DISCUSSION PAPER SERIES

DP13348
(v. 2)

Female autonomy generates superstars in long-term development: Evidence from 16th to 19th century Europe

Jörg Baten and Alexandra de Pleijt

ECONOMIC HISTORY

CEPR

ISSN 0265-8003

Female autonomy generates superstars in long-term development: Evidence from 16th to 19th century Europe

Jörg Baten and Alexandra de Pleijt

Discussion Paper DP13348
First Published 01 December 2018
This Revision 12 May 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Female autonomy generates superstars in long-term development: Evidence from 16th to 19th century Europe

Abstract

Many countries did not accumulate sufficient human capital to be successful, because they did not make use of the potential of the female half of their population. Other countries did the opposite and became “superstars” and pioneers in long-term human capital development. This view is supported by studying female autonomy and numeracy indicators of 27 countries and 153 regions in Europe between 1500 and 1900. We are using the demographic indicator age at marriage as a proxy indicator for female autonomy. We approach endogeneity issues by exploiting exogenous variation in gender-biased agricultural specialization.

JEL Classification: N13, N33, O40

Keywords: human capital formation, Female autonomy, Early modern growth

Jörg Baten - joerg.baten@uni-tuebingen.de
University of Tuebingen, CESifo and CEPR and CEPR

Alexandra de Pleijt - sandradepleijt@gmail.com
University of Oxford

Acknowledgements
We thank the conference and workshop audiences at the 2017 European Historical Economics Society Conference, the 18th World Economic History Conference, the CEPR/NYU/CAGE Economic History Conference, the WEast confrence in Vienna, as well as the seminar participants at the University of Oxford and Vienna University of Economics and Business for helpful comments and suggestions. We are grateful to Steve Broadberry, Bishnu Gupta, Markus Lampe, Avner Offer, Deb Oxley, Paul Sharp and Jan Luiten van Zanden for various useful suggestions. Alexandra de Pleijt has benefitted from funding made available by the Netherlands organisation for Scientific Research (NWO), project number 446-16-014. Baten thanks the German Science Foundation for their grant SFB 1070, DFG-No. 215859406.

Female autonomy generates superstars in long-term human capital development: Evidence from 16th to 19th century Europe

Joerg Baten, University of Tuebingen, CESifo and CEPR
joerg.baten@uni-tuebingen.de

Alexandra M. de Pleijt, Utrecht University
sandradepleijt@gmail.com

Abstract: Many countries did not accumulate sufficient human capital to be successful, because they did not make use of the potential of the female half of their population. Other countries did the opposite and became "superstars" and pioneers in long-term human capital development. This view is supported by studying female autonomy and numeracy indicators of 27 countries and 153 regions in Europe between 1500 and 1900. We are using the demographic indicator age at marriage as a proxy indicator for female autonomy. We approach endogeneity issues by exploiting exogenous variation in gender-biased agricultural specialization.

Keywords: Human capital formation, Female autonomy, Early modern growth.

JEL codes: N13, N33, O40.

Acknowledgements: We thank the conference and workshop audiences at the 2017 European Historical Economics Society Conference, the 18th World Economic History Conference, the CEPR/NYU/CAGE Economic History Conference, the WEast confrence in Vienna, as well as the seminar participants at the University of Oxford and Vienna University of Economics and Business for helpful comments and suggestions. We are grateful to Steve Broadberry, Bishnu Gupta, Markus Lampe, Avner Offer, Deb Oxley, Paul Sharp and Jan Luiten van Zanden for various useful suggestions. Alexandra de Pleijt has benefitted from funding made available by the Netherlands organisation for Scientific Research (NWO), project number 446-16-014. Baten thanks the German Science Foundation for their grant SFB 1070, DFG-No. 215859406.

1. Introduction

One important hypothesis about the prosperity or decline of nations says that many societies do not make use of the economic potential and talent of the female half of their population, and therefore remain much poorer and less educated than they could be (Klasen and Lamanna 2009; Sen 1990, among many others). In 2005, for example, the United Nations Secretary General Kofi Annan stated that gender equality is a prerequisite for eliminating poverty, reducing infant mortality and reaching universal education (United Nations, 2005). Traditional gender roles survive astonishingly long and result in women being restricted to household activities and girls having to marry young. Development economists debate this hypothesis nowadays, because experimental and other studies sometimes point to the opposite causal direction: poverty and low levels of human capital might lead to gender discrimination (for a review, see Duflo 2012).

