access, alternative approaches to standard errors, robustness of the instrumental variables estimates, alternative functional forms and specifications, accounting for coal, sample construction and possible outliers, and timing.

6.0.1. *Parallel trends.* A structural interpretation of equation (1) requires that ϵ_{ld} is uncorrelated with $\ln(MarketAccess)_{ld}$, conditional on controls x_l and fixed effects δ_l and η_d for β to be estimated correctly. Alternatively, interpreted in a potential outcomes framework, our two-way fixed effects estimation hinges on the strong parallel trends assumption (Callaway, Goodman-Bacon and Sant'Anna, 2021). We both validate and relax this assumption. First, to validate the assumption, we add forwards (F) and backwards (L) lags of market access in Table 8. While previous lags of market access may have predictive power, we find no evidence that future market access predicts current fertility.

To relax the assumption, we consider instead estimation in long differences. We estimate the following equation using OLS:

TABLE 8. Forwards and backwards lags

	(1) Total Fertility	(2) Total Fertility
F. ln Market Access: (P=1830, θ=8.22)	0.025	-0.007
	(0.041)	(0.039)
ln Market Access: (P=1830, θ=8.22)	0.102***	0.044*
	(0.026)	(0.025)
L. ln Market Access: (P=1830, θ=8.22)	0.143***	0.106***
	(0.027)	(0.027)
N	2,731	2,731
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE 9. Long differences: 1870 to 1910

	(1) Δ Total Fertility	(2) Δ Total Fertility
Δ ln Market Access	0.340***	0.235***
	(0.062)	(0.069)
N	347	347
Controls	No	Yes

***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Controls are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Robust standard errors in parentheses.

$$\Delta TotalFertility_l = \alpha + \beta \Delta \ln MarketAccess_l + x_l' \gamma + \epsilon_l \tag{8}$$

In equation 8, the dependent variable $\Delta TotalFertility_l$ is the change in fertility in location l between 1870 and 1910. α is a constant. The main independent variable $\Delta \ln MarketAccess_l$ is the change in the natural logarithm of market access over the same time period. We include the same controls in x_l that we interact with year fixed effects in our baseline specifications. We report robust, Huber-White standard errors. The results, in Table 9, confirm our main finding that increases in market access predict increases in – or in this context, smaller reductions in – fertility. We report equivalent IV results in Table F.1, using the change in the natural log of distant market access as an instrument. This exercise further confirms our results.

While our focus on the years 1870 to 1910 is due to the relative completeness of the data and the period's importance in Europe's fertility transition, we show in Figure F.1 that we obtain positive estimates of β when estimating equation 8 using most combinations of starting and ending years in the data. There are two broad exceptions: periods with a start year of 1840 and an end year of 1880 or earlier, prior to the widespread fertility declines in Europe, and periods with a start year of 1890 or later with an end point after the First World War. The First World War itself led to major disruptions in European fertility (Vandenbroucke, 2014).

6.0.2. *Alternative market access.* We next consider several alternative computations of our market access variable. We begin by changing the trade elasticity θ. In Table F.2, we show results with $\theta = 1$ and $\theta = 3.60$. In Table F.3, we show results with $\theta = 12.86$. These alternative values come from Donaldson and Hornbeck (2016). Our estimates of β remain positive and significant throughout.[17]

Next, we consider several alternative measures of market access based on differing sources of data on $P_{l'd}$. While we are not aware of sub-national census data with continental coverage dating back to 1830, census-based populations are available for several countries dating back to 1850 from Martí-Henneberg (2023). We use these data construct alternative estimates of $P_{l'd}$ in 1850 and re-construct our market access measures accordingly. For countries not covered by these data, we impure 1850 populations using the Klein Goldewijk, Beusen and Janssen (2010) data.[18] Results, indicated in Table F.3 as "$P = 1850$," are similar to our baseline estimates.

[17]While the magnitudes of β vary with θ, so does the standard deviation of the independent variable. In Table F.4, we report standardized estimates in which we normalize both the fertility and market access variables to have $N(0, 1)$ distributions. The differences in magnitude are now less dramatic, though they remain larger for lower values of θ.

[18]To construct an estimate of population in 1850, we map the populations from 1850 into the polygons in MPIDR (2013) using area weights. The 1850 population data are not available for Germany, Poland, Hungary, Romania, Serbia, Iceland, Greece, or Albania Macedonia and Tracia. For locations in these countries, we impute populations in 1850 using data from Klein Goldewijk, Beusen and Janssen (2010). To make this imputation, we start with locations in MPIDR (2013) whose areas overlap at least 90% with locations in the population data. We regress the natural logarithm of population in 1850 on the natural logarithm of population in Klein Goldewijk, Beusen and Janssen (2010). We use the resulting coefficient estimates to predict populations in 1850 out of sample for the remaining locations in MPIDR (2013).

While we use populations in 1830 in our baseline calculations to avoid potential endogeneity, we show in Table F.6 that estimates using contemporary populations – indicated in the table as "$P = t$" – are similar to our baseline estimates. We also show in this table that we obtain similar results if we re-scale values of P within each country so that they sum to the country-level totals Federico and Tena Junguito (2023) report for 1830. These results are indicated in the table as "$P = FT$."

We construct three further population estimates that make no use of the Klein Goldewijk, Beusen and Janssen (2010) data at all. The first computes $P_{l'd}$ by summing over the 1830 populations of the cities reported in Reba, Reitsma and Seto (2016).[19] Estimates using this measure are reported in Table F.7 as "$P =$ Cities in 1830" and remain similar to our baseline estimates. The second population estimate computes $P_{l'd}$ by summing over the populations in 1800 of the cities reported in Bosker, Buringh and Van Zanden (2013).[20] Estimates using this measure are reported in Table F.8 as "$P =$ Cities in 1800" and again remain similar to our baseline estimates. The third population estimate computes $P_{l'd}$ by summing over the populations in 1850 of the cities reported in Martí-Henneberg (2023).[21] Estimates using this measure are reported in Table F.9 as "$P =$ Cities in 1850" and again remain similar to our baseline estimates.

We also consider alternative constructions of the transportation network when constructing market access. First, we show robustness to the assumption of a travel speed of 60 kilometers per hour by rail. Estimates using a speed of 30 kilometers per hour are presented in Table F.10, while estimates considering a speed of 120 kilometers per hour are presented in Table F.11. Results remain similar to our baseline estimates, and reflect the fact that it is not the specific speed of rail that matters, but rather the fact that rail was an order of magnitude faster than other available modes of transportation. In Table F.12 we consider the possible complication of borders, allowing travel speeds to slow to 1 kilometer per hour when crossing a country border. Our results are unchanged.

Our baseline estimates of market access consider only travel by rail, walking, or traditional seafaring. Our estimates should be interpreted, then, not as the impacts of market access *per se*, but of the changes in market access wrought by the expansion in the railroad network. Here, we consider the addition of two alternative modes of transportation. The first is roads. In our context, this means travel by horse and not by automobile. We begin with a measure of the road network in 1820 from Martí-Henneberg (2023). As

[19]These data are based largely on Chandler (1987) and Modelski (2003). Because these data form a very unbalanced panel of city populations, we follow Dincecco, Fenske and Menon (2023) and impute city populations geometrically between reported years. For years prior to the first time a city population appears, we assume a city grows geometrically from an initial size of 1 in the first year in the data – 2250BCE. We further assume that a city stops growing after the final reported city size.

[20]This source draws heavily on Bairoch (1988) and requires no imputation.

[21]This source is based on census data and requires no imputation.

speeds along Britain's turnpikes had risen to 8 miles per hour by 1829 (Bogart, 2005), we allow for a speed of 13 kilometers per hour along these roads. We then compute an alternative measure of market access. In Table F.13 we show results using this measure, and the results are similar to our baseline, albeit smaller. The second alternative is travel by steamship. To construct this alternative, we allow travel in open water to have a speed of 16 kilometers per hour and compute an alternative measure of market access.[22] Results using this measure are in Table F.14, and are similar to our baseline estimates.

6.0.3. *Alternative standard errors.* We consider possible spatial correlation in the error term in Table F.15, using the Colella et al. (2019) implementation of Conley (1999) standard errors with distance cutoffs ranging from 250 to 1500 kilometers. Our results are robust at the 10% level in all specifications, at the 5% level in all but one specification, and become more precisely estimated as the cutoff distance expands.

6.0.4. *Robustness of instrumental variables.* We also consider the robustness of our instrumental variables approach. The decision to use 500 kilometers as a cutoff in defining the instrument is ultimately arbitrary. In Figure F.2, however, we show that results are largely unchanged using other distance cutoffs. While the fact that our instrument leverages variation based on distant markets gives us *a priori* grounds to believe that exclusion restriction is satisfied, we also implement the D'Haultfœuille, Hoderlein and Sasaki (2021) control function approach to testing this assumption directly. Without controls, the test gives a p value of 0.721 while, with controls, the test gives a p value of 0.414. Both cases suggest that the exclusion restriction is satisfied in our data.

6.0.5. *Alternative functional forms and specifications.* We consider several issues of specification and functional form. Our baseline specification assumes a linear-logarithmic relationship between fertility and market access, a relationship validated by the binned scatterplot in Figure C.1. In Table F.16, we show the results of using the natural logarithm of our fertility measure instead. We confirm the sign and significance of our main result, though the magnitude – including the standardized magnitude – is now smaller. Further, we change our regression specification in Tables F.17 and F.18, adding country-specific linear time trends and country-by-year fixed effects respectively. While the latter exercise reduces the magnitude of our results, consistent with the usual exacerbation of attenuation bias due to fixed effects, both exercises confirm the sign and significance of our results.