In this study, we assess the hypothesis that higher female autonomy allowed remarkable success in developing numeracy, which is a core component of human capital. As the setting of the study, we chose Europe in its development process from the 16th to the 19th centuries. We contribute new empirical evidence in order to provide a long-term perspective of this core question of global development. As the debate centered on the direction of causality issue, we carefully study endogeneity by employing instruments that can be identified best over long-time horizons. We define the concept "female autonomy" as the capacity for women to achieve high standards of well being and assume a substantive significant role in decision-making, and find evidence that female autonomy was actually decisive for numeracy formation. We are using the demographic indicator "age at marriage" as a proxy indicator for female autonomy. Low age at marriage is substantially correlated with low female autonomy, as we discuss below (Gruber and Szołtysek 2014). Regions in Europe with low marriage age might not have been as extreme as India around 1900, where girls married as early as age 13 in some regions (Krishnan 1977). But also in Europe, there were vast differences between Eastern and Southeastern regions with age at marriages of around 17 or 18, and Denmark with average marriage ages above 29.

We also expand the theoretical approach of the debate about the mechanism by adding the observation that, as women were traditionally responsible for the human capital formation of their offspring, they had to take care of education which took mostly place in households, at least in early societies. In the early modern period, women typically left the labor market when they married. Hence, if women married early, they were not able to gain much independent work experience and they could not provide many relevant labor market skills such as numerical competency (or other skills) to their offspring. Although this might have been the typical pattern of early societies, important differences existed. In some societies, women had more autonomy than in others. For example, they could marry later and therefore potentially develop more labor-market-related skills. A typical example is found in dairy-farming oriented economies in Northern Europe, the Alpine regions, or – to a more limited extent – in Eastern Russia: here, women contributed more to overall household income than in societies that practiced more grain-focused agriculture, in which male upper-body strength was a comparative advantage (such as in other parts of Europe).

For an assessment of the direction of causality in long-term perspective, consistent data had not been available before. Due to this lack of evidence, the link between female autonomy and human capital formation in early modern Europe has not yet been formally tested in a dynamic model (for Eastern Europe see Baten, Szoltysek and Campestrini 2017; and see de Pleijt et al. 2016b for a cross-section). De Moor and van Zanden (2010) have put forward the hypothesis that female autonomy had a strong influence on European history, basing their argument on a historical description of labor markets and the legacy of medieval institutions. They argued that female marriage ages, amongst other components of demographic behavior, might have been a crucial factor for early development in

Northwestern European countries (for a critique, especially on endogeneity issues see Dennison and Ogilvie 2014 and 2016; reply: Carmichael et al. 2016).[1] In a similar vein, Diebolt and Perrin (2013) argued, theoretically, that gender inequality retarded modern economic growth in many countries. Our study is the first to directly assess the growth effects of female autonomy in a dynamic historical context. While we can cover a long time span, the number of countries in the panel analysis is admittedly limited. We address potential issues of representativeness and find our sample to be unbiased for the countries of Europe (see Appendix A). Moreover, the small number of cross-sectional units in the panel is the motivation for adding a second analysis of 153 European regions.

Given the obviously crucial role of endogeneity issues in this debate, we carefully consider the causal nature of the relationship. More specifically, we exploit relatively exogenous variation of (migration adjusted) lactose tolerance and pasture suitability as instrumental variables for female autonomy. The idea is that regions with high lactose tolerance had a high demand for dairy products and allowed cattle farming to cover a high share of total agricultural production (Boehm 1995). In dairy farming, women traditionally had a strong role; this allowed them to participate substantially in income generation (Boserup 1970). In contrast, female participation was limited in grain farming, as it requires substantial upper-body strength (Alesina et al. 2013). Hence, the genetic factor of lactose tolerance and pasture suitability influences long-term differences in gender-specific agricultural specialization. In our instrumental variable regressions, we show that the relationship between female autonomy and human capital is likely to be causal. We also discuss potential violations of the exclusion restriction intensively below. Moreover, we gain insights into identification

[1] Moreover, Denison and Ogilvie (2016) found that the highest marriage ages were not observed in England, which was undergoing the industrial revolution at the time, but rather in Northern and Central Europe.

issues by applying Oster ratio strategies and find that omitted variables would probably not eliminate the female autonomy effect.

In order to solve the issue of missing data, we use age-heaping-based numeracy estimates, as these are available for many countries since the 16^{th} century. Moreover, they reflect a crucial component of human capital formation. Recent evidence documents that numerical skills are the ones that matter most for economic growth. Hanushek and Woessmann (2012) argued that math and science skills were crucial for economic success in the 20th century. They observed that these kinds of skills outperform simple measures of school enrolment in explaining economic development. They apply very sophisticated approaches to deal with causality issues, including special migration analysis. Hanushek and Woessmann (2012, 2020) verify that causality runs from math and science skills to growth and not vice versa, as an old debate about general schooling and growth assumed (Bils and Klenow 2000). Until now, no rejoinder criticized their finding about this direction of causality. This is consistent with recent data on natural experiments. A very convincing historical natural experiment consisted of the Cherokee land distribution, which took place in Georgia in the Southern US in 1832 (Bleakley and Ferrie 2016). The authors studied the winners of this lottery for whom the previous wealth almost doubled. Bleakley and Ferrie traced their children and grandchildren over the 19^{th} century and found that persons who received an unexpected income did not increase the schooling of their children. They used their additional income for other purposes.

Hence, in this study we focus on math-related indicators of basic numeracy. We use two different datasets: firstly, a panel dataset of European countries from 1500 to 1850, which covers a long time horizon. Secondly, we study 153 regions in Europe, stretching from the Ural Mountains in the East to Spain in the West. Using regional evidence has the advantage of avoiding the problem of aggregation on national entities, as well as expanding the relatively small number of cross-sectional units in our panel analysis.

We contribute to the development economics literature about the effect of gender inequality on slower development. Amartya Sen (1990) estimated a large number of "missing women"[2] which resulted in skewed sex ratios, and argued that this has been one of history's crucial development hurdles. Stephan Klasen, with various co-authors, used macroeconomic regressions to show that gender inequality has usually been associated with lower GDP growth in developing countries during the last few decades (Klasen and Lamanna 2009; Gruen and Klasen 2008). This resulted in development policies targeted specifically at women. In recent periods, however, a number of doubts have been made public by development economists. Esther Duflo (2012) suggested that there is no automatic effect of gender equality on poverty reduction. She cites, for example, work by Deaton (1989, 1997) who suggested, in a study of India, Côte d´Ivoire and Pakistan, that the overall amount of spending on adult goods (alcohol, cigarettes, adult clothing) was not reduced significantly more after the birth of a boy than when a girl was born. This indirectly suggests that expenditure on children of both genders was roughly equal. Similarly, Khanna et al. (2003) studied mortality rates of boys and girls, and found that girls' mortality deteriorated in crisis situations (see also Rose 1999). Poverty leads to gender inequality, not vice versa, according to this view.

In sum, the chief contribution of our study is to argue for a strong role of female autonomy in the early European human capital revolution and the development of numeracy in particular. More specifically, we find that average age at marriage explains almost 50% of the variation in numeracy between 1500 and 1850. Therefore our model provides crucial insights into understanding the roots of the Industrial Revolution in Northwestern Europe, as well as the convergence of Scandinavia and Central Europe during the "Second Industrial Revolution". It also contributes to the recent literature about growth and development

---

[2] It should be noted that the "missing women" issue is specifically for the case of India.

determinants in the long run (Acemoglu et al. 2005, 2011; Gennaioli and Voth 2015; and many others).

The remainder of our paper is organized as follows. Section 2 presents our estimates on numeracy and discusses its long-run trends. Section 3 discusses the relationship between female autonomy and human capital in pre-industrial Europe. Section 4 introduces our dataset and section 5 presents the panel results. Section 6 studies the sample of 153 regions and provides Oster ratio estimates. Finally, Section 7 summarizes our main findings and discusses the implications.