6.0.6. *Accounting for coal.* We show that accounting for coal does not affect our results. In Table F.19, we include two time-invariant controls that we interact with our decade fixed effects. Both controls come from Fernihough and O'Rourke (2021). The first, "Coal

[22]Kelly and Ó Gráda (2014) give a range of 8 to 13 knots for steam in the mid-nineteenth century.

Share," is the share of a location's land area in which there are coal deposits. The second, "Carbon Share," is the share of a location's land area in which there is carboniferous rock. Accounting for these controls does little to our magnitudes or standard errors.

6.0.7. *Sample construction and possible outliers.* We next consider several issues related to sample construction. We begin by estimating our results separately by country. We plot the resulting coefficients in Figure F.3 and map them in Figure F.4. These patterns suggest, consistent with the results in Section 5.2, that our main results are driven by countries that had achieved higher levels of development by the start of the First World War. Less-developed countries, notably Ireland, Sweden and countries in Southeastern Europe, work against our main result. Belgium, however, stands out as an outlier with an unusually large and positive coefficient estimate. We show in Table F.20, however, that dropping Belgium from the sample does little to our results.

Another possible set of outliers is locations with large areas; Brown and Guinnane (2007) emphasize that fertility may be particularly poorly measured in these places. We show in Table F.21, however, that discarding the top 40% of locations by land area from our data does little to our results. Similarly, we would be concerned if locations with poor data that appear only infrequently in our data were to drive our results. We show in Figure F.5 the sensitivity of our results to keeping locations that only appear at least once, at least twice, and so forth. The results survive retaining locations for which fertility is reported in at least 9 time periods. The maximum, 10 time periods, represents less than 20% of our sample.

6.0.8. *Timing.* Finally, we consider timing. This consideration is for two reasons. First, we have framed our discussion in terms of Europe's fertility transition even though our sample covers a broad time period. We split our sample by time period in Table F.22 and show that it is indeed the years from 1870 to 1914 – conventionally considered the high point of the European fertility transition – that drive our result. This result is evidence that our discussion is correctly framed. Second, the spread of the automobile, particularly after 1914, could make our market access calculations, which ignore this new technology, invalid. In Table F.23, however, we show that discarding years after 1914 does little to our results.

7. Conclusion

We have shown that the expansion of railways in Europe, rather than speeding up the fertility transition, delayed it. We use fixed effects and instrumental variables approaches, and find that, had market access in Europe remained at its 1830 level, fertility would have been 8% lower on the eve of the First World War. We document greater nuptiality of young women as a proximate mechanism, while the most plausible deeper economic mechanism is income gains with children as a normal good. Consistent with this interpretation, we show that market access indeed predicts higher income, that our results are driven by locations that ultimately attained higher incomes, and that they are driven by locations were countervailing mechanisms such as human capital and female labor force participation remained relatively low.

Our findings have several implications for the existing literature. First, we demonstrate the importance of economic factors in understanding European fertility during the period of the continent's fertility transition. Second, we show that greater living standards are not enough, at least in our context, to lead to declining fertility – they must be accompanied by additional forces such as returns to skill and labor market opportunities for women. Third, we have reinforced recent findings in economic geography that the impacts of transportation infrastructure cannot be understood without general-equilibrium concepts such as market access.

Our study does, of course, have limitations. The external validity of a continent-wide focus has come at the expense of confirming these results using the often rich but context-laden micro data that exists for some European countries. Similarly, our aggregate data prevent us from exploring heterogeneity in the fertility response within regions, for example by occupational class or by religion. The data on sub-national incomes in general and wages in particular are, at present, too sparse for us to push our analysis of income as a mechanism into the earlier periods in our data. The same is true for countervailing mechanisms such as human capital and female labor force participation. We leave these tasks to future research.

References

Aaronson, Daniel, Fabian Lange and Bhashkar Mazumder. 2014. "Fertility transitions along the extensive and intensive margins." *American Economic Review* 104(11):3701–3724.

Ager, Philipp, Casper Worm Hansen and Peter Sandholt Jensen. 2018. "Fertility and early-life mortality: Evidence from smallpox vaccination in Sweden." *Journal of the European Economic Association* 16(2):487–521.

Alesina, Alberto, Paola Giuliano and Bryony Reich. 2021. "Nation-Building and Education." *The Economic Journal* 131(638):2273–2303.

Allen, Robert C. 2001. "The great divergence in European wages and prices from the Middle Ages to the First World War." *Explorations in Economic History* 38(4):411–447.

Amatulli, Giuseppe, Sami Domisch, Mao-Ning Tuanmu, Benoit Parmentier, Ajay Ranipeta, Jeremy Malczyk and Walter Jetz. 2018. "A suite of global, cross-scale topographic variables for environmental and biodiversity modeling." *Scientific data*

5(1):1–15.

Andersson, David, Thor Berger and Erik Prawitz. 2023. "Making a market: Infrastructure, integration, and the rise of innovation." *The Review of Economics and Statistics* 105(2):258–274.

Angeles, Luis. 2010. "Demographic transitions: analyzing the effects of mortality on fertility." *Journal of Population Economics* 23:99–120.

Ashraf, Quamrul and Oded Galor. 2011. "Dynamics and stagnation in the Malthusian epoch." *American Economic Review* 101(5):2003–2041.

Atack, Jeremy, Fred Bateman and Robert A. Margo. 2010. "Did railroads induce or follow economic growth? Urbanization and population growth in the American midwest, 1850-60." *Social Science History* 34:171–197.

Atack, Jeremy and Robert A. Margo. 2011. "The impact of access to rail transportation on agricultural improvement: The American Midwest as a test case, 1850–1860." *Journal of Transport and Land Use* 4:5–18.

Atack, Jeremy, Robert A. Margo and Paul W. Rhode. 2022. "Industrialization and urbanization in nineteenth century America." *Regional Science and Urban Economics* 94:103678.

Bairoch, Paul. 1988. *Cities and economic development: from the dawn of history to the present.* University of Chicago Press.

Basso, Alberto and David Cuberes. 2017. "Deep-rooted determinants of the fertility transition across countries." *Macroeconomic Dynamics* 21(7):1712–1751.

Beach, Brian and Walker Hanlon. 2023. "Culture and the historical fertility transition." *The Review of Economic Studies* 90(4):1669–1700.

Becker, Gary S and H Gregg Lewis. 1973. "On the interaction between the quantity and quality of children." *Journal of Political Economy* 81(2, Part 2):S279–S288.

Becker, Gary Stanley. 1960. An economic analysis of fertility. In *Demographic and Economic Change in Developed Countries*, ed. Gary Stanley Becker. Princeton University Press pp. 61–89.

Becker, Sascha O S, Francesco Cinnirella and Ludger Woessmann. 2013. "The trade-off between fertility and education: evidence from before the demographic transition." *European Review of Economic History* 17(1):24–44.

Becker, Sascha O S and H Gregg Lewis. 2010. "The trade-off between fertility and education: evidence from before the demographic transition." *Journal of Economic Growth* 15(4):177–204.

Black, Dan A, Natalia Kolesnikova, Seth G Sanders and Lowell J Taylor. 2013. "Are children "normal"?" *The Review of Economics and Statistics* 95(1):21–33.

Blanc, Guillaume. 2021. "The cultural origins of the demographic transition in France." *Working paper, University of Manchester.* .

Blanc, Guillaume and Romain Wacziarg. 2020. "Change and persistence in the age of modernization: Saint-Germain-d'Anxure, 1730–1895." *Explorations in Economic History* 78:101352.

Bleakley, Hoyt and Fabian Lange. 2009. "Chronic disease burden and the interaction of education, fertility, and growth." *The Review of Economics and Statistics* 91(1):52–65.

Bogart, Dan. 2005. "Turnpike trusts and the transportation revolution in 18th century England." *Explorations in Economic History* 42(4):479–508.

Bogart, Dan, You Xuesheng, Alvarez-Palau Eduard J., Satchell Max and Shaw-Taylor Leigh. 2022. "Railways, divergence, and structural change in 19th century England and Wales." *Journal of Urban Economics* 128:103390.

Bolt, Jutta and Jan Luiten van Zanden. 2020. "Maddison Project Database, version 2020. Maddison style estimates of the evolution of the world economy. A new 2020 update.".

Boserup, Ester. 1970. *Woman's Role in Economic Development.* Allen & Unwin.

Bosker, Maarten, Eltjo Buringh and Jan Luiten Van Zanden. 2013. "From Baghdad to London: Unraveling urban development in Europe, the Middle East, and North Africa, 800–1800." *Review of Economics and Statistics* 95(4):1418–1437.

Brown, John C and Timothy W Guinnane. 2002. "Fertility transition in a rural, Catholic population: Bavaria, 1880–1910." *Population studies* 56(1):35–49.

Brown, John C and Timothy W Guinnane. 2007. "Regions and time in the European fertility transition: problems in the Princeton Project's statistical methodology." *The Economic History Review* 60(3):574–595.

Callaway, Brantly, Andrew Goodman-Bacon and Pedro HC Sant'Anna. 2021. "Difference-in-differences with a continuous treatment." *arXiv preprint arXiv:2107.02637* .

Cameron, A Colin, Jonah B Gelbach and Douglas L Miller. 2011. "Robust inference with multiway clustering." *Journal of Business & Economic Statistics* 29(2):238–249.

Chan, Jeff. 2023. "The long-run effects of childhood exposure to market access shocks: Evidence from the US railroad network expansion." *Explorations in Economic History* p. 101503.

Chandler, Tertius. 1987. *Four Thousand Years of Urban Growth.* St David's.