2. Age-heaping as an indicator of human capital: methods and data

Measuring the production factor "human capital" has never been simple as the concept is broad – comprising health, cognitive abilities, knowledge, and physical skills. Data limitations for 19th century Europe have forced economists to rely on narrow indicators such as school enrolment rates and self-reported literacy rates. Reis (2005) presents literacy estimates for 15 European countries for around 1800. They range widely for males from over 60% in Northwestern Europe to below 20% in parts of Italy and under 10% in Hungary. For the period before 1800, not much evidence is available and it is often based on regionally limited samples and special social groups. Graff (1987) has shown improvements in literacy rates over the 17th and 18th centuries, but only for a handful of European countries: Britain, France, Spain, Sweden and the Netherlands. For the rest of Europe, and for earlier periods, evidence is harder to come by.

We therefore make use of numeracy as an indicator of human capital formation in early modern Europe. Numeracy is available for a substantial set of 27 European countries and 153 regions.

Crayen and Baten (2010) found that the relationship between age-heaping and other human capital indicators is very close for less developed countries after 1950. They calculated age-heaping and illiteracy rates for no less than 270,000 individuals who were organized into 416 regions, ranging from Latin America to Oceania. Their findings indicated that the correlation coefficient with illiteracy was as high as 0.70 and that the correlation with modern student test results for numerical skills was as high as 0.85. They therefore concluded that the age-heaping measure is more strongly correlated with numerical skills than with other educational indicators. Recently, Baten (2021) found a very close correlation of age-heaping based numeracy estimates and math tests of children for a large number of African regions. Age-heaping based numeracy is an established technique in economic history and development economic that was used by hundreds of studies (reviews: Tollnek and Baten 2017, Baten 2021).

In both industrial and agricultural economies, numeracy was clearly a core component of human capital. In agricultural societies, individuals making decisions about the timing of activities had to take a number of issues into account, such as the weather, the status of plants and animals, and other variables (Baten 2016). Weber (1930/1976) and Schumpeter (1950) pointed out that quantitative calculation was at the very heart of modern, rational capitalism (reviewed in Carruthers and Espeland 1991). They traced its roots to the invention of double-entry bookkeeping in late medieval Italy. Goldthwaite (1972) has, moreover, shown that numerous *scuole d'abbaco* thrived in Renaissance Florence. The young sons of the commercial classes already studied a mathematics curriculum in the 15th century that would change little before the 19th century. Likewise, when England started to engage in international trade and shipping in the 17th century, many secondary schools started to offer courses in mathematics, bookkeeping and mensuration.

The age-heaping methodology is based on the tendency of poorly educated people to round their age erroneously. For example, less-educated people are more likely than people with greater levels of human capital to state their age as "30" even if they are in fact 29 or 31 years old. The calculation of the ABCC Index for numeracy is shown here as a derivation of the Whipple Index (Wh):

(1) $$Wh = \left(\frac{(Age25 + Age30 + Age35 + ... + Age60)}{1/5 * (Age23 + Age24 + Age25 + ... + Age62)} \right) \times 100$$

(2) $$ABCC = \left(1 - \frac{(Wh - 100)}{400} \right) \times 100 \text{ if } Wh \geq 100 \text{; else } ABCC = 100$$

In Table 1, numeracy estimates range from 35 to almost 100. The estimates on numeracy for the sub-periods 1500-49, 1600-49, 1700-49 and 1800-49 are mapped in Figures (1) – (4). The early 16th century saw high rates in Central Europe and the Netherlands, whereas Spain and Hungary had lower values. Before the British Industrial Revolution, the highest numeracy rates in 1700-1749 could be found in North-western Europe. Another finding is that Scandinavian countries stand out in terms of numeracy (see also Sandberg 1979). Compared to North-western European countries, human capital formation in Eastern Europe was relatively slow. By 1800-1849, numeracy rates had a strong North-West/South-East gradient, but also deviations from this pattern, for example, Portugal.[3]