Chaudhary, Latika and James Fenske. 2023. "Railways, development, and literacy in India." *Forthcoming in the Journal of Economic History* .

Ciccarelli, Carlo and Alessandro Nuvolari. 2015. "Technical change, non-tariff barriers, and the development of the Italian locomotive industry, 1850-1913." *Journal of Economic History* pp. 860–888.

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli and Mathias Thoenig. 2019. "Inference with arbitrary clustering.".

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92(1):1–45.

Crafts, Nicholas FR. 1989. "Duration of marriage, fertility and women's employment opportunities in England and Wales in 1911." *Population Studies* 43(2):325–335.

Crafts, Nicholas and Terence C Mills. 2020. "The race between population and technology: real wages in the first industrial revolution." *Centre for Economic Policy Research Discussion Paper No. DP15174* .

Daudin, Guillaume, Raphaël Franck and Hillel Rapoport. 2019. "Can internal migration foster the convergence in regional fertility rates? Evidence from 19th century France." *The Economic Journal* 129(620):1618–1692.

De La Croix, David and Faustine Perrin. 2018. "How far can economic incentives explain the French fertility and education transition?" *European Economic Review* 108:221–245.

de la Escosura, Leandro Prados. 2022. *Human Development and the Path to Freedom: 1870 to the Present.* Cambridge University Press.

Diebolt, Claude, Audrey-Rose Menard and Faustine Perrin. 2017. "Behind the fertility-education nexus: what triggered the French development process?" *European Review of Economic History* 21:357–92.

Dincecco, Mark, James Fenske and Anil Menon. 2023. "The Columbian exchange and conflict in Asia." *Working Paper. Available at SSRN 3750813* .

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the impact of transportation infrastructure." *American Economic Review* 108(4-5):899–934.

Donaldson, Dave and Richard Hornbeck. 2016. "Railroads and American economic growth: A "market acces" approach." *The Quarterly Journal of Economics* 131(2):799–858.

D'Haultfœuille, Xavier, Stefan Hoderlein and Yuya Sasaki. 2021. "Testing and relaxing the exclusion restriction in the control function approach." *Forthcoming in the Journal of Econometrics* .

Enflo, Kerstin et al. 2014. "Finland's regional GDPs 1880-2010: estimates, sources and interpretations." *Lund Papers in Economic History. General Issues; No. 135* .

Federico, Giovanni and Antonio Tena Junguito. 2023. "How many people on earth? World population 1800-1938." *Universidad Carlos III de Madrid. Instituto Figuerola de Historia y Ciencias Sociales. Working Papers in Economic History. 23-02. ISSN: 2341-2542.* .

Felice, Emanuele. 2009. "Estimating regional GDP in Italy (1871-2001): sources, methodology and results." *IFCS - Working Papers in Economic History.WH wp09-07, Universidad Carlos III de Madrid. Instituto Figuerola.* .

Fernihough, Alan. 2017. "Human capital and the quantity–quality trade-off during the demographic transition." *Journal of Economic Growth* 22:35–65.

Fernihough, Alan and Kevin Hjortshøj O'Rourke. 2021. "Coal and the European industrial revolution." *The Economic Journal* 131(635):1135–1149.

Fogel, Robert William. 1964. *Railroads and American economic growth.* Johns Hopkins Press Baltimore.

Forero, Andrés, Francisco A. Gallego and Matías Tapia. 2021. "Railroads, specialization, and population growth: evidence from the first globalization." *Journal of Population Economics* 34:1027–1072.

Friedlander, Dov, Jona Schellekens and Eliahu Ben-Moshe. 1991. "The transition from high to low marital fertility: cultural or socioeconomic determinants?" *Economic Development and Cultural Change* 39(2):331–351.

Galor, Oded. 2011. *Unified Growth Theory.* Princeton University Press.

Galor, Oded. 2022. *The Journey of Humanity: And the Keys to Human Progress.* Random House.

Galor, Oded and Andrew Mountford. 2008. "Trading population for productivity: theory and evidence." *The Review of Economic Studies* 75(4):1143–1179.

Galor, Oded and David N Weil. 1999. "From Malthusian stagnation to modern growth." *American Economic Review* 89(2):150–154.

Galor, Oded and David Weil. 1996. "The Gender Gap, Fertility, and Growth." *American Economic Review* 86(3):374–87.

Galor, Oded and Ömer Özak. 2015. "Land Productivity and Economic Development: Caloric Suitability vs. Agricultural Suitability." *Working Paper* .

Galor, Oded and Ömer Özak. 2016. "The Agricultural Origins of Time Preference." *American Economic Review* 106(10):3064–3103.

Glaeser, Edward L. 2011. *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier.* MacMillan.

Goldin, Claudia. 1995. The U-Shaped Female Labor Force Function in Economic Development and Economic History. In *Investment in Women's Human Capital*, ed. Paul T. Schultz. University of Chicago Press pp. 61–89.

Guinnane, Timothy W. 2011. "The historical fertility transition: A guide for economists." *Journal of Economic Literature* 49(3):589–614.

Guinnane, Timothy W, Barbara S Okun and James Trussell. 1994. "What do we know about the timing of fertility transitions in Europe?" *Demography* 31(1):1–20.

Guldi, Melanie and Ahmed Rahman. 2022. "Little Divergence in America – Market Access and Demographic Transition in the United States." *IZA - Institute of Labor Economics Working Paper No. 15215* .

Hansen, Casper Worm, Peter Sandholt Jensen and Lars Lønstrup. 2018. "The fertility decline in the United States: schooling and income." *Macroeconomic Dynamics* 22(6):1584–1612.

Hornbeck, Richard and Martin Rotemberg. 2021. "Growth Off the Rails: Aggregate Productivity Growth in Distorted Economies." *Working Paper, University of Chicago Booth School of Business* .

Hornung, Erik. 2015. "Railroads and growth in Prussia." *Journal of the European Economic Association* 13(4):699–736.

Humpries, Jane and Carmen Sarasúa. 2012. "Off the Record: Reconstructing Women's Labor Force Participation in the European Past." *Feminist Economics* 18(4):39–67.

Janisse, Kari Anne, Peter Sandholt Jensen, Cristina Victoria Radu and Paul Richard Sharp. 2018. Regional GDP in Denmark, 1850–2010. In *The Economic Development of Europe's Regions.* Routledge pp. 85–102.

Jedwab, Remi and Adam Storeygard. 2022. "The average and heterogeneous effects of transportation investments: Evidence from Sub-Saharan Africa 1960–2010." *Journal of the European Economic Association* 20(1):1–38.

Jensen, Robert. 2012. "Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India." *The Quarterly Journal of Economics* 127(2):753–792.

Katz, Ori. 2018. "Railroads, Economic Development, and the Demographic Transition in the United States." *MPRA Paper No. 88869* .

Kelly, Morgan and Cormac Ó Gráda. 2014. "Speed under sail, 1750-1850." *UCD Centre for Economic Research Working Paper Series, No. WP14/10* .

Kleibergen, Frank and Richard Paap. 2006. "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics* 133(1):97–126.

Klein Goldewijk, Kees, Arthur Beusen and Peter Janssen. 2010. "Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1." *The Holocene* 20(4):565–573.

Klemp, Marc and Jacob Weisdorf. 2019. "Fecundity, Fertility and The Formation of Human Capital." *The Economic Journal* 129:925–960.

Lardner, Dionysius. 1850. *Railway Economy: A treatise on the new art of transport, its management, prospects, and relations, commercial, financial, and social.* London: Taylor, Walton, and Maberly.

Livi Bacci, Massimo. 1977. *A History of Italian Fertility During the Last Two Centuries*. Princeton University Press.

Livi Bacci, Massimo. 1986. "Fertility, Nutrition, and Pellagra: Italy during the Vital Revolution." *The Journal of Interdisciplinary History* 16(3):431–454.

Mammen, Kristin and Christina Paxson. 2000. "Women's work and economic development." *Journal of Economic Perspectives* 14(4):141–164.

Martí-Henneberg, Jordi. 2013. "European integration and national models for railway networks (1840–2010)." *Journal of Transport Geography* 26:126–138.

Martí-Henneberg, Jordi. 2021. *European Regions, 1870-2020*. Springer.

Martí-Henneberg, Jordi. 2023. *Creative Ways to apply Historical GIS: Promoting Research and Teaching about Europe*. Springer Nature.

Melander, Eric. 2021. "Transportation Technology, Individual Mobility and Social Mobilisation." *Cage Working Paper no. 471* .

Modelski, George. 2003. *World Cities*. Faros 2000.

MPIDR. 2013. "[Max Planck Institute for Demographic Research] and CGG [Chair for Geodesy and Geoinformatics, University of Rostock]: MPIDR Population History GIS Collection – Europe (partly based on EuroGeographics for the administrative boundaries). Rostock.".

Murphy, Tommy E. 2015. "Old habits die hard (sometimes) Can département heterogeneity tell us something about the French fertility decline?" *Journal of Economic Growth* 20:177–222.

Murtin, Fabrice. 2013. "Long-term determinants of the demographic transition, 1870–2000." *Review of Economics and Statistics* 95(2):617–631.

Nicolas, P. 1973. "Brief history of railway speed progress." *Federation Nationale des Trav Publ & des Synd Aff* .

Nilsson, Charlie, Kerstin Enflo, Håkan Lobell and Olle Krantz. 2023. "Swedish Historical Regional Accounts: 2023 Update, Revision and Underlying Principles." *Ekonomisk-historiska institutionen, Lunds universitet* .

O'Brien, Patrick. 1983. *Railways and the Economic Development of Western Europe, 1830-1914* . St. Martin's Press.

Olivetti, Claudia. 2013. "The female labor force and long-run development: the American experience in comparative perspective." *NBER Working Paper No. w19131* .

Özak, Ömer. 2010. "The Voyage of Homo-Economicus: Some Economic Measures of Distance." *Working Paper, Brown University* .

Özak, Ömer. 2018. "Distance to the Pre-Industrial Technological Frontier and Economic Development." *Journal of Economic Growth* 23(2):175–221.

Reba, Meredith, Femke Reitsma and Karen Seto. 2016. "Spatializing 6,000 Years of Global Urbanization from 3700 BC to AD 2000." *Scientific Data* 3(1):1–16.

Redding, Stephen J and Daniel M Sturm. 2008. "The costs of remoteness: Evidence from German division and reunification." *American Economic Review* 98(5):1766–1797.

Redding, Stephen J and Esteban Rossi-Hansberg. 2017. "Quantitative spatial economics." *Annual Review of Economics* 9:21–58.

Robert, Gianni. 1964. *Le ferrovie nel mondo.* Vallardi.

Rosés, Joan and Nikolaus Wolf. 2020. "Rosés-Wolf-database." *Version 6* .

Schivelbusch, Wolfgang. 1978. "Railroad space and railroad time." *New German Critique* (14):31–40.

Schultz, T Paul. 1985. "Changing world prices, women's wages, and the fertility transition: Sweden, 1860-1910." *Journal of Political Economy* 93(6):1126–1154.

Spolaore, Enrico and Romain Wacziarg. 2022. "Fertility and modernity." *The Economic Journal* 132(642):796–833.

Tang, John P. 2017. "The engine and the reaper: industrialization and mortality in late nineteenth century Japan." *Journal of Health Economics* 56:145–162.

Vandenbroucke, Guillaume. 2014. "Fertility and wars: The case of World War I in France." *American Economic Journal: Macroeconomics* 6(2):108–136.

Voigtländer, Nico and Hans-Joachim Voth. 2013. "How the West "invented" fertility restriction." *American Economic Review* 103(6):2227–2264.

Wanamaker, Marianne H. 2012. "Industrialization and fertility in the nineteenth century: Evidence from South Carolina." *The Journal of Economic History* 72(1):168–196.

Watkins, Susan Cotts and Ansley Johnson Coale. 1986. *The decline of fertility in Europe.* Princeton University Press.

Weir, David R. 1984. "Life under pressure: France and England, 1670–1870." *The Journal of Economic History* 44(1):27–47.

Zimran, Ariell. 2020. "Transportation and health in the antebellum United States, 1820–1847." *The Journal of Economic History* 80(3):670–709.

APPENDIX B. FURTHER DATA DESCRIPTION

FIGURE B.1. Fertility data: availability

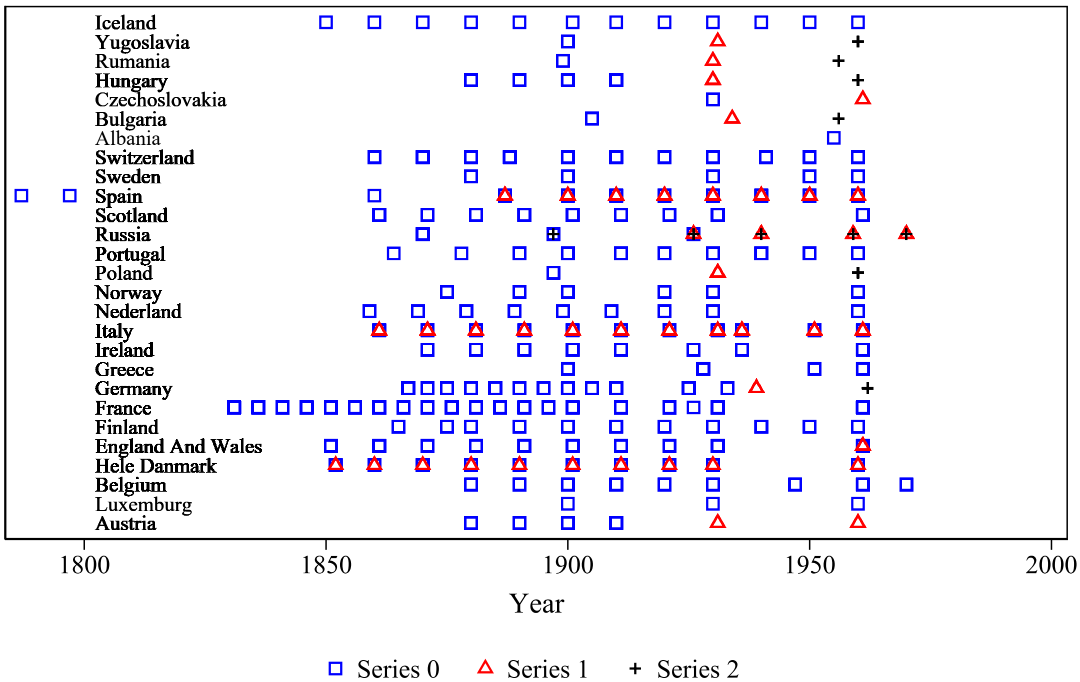
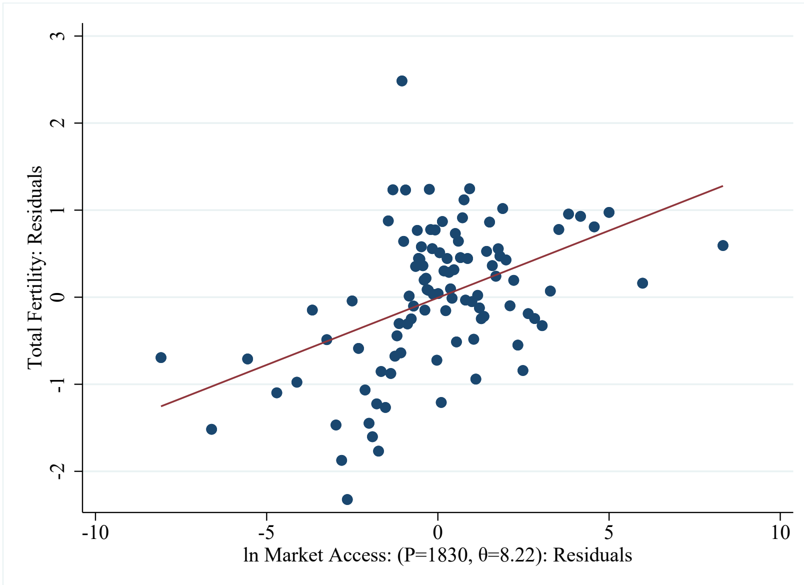


TABLE B.1. Summary statistics

	(1) mean	(2) sd	(3) min	(4) max	(5) N
Total Fertility	30.2	8.93	5.90	68.1	4,104
ln Market Access: (P=1830, θ=8.22)	6.98	7.91	-22.1	27.7	4,206
Latitude	49.6	5.75	35.2	70.0	4,206
Longitude	5.14	9.00	-18.6	29.7	4,206
Coast Distance	97.7	134	0	614	4,206
River	0.53	0.50	0	1	4,206
Altitude	318	340	-144	2,186	4,206
Population Density 1830	77.8	179	0.15	2,599	4,206
Area	7,138	12,013	12.3	166,762	4,206
Caloric Suitability	8,000	2,653	0	14,514	4,206
Barley Suitability	7,442	2,230	0	10,604	4,206
Maize Suitability	3,493	3,994	0	14,527	4,206
Rye Suitability	4,567	1,361	0	6,383	4,206
Oat Suitability	2,992	797	0	3,681	4,206
Wheat Suitability	7,286	2,158	0	10,303	4,206
Average Precipitation	73.5	25.2	28.7	231	4,206
Ruggedness	13.7	13.0	0.41	82.0	4,206
Year	1,895	25.1	1,840	1,940	4,206

Appendix C. Further results

Figure C.1. Binned scatterplot: net of fixed effects



This figure represents estimates of equation (1) as a binned scatterplot with 100 bins. We residualize both fertility and market access to be net of both location and decade fixed effects. The corresponding binned scatterplot and best linear fit are shown.

Table C.1. Results: first stage

	(1) ln Market Access: (P=1830 θ=8.22)	(2) ln Market Access: (P=1830 θ=8.22)
ln Distant Market Access: 500 km	1.225***	1.705***
	(0.053)	(0.074)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

FIGURE C.2. Counterfactual fertility with 1830 market access

This figure represents the counterfactual exercise explained in Section 4.