3. Relationship between female autonomy and human capital formation

In the empirical analysis of our study, we use the average age at marriage as a proxy for female autonomy. Low age at marriage is usually associated with low female autonomy – one

[3] Perhaps surprisingly, England only had numeracy rates in the second highest category in the early 19th century. It is beyond the scope of this article to discuss this in great detail, but this can probably be attributed to the deskilling phenomenon of the early phase of the Industrial Revolution. The implementation of early technologies in England in the 18th century reduced the demand for skilled workers (de Pleijt and Weisdorf 2017, de Pleijt et al. 2018).

of the extreme cases in 1900 was probably India, where girls married as early as age 13 in some regions, and they had very low female autonomy (Krishnan 1977). But also in Europe, there were vast differences between Russian, Serbian and Bulgarian regions with age at marriages of around 17 or 18, and Denmark with average marriage ages above 29 (the latter corresponding with high female autonomy). We show below that age at marriage is highly correlated with other indicators of female autonomy. In addition to the general indicator function, we argue that age at marriage is particularly interesting because of the microeconomic channel that runs from labor experience to an increase in women's human capital: After marriage, women typically dropped out of the labor market, and switched to work in the household economy (Diebolt and Perrin 2013). Consequently, after early marriage women provided less teaching and self-learning encouragement to their children, including numeracy and other skills. Early-married women sometimes also valued these skills less because they did not "belong to their sphere", i.e., these skills did not allow identification (Baten et al. 2017).

We should note that the skills which both males and females could obtain in early modern labor markets were not very sophisticated. Agriculture represented more than 80% of most economies, and the skills that could be obtained there were not advanced from the modern point of view. We speak here about milk maids, agricultural help of various types, household services and similar work. However, it made a difference whether a woman could participate relatively independently in the labor market, negotiate about contractual issues and gain experience in forming labor teams and solving conflicts. We would argue that this required skills in organizing cooperation, which was more challenging outside of the family. Within families, in contrast, rules were often automatically set (and not depending on abilities).

It is actually not crucial for our theoretical approach whether female autonomy increased labor market participation or whether perhaps – in the alternative scenario – the demand for female labor might have increased the age at marriage, because more skilled female labor was needed. Both factors might go hand in hand, important is the effect on human capital formation.

Below we add some of the main underlying determinants to this causal chain, as we can use these as instrumental variables. Apart from culturally idiosyncratic factors[4], geographic and climatic conditions allowing for dairy farming (see introduction) were among the underlying determinants, and this complements the causal chain. In a nutshell, we would argue: suitability for dairy farming (and other factors) led to higher female labor force participation, and this resulted in more female autonomy (reflected in age at marriage), and this in turn enabled better numeracy for both genders in some of the countries and regions studied here.

4. Data and its potential selectivities

Our data set of numeracy estimates is mostly based on the collection of census and census-similar sources that were published in a large number of studies, and has been compiled in the clio-infra database (www.clio-infra.eu, reviewed by Tollnek and Baten 2017). It applies a common standard of source evaluation, which guarantees that sources are not socially or regionally selective (at least not to a degree that would lead to substantial distortions). Also, composition by sex, age and urban-rural composition has been taken into account. For example, the gender share for the samples included is close to 50 percent females for the age group 23-72, as we would expect for this time and geography (Tollnek and Baten 2017). Tollnek and Baten also assessed the urban share for the most relevant samples and compared this with representative background data. The data were weighted so that the true

[4] For example the religious attitudes of muslims had a persistent effect on the regions in South-western Europe that were ruled by the Ottoman Empire for centuries.

urban share underlying our study comes as close as possible to historical reality. We added a number of cases applying the same standards (on the sources: see Appendix B). We also include some sources which are not fully compatible with these standards, court records of witchcraft accusations and mortality registers. These sources will require systematic robustness tests below. European historical demographers provided a rich fundus of age at marriage data for almost all European countries (recently compiled by Dennison and Ogilvie 2013). We added a set of Portuguese, Eastern Central and Eastern European estimates for early periods, which were collected by Mikolaj Szoltysek, a renowned specialist for Eastern Central European demography (published in Baten et al. 2017; Portugal: Botão Rego et al. 2016). Again, this evidence is assessed for regional, social or gender composition representativeness (Baten 2017).

In order to cross-validate our evidence, we now compare the age at marriage indicator with other proxies for female autonomy. Several indicators have been suggested to approximate female autonomy (Gruber and Szoltysek, 2014). The most relevant ones are: (1) the share of female-headed households reflects whether societies allowed women to take this leading role (even if often only after the husband's death). (2) Whether males only accepted younger wives, reflecting a male power structure, or sometimes also couples with older wives appeared ("older wives" below) and (3) whether some younger women lived independently from their family in other households, for example, when they worked as farmhands or household aids.[5]

Gruber and Szołtysek (2014) have studied the correlation between these indicators for Europe -- mainly between the eighteenth and early twentieth century -- using not less than 700,000 observations (Table 2). They aggregated these on the place and time level. Our age at

[5] Unfortunately, most of these indicators cannot be traced back to the 16th century, when many of the differences of female autonomy and human capital can already be observed, as we will show below.

marriage indicator is almost perfectly negatively correlated with their "share of married women aged 15 to 19", as both proxies measure the same aspects (the correlation is as high as −0.91, p=0.00, not shown in Table 2). More interesting is the comparison with conceptually different indicators of female autonomy. For example, the share of couples in which the wife was older than the husband correlates closely with the share of married women age 15 to 19, with a coefficient of -0.73. The correlation of the same variable with the share of females living outside their family household ("non-kin") is as high as -0.76. Finally, the correlation with the share of female household heads is substantial (though slightly smaller: -0.40, p=0,00). The main result of this cross-validation is that the correlation with other indicators of female autonomy is quite high by early modern standards. In an ideal data situation, we would use all these indicators to form a joint female autonomy index (perhaps combined via principle components analysis). However, in reality only the age at marriage indicator is available for a substantial number of countries. Especially the earlier centuries are not covered by these other indicators in a representative way. Fortunately, given the correlation between the variables, age at marriage is a reasonable indicator that reflects various aspects of female autonomy.

In Figure (5) – (8) we map ages at marriage in four periods between 1500-49 and 1800-49. In the early 16th century map of age at marriage in Europe, we observe that the UK was far ahead. This might have already forecasted some of the rapid growth taking place later on in the UK, which ultimately led to the Industrial Revolution of the late 18th and early 19th century (Kelly et al. 2013). In the middle group, the Scandinavian and Central European countries were situated, whereas Southern Europe had lower age at marriage. In the early 17th century, Germany clearly fell back behind Scandinavia, Belgium, the Netherlands, and the UK, which might have been caused by the Thirty Years War that was particularly devastating in Central Europe. The South of Europe fell back relative to the North, which foreshadows the "Little Divergence" (De Plejit and van Zanden 2016). In the early 18th century, Central

Europe starts to recover in terms of female autonomy, whereas now some of the East Central European economies enter the picture with relatively low rates. This continues in the early 19th century, when especially South Eastern Europe had low rates. In contrast, Scandinavia and Central Europe, including Switzerland and Austria, had quite high female autonomy values. These are exactly those countries which became superstars in the Second Industrial Revolution, which took place shortly thereafter in the late 19th and early 20th century.

The literature has identified other important potential determinants of human capital formation in the early modern period, such as conflict, institutional settings and religion. Both interstate wars and civil and religious wars were potentially devastating (such as the Thirty Years War). Hence, we control for civil war and interstate war, as both were potentially detrimental (Baten and Mumme 2013).

The development of the “Second Serfdom” has been identified as a source of slow economic growth in Eastern Europe by Kula (1976), Millward (1982), and others. Eastern European landowners expanded their previously modest familial manor farms into large-scale domanial economies in the 16th century designed to produce surpluses for sale in the urban markets of Western Europe. This type of seigneurialism led landlords to demand from their peasant subjects not only rents in cash and kind but, above all, in labour services. Serfs, therefore, did have few incentives or opportunities to invest in basic education compared to free farmers. We control for differences in exploitation by including a dummy variable to identify countries that had an “extreme” form of serfdom (more than 30 per cent serfs with severe labor obligations (corvee), following Baten et al. 2017).