Appendix D. Further evidence: mechanisms

Table D.1. Marital and non-marital fertility

	(1) Marital Fertility	(2) Marital Fertility	(3) Non-Marital Fertility	(4) Non-Marital Fertility
ln Market Access: (P=1830, θ=8.22)	0.268***	0.197***	0.046***	0.033***
	(0.055)	(0.052)	(0.011)	(0.011)
N	4,040	4,040	4,034	4,034
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

Table D.2. Urban fertility (1/2)

	(1) Urban Fertility	(2) Urban Fertility	(3) Urban Marital Fertility	(4) Urban Marital Fertility
ln Market Access: (P=1830, θ=8.22)	0.024	0.027	0.102	0.016
	(0.061)	(0.064)	(0.114)	(0.111)
N	687	687	718	718
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.3. Urban fertility (2/2)

	(1) Urban Non-Marital Fertility	(2) Urban Non-Marital Fertility	(3) Urban Nuptiality	(4) Urban Nuptiality
ln Market Access: (P=1830, θ=8.22)	-0.044*	0.029	-0.132**	-0.017
	(0.023)	(0.027)	(0.057)	(0.047)
N	687	687	737	737
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.4. Rural fertility (1/2)

	(1) Rural Fertility	(2) Rural Fertility	(3) Rural Marital Fertility	(4) Rural Marital Fertility
ln Market Access: (P=1830, θ=8.22)	-0.007	-0.002	0.168**	0.022
	(0.046)	(0.049)	(0.075)	(0.075)
N	888	888	936	936
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.5. Rural fertility (2/2)

	(1) Rural Non-Marital Fertility	(2) Rural Non-Marital Fertility	(3) Rural Nuptiality	(4) Rural Nuptiality
ln Market Access: (P=1830, θ=8.22)	-0.020	-0.024	-0.121**	0.041
	(0.024)	(0.029)	(0.048)	(0.046)
N	888	888	984	984
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.6. Intermediate outcomes (1/2)

	(1) Nuptiality	(2) Nuptiality	(3) Infant Mortality	(4) Infant Mortality
ln Market Access: (P=1830, θ=8.22)	0.033	0.048*	-0.015	0.059
	(0.027)	(0.026)	(0.039)	(0.039)
N	4,074	4,074	1,606	1,606
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.7. Intermediate outcomes (2/2)

	(1) Pct. Married by 50	(2) Pct. Married by 50	(3) Pct. Urban	(4) Pct. Urban
ln Market Access: (P=1830, θ=8.22)	0.067**	0.030	-0.022	0.206
	(0.027)	(0.028)	(0.136)	(0.000)
N	2,401	2,401	898	898
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.8. Results by GDP in 1880

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.293***	0.235***	0.044	0.012
	(0.037)	(0.049)	(0.048)	(0.044)
N	1,666	1,666	1,626	1,626
Sample	GDP in 1880 with imputed Above Median	GDP in 1880 with imputed Above Median	GDP in 1880 with imputed Below Median	GDP in 1880 with imputed Below Median
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.9. Results by population in 1900

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.254***	0.161***	0.035	-0.001
	(0.040)	(0.054)	(0.044)	(0.039)
N	1,757	1,757	1,749	1,749
Sample	ln RW Population Above Median in 1900	ln RW Population Above Median in 1900	ln RW Population Below Median in 1900	ln RW Population Below Median in 1900
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.10. Results by labor share in agriculture in 1900

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	-0.043	0.007	0.219***	0.225***
	(0.047)	(0.039)	(0.049)	(0.054)
N	1,740	1,740	1,766	1,766
Sample	Agriculture Share Above Median in 1900	Agriculture Share Above Median in 1900	Agriculture Share Below Median in 1900	Agriculture Share Below Median in 1900
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.11. Results by labor share in industry in 1900

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.207***	0.208***	0.021	0.085*
	(0.048)	(0.053)	(0.049)	(0.044)
N	1,748	1,748	1,758	1,758
Sample	Industry Share Above Median in 1900	Industry Share Above Median in 1900	Industry Share Below Median in 1900	Industry Share Below Median in 1900
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.12. Results by labor share in services in 1900

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.248***	0.261***	-0.013	0.004
	(0.050)	(0.059)	(0.044)	(0.034)
N	1,760	1,760	1,746	1,746
Sample	Services Share Above Median in 1900	Services Share Above Median in 1900	Services Share Below Median in 1900	Services Share Below Median in 1900
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.13. Results by years of schooling in 1870

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.121**	0.033	0.223***	0.141***
	(0.048)	(0.046)	(0.035)	(0.038)
N	1,234	1,234	2,810	2,810
Sample	Above Median Years of Education in 1870	Above Median Years of Education in 1870	Below Median Years of Education in 1870	Below Median Years of Education in 1870
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.14. Results by numeracy in 1880

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.009	0.030	0.291***	0.177***
	(0.041)	(0.039)	(0.047)	(0.040)
N	1,932	1,932	2,022	2,022
Sample	Above Median Numeracy in 1880	Above Median Numeracy in 1880	Below Median Numeracy in 1880	Below Median Numeracy in 1880
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE D.15. Results by numeracy in 1900

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	-0.059	0.009	0.191***	0.138***
	(0.043)	(0.000)	(0.059)	(0.040)
N	1,740	1,740	1,726	1,726
Sample	Above Median Numeracy in 1900	Above Median Numeracy in 1900	Below Median Numeracy in 1900	Below Median Numeracy in 1900
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

APPENDIX E. EVIDENCE: ALTERNATIVE MECHANISMS

TABLE E.1. Fertility access

	(1) Total Fertility	(2) Total Fertility
ln Weighted Fertility Access (θ=8.22)	0.287	0.666
	(0.640)	(0.437)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE E.2. Pairwise results

	(1) AD. Total Fertility	(2) AD. Total Fertility	(3) ln Fertility Difference	(4) ln Fertility Difference
Travel Time	-0.031***		-0.006***	
	(0.005)		(0.001)	
ln Travel Time		-1.359***		-0.238***
		(0.283)		(0.046)
N	901,455	901,455	897,947	897,947
Pair and Year FE	Yes	Yes	Yes	Yes

***Significant at 1%, **Significant at 5%, *Significant at 10%. This table reports estimates of equation (7), which is described in the text.

TABLE E.3. Mortality access

	(1) Total Fertility	(2) Total Fertility
ln Weighted Mortality Access (θ=8.22)	-3.529**	-1.865
	(1.396)	(1.332)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE E.4. By life expectancy in 1870

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.276***	0.272***	0.158***	0.023
	(0.043)	(0.047)	(0.032)	(0.038)
N	1,753	1,753	4,056	2,291
Sample	Above Median Life Expectancy in 1870	Above Median Life Expectancy in 1870	Below Median Life Expectancy in 1870	Below Median Life Expectancy in 1870
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE E.5. Control for railway density

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.153***	0.146***
	(0.031)	(0.030)
Rail Density	-25.722***	-21.105***
	(7.161)	(5.700)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE E.6. Sectoral shares as outcomes

	(1) Agriculture Share	(2) Agriculture Share	(3) Industry Share	(4) Industry Share	(5) Services Share	(6) Services Share
ln Market Access: (P=1830, θ=8.22)	-0.002*	-0.001	0.002	0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	1,743	1,743	1,743	1,743	1,743	1,743
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE E.7. Control for the urbanization rate

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.140***	0.115***
	(0.032)	(0.030)
HYDE Urbanization Rate	-16.080***	-10.595***
	(2.263)	(2.174)
N	4,026	4,026
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses. All specifications also control for urbanization.

TABLE E.8. Control for country-level education

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.135***	0.105***
	(0.043)	(0.037)
N	3,665	3,665
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses. All specifications also control for country-level years of schooling.

TABLE E.9. Control for country-level numeracy

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.163***	0.105***
	(0.031)	(0.029)
N	3,463	3,463
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses. All specifications also control for country-level numeracy.

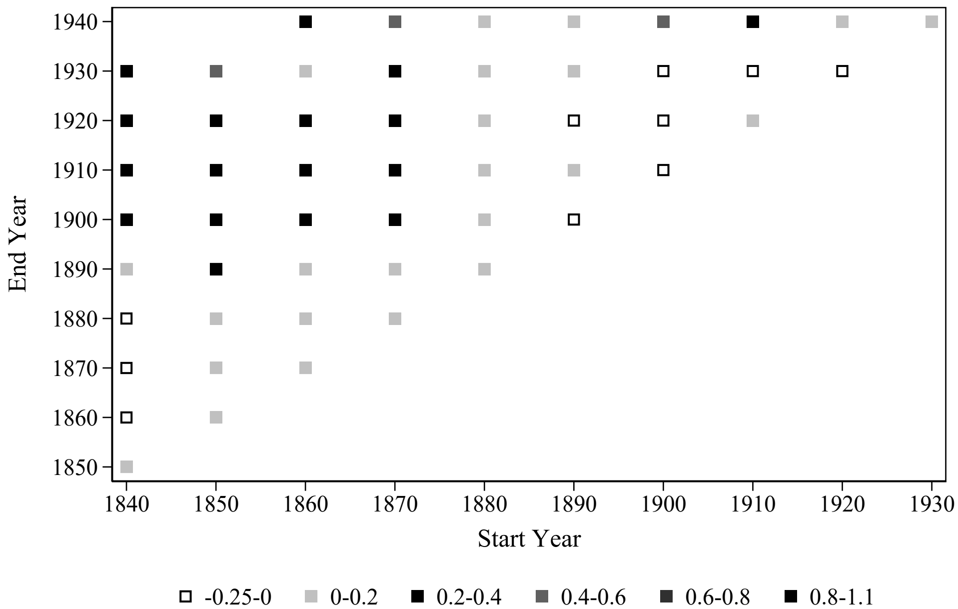
TABLE E.10. Drop capitals

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.161***	0.121***
	(0.033)	(0.030)
N	3,925	3,925
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

APPENDIX F. ADDITIONAL ROBUSTNESS

FIGURE F.1. Every Long Difference



This figure shows estimates of β obtained by estimating equation (8) for every pair of start and end years in the dataset.

TABLE F.1. Long differences IV: 1870 to 1910

	(1) Δ Total Fertility	(2) Δ Total Fertility
Δ ln Market Access	0.562***	0.336***
	(0.104)	(0.110)
N	347	347
Controls	No	Yes
KPF	89.13	141.4

***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Controls are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Robust standard errors in parentheses.

TABLE F.2. Alternative market access (1/6)

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=1)	3.721***	4.395***		
	(0.548)	(0.827)		
ln Market Access: (P=1830, θ=3.60)			0.526***	0.437***
			(0.092)	(0.095)
N	4,056	4,056	4,056	4,056
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.3. Alternative market access (2/6)

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=12.86)	0.095***	0.075***		
	(0.019)	(0.018)		
ln Market Access: (P=1850, θ=8.22)			0.157***	0.124***
			(0.032)	(0.030)
N	4,056	4,056	4,056	4,056
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.4. Standardized β by θ

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.111***			
	(0.026)			
ln Market Access: (P=1830, θ=1)		0.272***		
		(0.051)		
ln Market Access: (P=1830, θ=3.60)			0.149***	
			(0.032)	
ln Market Access: (P=1830, θ=12.86)				0.106***
				(0.025)
N	4,056	4,056	4,056	4,056
Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses. Standardized coefficients reported.

TABLE F.5. Alternative Market Access (3/6)

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=t, θ=8.22)	0.131***	0.103***
	(0.032)	(0.030)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.6. Alternative market access (4/6)

	(1) Total Fertility	(2) Total Fertility
ln Market Access (θ=8.22, P=FT 1830)	0.158***	0.125***
	(0.032)	(0.030)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.7. Alternative market access (5/6)

	(1) Total Fertility	(2) Total Fertility
ln Market Access (θ=8.22, P=Cities in 1830)	0.173***	0.131***
	(0.032)	(0.031)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.8. Alternative market access (6/6)

	(1) Total Fertility	(2) Total Fertility
ln Market Access (θ=8.22, P=Cities in 1800)	0.178***	0.130***
	(0.033)	(0.033)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.9. Market access with cities in 1850

	(1) Total Fertility	(2) Total Fertility
ln Market Access (θ=8.22, P=Cities 1850)	0.159***	0.129***
	(0.032)	(0.030)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.10. Rail speed of 30km per hour

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22) 30km h	0.211***	0.169***
	(0.045)	(0.041)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.11. Rail speed of 120km per hour

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22) 120km h	0.122***	0.093***
	(0.024)	(0.023)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.12. With border costs

	(1) Total Fertility	(2) Total Fertility
ln Market Access with borders (θ=8.22, P=1830)	0.157***	0.124***
	(0.031)	(0.030)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.13. With roads

	(1) Total Fertility	(2) Total Fertility
ln Market Access with roads (θ=8.22, P=1830)	0.131***	0.107***
	(0.031)	(0.029)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.14. With steam travel over water

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22) with steam	0.156***	0.123***
	(0.032)	(0.030)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

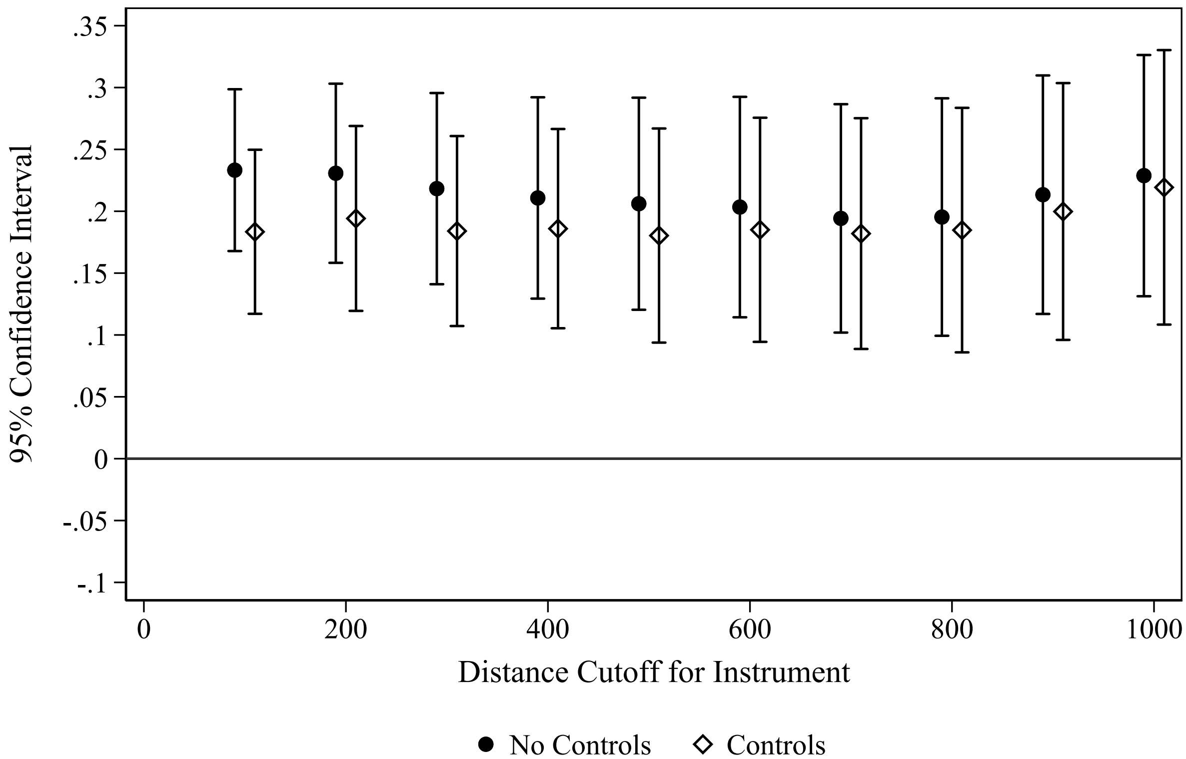
Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.15. Conley standard errors

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.158*	0.125***
250 km	(0.086)	(0.043)
500 km	(0.076)	(0.050)
750 km	(0.050)	(0.050)
1000 km	(0.020)	(0.049)
N	4,104	4,104
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors in parentheses computed using the Colella et al. (2019) implementation of Conley (1999) standard errors with distance cutoffs as indicated.

FIGURE F.2. Alternative IV cutoff distances



These results depict estimates of estimating equation (1) by IV, using a distant market access instrument that employs the distance cutoff given on the horizontal axis.

TABLE F.16. ln Fertility

	(1) ln Total Fertility	(2) ln Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.007***	0.005***
	(0.001)	(0.001)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.17. Country trends

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.055**	0.063**
	(0.027)	(0.026)
N	4,056	4,056
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses. All specifications also control for country-specific linear trends.

TABLE F.18. Country-year fixed effects

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.092***	0.084***
	(0.025)	(0.026)
N	4,044	4,044
Fixed Effects	Yes	Yes
Controls	No	Yes

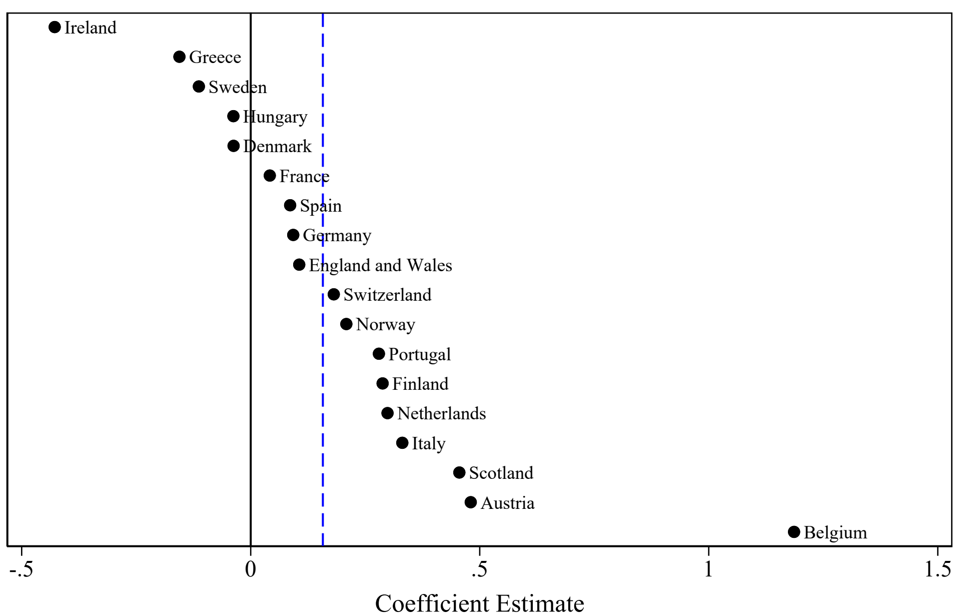
Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses. All specifications also control for country-by-year fixed effects.

TABLE F.19. Controlling for coal

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.158***	0.124***	0.155***	0.125***
	(0.031)	(0.030)	(0.032)	(0.030)
N	4,056	4,056	4,056	4,056
Coal Control	Coal Share	Coal Share	Carbon Share	Carbon Share
Fixed Effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

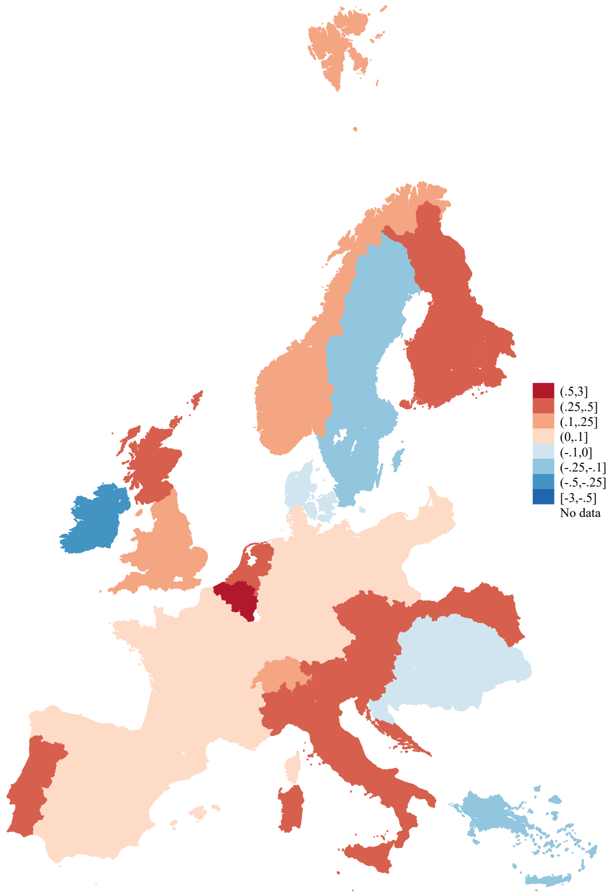
Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

FIGURE F.3. Results by country: Plot



These results depict estimates of estimating equation (1) separately by country.