Becker and Woessmann (2009) found a link between religion (notably Protestantism) and human capital formation (see also Baten and van Zanden 2008, de Pleijt and van Zanden 2016). In the analysis below, we capture this by controlling for mostly protestant countries.

There are several other control variables such as the level of national income, engagement in international trade and shipping, population density and the quality of institutions. These variables are endogenous and therefore “bad controls”. For this reason we do not include them in the main specification of the models, but we do however include them in several robustness-checks below to study the robustness of our regression results.

5. Empirical analysis

To determine the importance of female autonomy for human capital formation, we explore the empirical relationship between average age at marriage and numeracy, while controlling for potential confounding factors as discussed in the previous section. Our regression analysis consists of two parts. In the first part of the analysis we examine the relationship between female autonomy and numeracy formation in Europe between 1500 and 1850. The unit of observation are countries at intervals of approximately half a century. The periods include the half centuries beginning with 1500, 1600, 1650, 1700, 1750 and 1800 (for 1550, the data source was too insufficient). The number of countries for the period 1500-49 is small, but already 1600-49 is covered by a substantial number of countries (see Table 3). We are using the Maddison strategy to aggregate all available regional data on a national level, using 1990 borders – this allows easier comparisons over time (moreover, the Clio-Infra database on which we partly rely has followed this strategy as well).

In the second part of the analysis we study the relationship between female autonomy and numeracy formation at the regional level in the 19th century. We have very detailed data on age at marriage for 17 countries and Empires (which include several modern countries) – i.e. Bulgaria, Serbia, Hungary (including modern Hungary, Slovakia, parts of Romania, Croatia, Serbia, Ukraine), Italy, and the Russian Empire (including modern Russia, Estonia, Latvia, Lithuania, Belarus, Moldavia, parts of Ukraine) – which allows us to examine the relationship in more detail.

Starting with the early modern period, Figure 9 depicts a strong and positive relationship between age at marriage and numeracy for five periods following 1600. Most countries are close to the regression line. Denmark, the Netherlands, Germany, Sweden and other countries had high values of female autonomy and numeracy – interestingly, many of the countries of the "Second Industrial Revolution" of the late 19th century. In contrast, Russia, Poland, Slovakia, Italy, Spain and Ireland had low values in both periods. There are modest deviations from the regression line: Hungary, and to a lesser extent the UK and Scandinavia had higher numeracy relative to the female autonomy proxy of marriage age. On the lower right side of the regression line, Portugal, Poland, Ireland, Italy, France and Belgium had sometimes lower numeracy than expected based on the female autonomy proxy. But in general, these deviations from the regression line were not substantial.

Of course, the observed relationship between female autonomy and numeracy is not necessarily causal. Higher numeracy and age at marriage may have existed independently, governed by common forces of economic development. To address the issue of endogeneity, we use exogenous variation in migration-adjusted lactose tolerance and the relative soil suitability for pasture as instruments for average marriage ages.

We base the construction of our instruments on the studies by Alesina et al. (2013) and of Voigtländer and Voth (2013) and a vast related literature, who all argued that agricultural specialization influences the relative position of women within the family and in the labor market. Alesina et al. (2013) find that in areas where plough cultivation was widespread, women had a relative disadvantage because this cultivation requires more upper-body strength. Plough cultivation also decreases the female position in the family labor participation due to its low compatibility with other activities, such as childcare. In Europe, the alternative to grain-oriented agriculture (using ploughs) was cattle farming, which was typically associated with a more active role for women. Voigtländer and Voth (2013) similarly suggest that the relative prevalence of animal husbandry over grain cultivation might be an important determinant of differences in age at marriage (Baten et al. 2017 explain this in detail). Animal husbandry benefits the relative bargaining position of women in their society because in this activity, upper-body strength is of smaller relevance. Dairy farming required specific skills that were transferred from mother to daughter (disease prevention, hygienic behavior) were also a substantial advantage.

In general, countries that are characterized by having a relatively high tolerance for lactose were more likely to specialise in dairy farming, because dairy products could be more easily consumed (dairy countries: see Lampe and Sharp 2015).[6] Hence, we would expect higher levels of female autonomy in lactose-tolerant populations.