FIGURE F.4. Results by country: Map



These results depict estimates of estimating equation (1) separately by country.

TABLE F.20. Drop Belgium

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.159*** (0.032)	0.125*** (0.030)
N	3,993	3,993
Fixed Effects	Yes	Yes
Controls	No	Yes

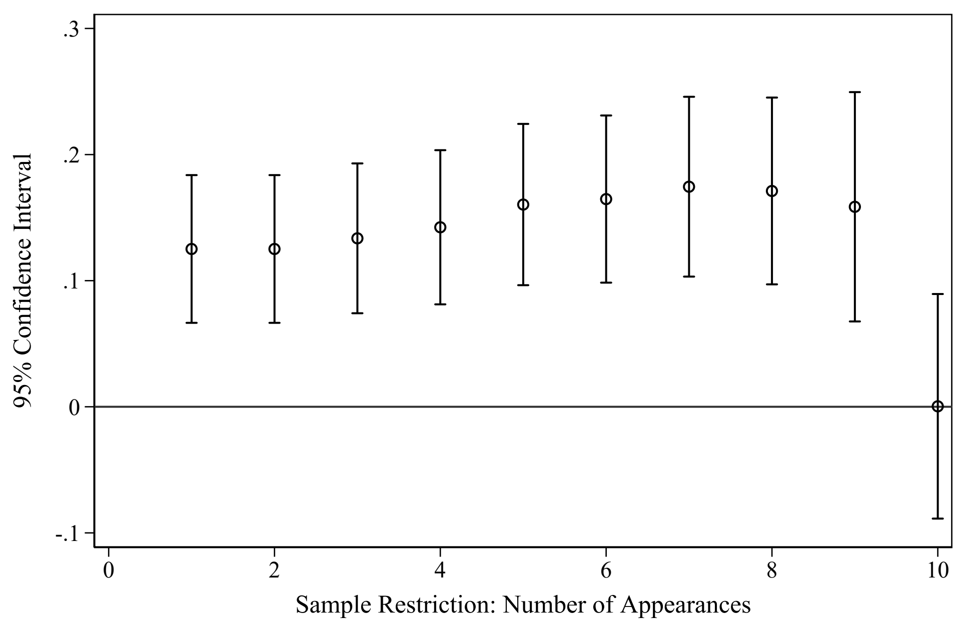
Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.21. Smaller regions only

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.231***	0.173***
	(0.038)	(0.041)
N	2,437	2,437
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

FIGURE F.5. Restrict sample by appearances



These results depict estimates of estimating equation (1) on the sub-sample of locations that appear at least x times in the data, where x is the number on the horizontal axis.

TABLE F.22. Results by time period

	(1) Total Fertility	(2) Total Fertility	(3) Total Fertility	(4) Total Fertility	(5) Total Fertility	(6) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.242***	0.102***	-0.089	0.012	-0.039**	0.003
	(0.039)	(0.038)	(0.133)	(0.057)	(0.016)	(0.019)
N	2,447	2,447	1,005	1,005	416	416
Period	1870 to 1910	1870 to 1910	After 1910	After 1910	Before 1870	Before 1870
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

TABLE F.23. Before 1914

	(1) Total Fertility	(2) Total Fertility
ln Market Access: (P=1830, θ=8.22)	0.199***	0.127***
	(0.024)	(0.026)
N	2,996	2,996
Fixed Effects	Yes	Yes
Controls	No	Yes

Notes: ***Significant at 1%, **Significant at 5%, *Significant at 10%. All specifications include a constant. Fixed effects are for location and decade. Time-invariant controls interacted with decade fixed effects are latitude, longitude, caloric suitability, coast distance, river, altitude, population density in 1830, area, average precipitation, ruggedness, and suitability for barley, maize, rye, oats and wheat. Standard errors clustered by location in parentheses.

APPENDIX G. MAPS OF CONTROLS

FIGURE G.1. Latitude

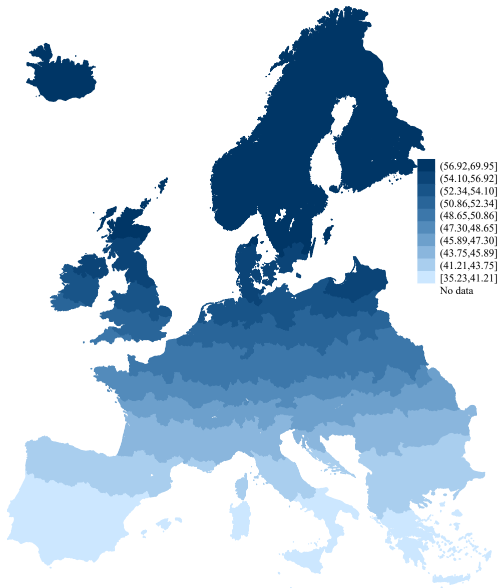


FIGURE G.2. Longitude

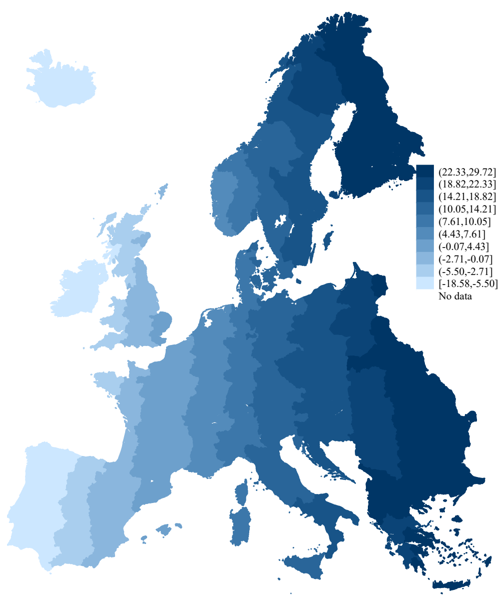


FIGURE G.3. Caloric Suitability

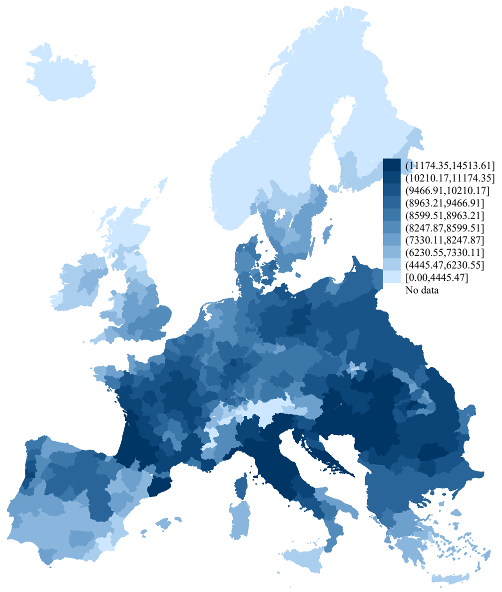


FIGURE G.4. Wheat Suitability

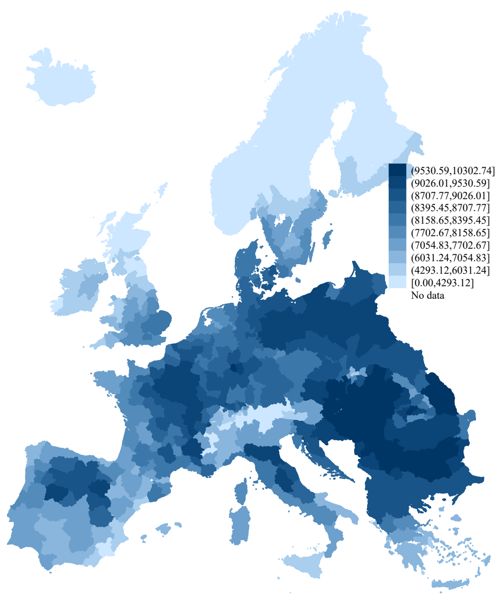


FIGURE G.5. Coast Distance

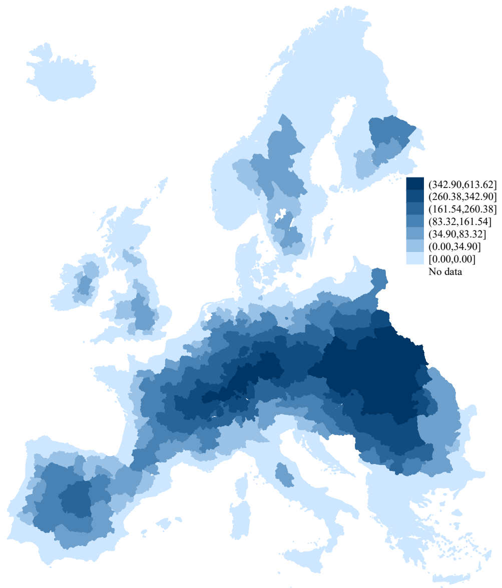


FIGURE G.6. Average Precipitation

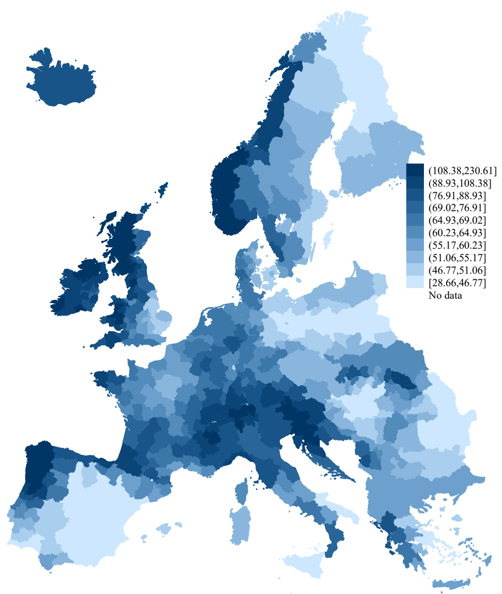


FIGURE G.7. Altitude

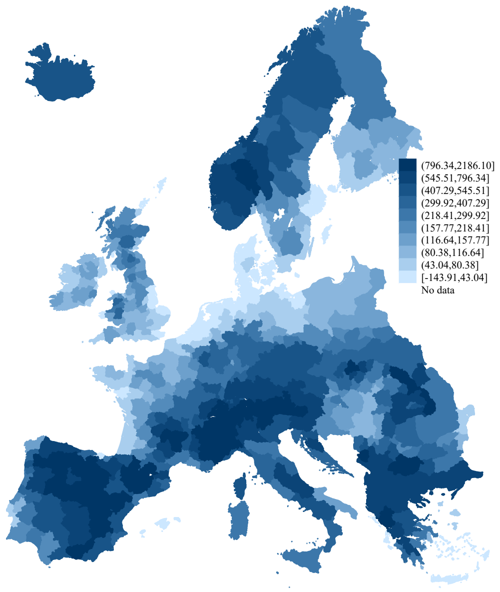


FIGURE G.8. Barley Suitability

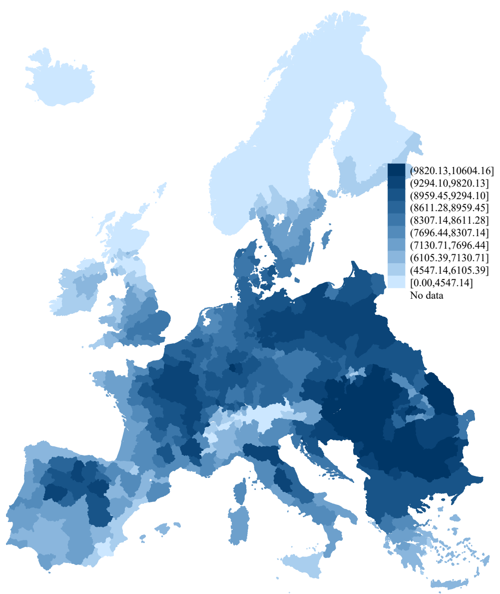


FIGURE G.9. Maize Suitability

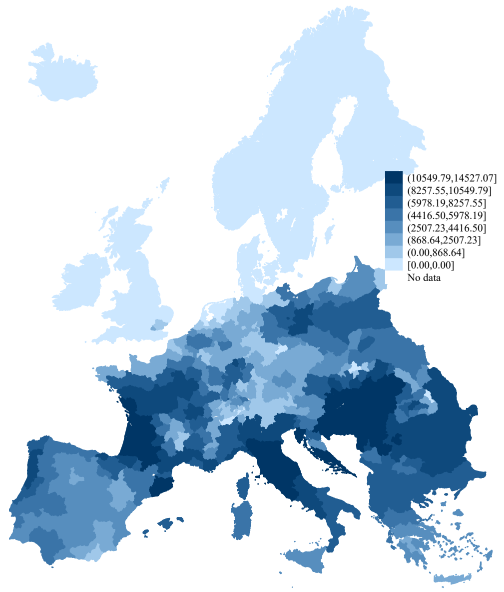


FIGURE G.10. Oat Suitability

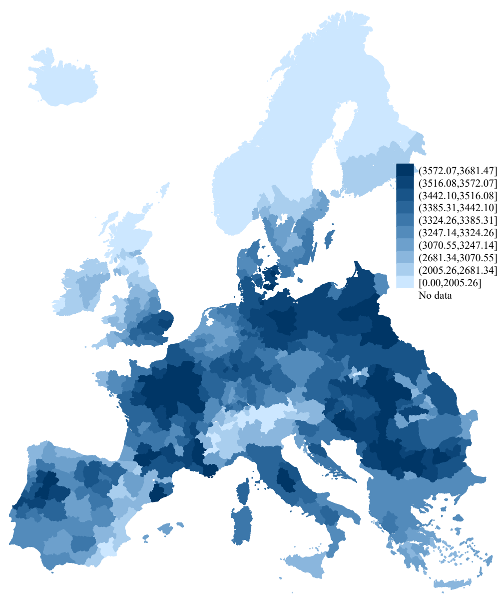


FIGURE G.11. Rye Suitability

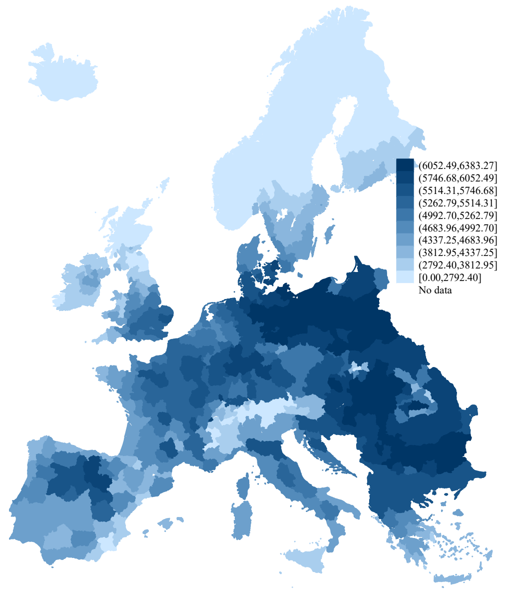


FIGURE G.12. Pop. Density 1830

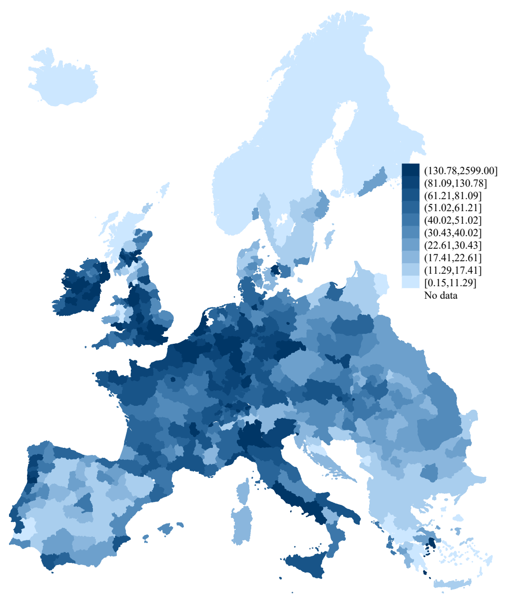


FIGURE G.13. River

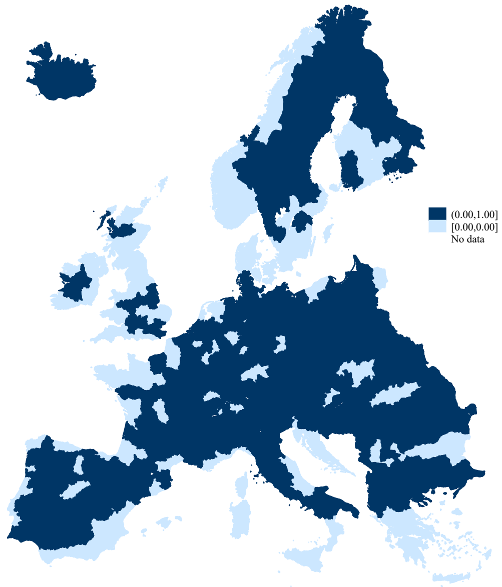


FIGURE G.14. Area

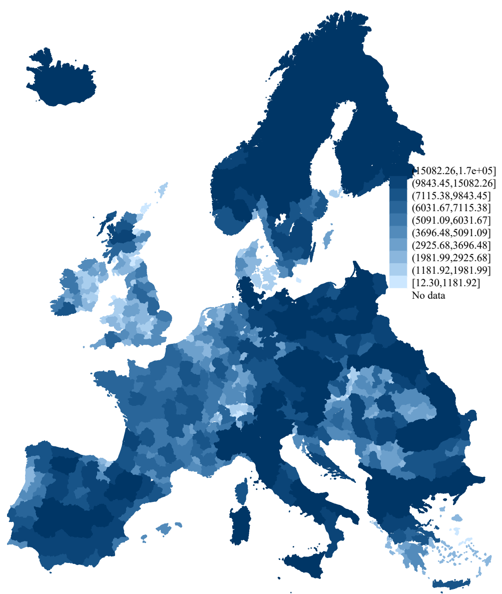


FIGURE G.15. Ruggedness