We compiled the data of Ingram et al. (2009) and Flatz (1995). Ingram et al. (2009) listed almost 450 studies on countries and regions within countries. These studies tested (a) whether the consumption of milk contributed to increases in blood glucose levels and (b) to what extent hydrogen could be measured in exhaled air, which indicates lactose intolerance. Why would lactose tolerance be a good instrument? If a large share of the local population can consume milk, it is more economic to specialize on cattle farming. The background is that most human bodies could not consume substantial amounts of unprocessed cow milk, after passing the weaning age. However, a certain share of human beings experienced a genetic modification which enabled their bodies to produce an enzyme called lactase. This enzyme allowed to consume unprocessed cow milk. Biologists suggest that this modification took place in the phase after the domestication of the cow mostly between 6000 and 4000 years before present, with some regional variation (Rosenstock et al. 2015). Especially in regions particularly suitable for dairy agriculture, such as Northwestern Europe, human beings who inherited this genetic modification had more surviving offspring. Hence, more lactose tolerant persons inhabited the dairy farming environment. This selective survival took place already in the first phase after the domestication of the cow. The shares of lactose tolerant people did not

[6] Boehm (1995) studied the share of milk and found that Central and Northern European agriculture was characterized by a very large share of milk in total value added.

change much over the past four thousand years (Rosenstock et al. 2015). Rosenstock et al. (2015) observed that in England a dramatic increase of height took place during the Copper and Bronze Age (between 6000 and 4000 years before present), which they explain by the genetic change of the population from mostly lactose intolerant to mostly lactose tolerant. Since then, the genetic ability is inherited and mostly constant across regions. During the time frame we are analysing here (1500-1900), lactose tolerance did not change much anymore and was probably exogenous. Migration might have mattered, but Cook (2014) provided data on lactose tolerance levels that was adjusted for migration over the last 500 years. Hence his migration-adjusted evidence for lactose intolerance around 1500 can be reliably used for historical studies in order to avoid potentially endogenous migration effects.

Figure 11 shows our instrument and its geography which is obviously correlated to age at marriage. Northwestern Europe in general was mostly lactose-tolerant (UK 94.6%, Netherlands 85.4%), whereas the East and South of Europe could only digest smaller amounts of milk sugar (Italy 47.6%, Serbia 48.2%). Interesting are again the deviations from this broad pattern. For example, the Czech lands had a relatively high lactose tolerance of 76.2%, which corresponded with the quite high age at marriage and early industrialisation (Komlos 1989). In the IV regressions below, we demonstrate that lactose tolerance is highly correlated with age at marriage. We also assess the exclusion restriction of our instrument by controlling for other potential channels between lactose tolerance and numeracy. The results, however, show that the effect of lactose tolerance is very likely to have run via the female autonomy channel. For the regional regressions below, we use relative soil and climatic suitability of pasture as instrumental variable. We calculate the ratio between the pasture suitability of a region (=good for animal husbandry) over its cereal suitability, using a similar argument for this intrument as for lactose tolerance: it was more economic to specialize in cattle farming there.

We begin our analysis with Least Square Dummy Variable (LSDV) regressions, regressing numeracy by country and half century on age at marriage and the confounding factors discussed above:

$$N_{i,c} = \alpha + \beta_1 MA_{i,c} + X'\gamma + \mu_c + \eta_s + \varphi_c + \varepsilon_{i,c} \,, \qquad (1)$$

where $N_{i,c}$ captures numeracy (of both genders) in country i in half century c. $MA_{i,c}$ is the main variable of interest: average age at marriage in country i in century c. Our data on numeracy is by country and half century, whereas age at marriage is available only for centuries. This means that in the regressions below we assign age at marriage values to two numeracy observations each. We therefore cluster the robust standard errors at the country and century levels which solves the econometric issue related to this. We also experimented with a) linear interpolation, so that each half century has a different value and b) aggregating to centuries, hence, reducing the total number of cases. Both strategies yielded very similar results. The resulting number of clusters is 63 (Table 4, Col. 1). Although this is not a "small" number of clusters according to Cameron and Miller (2015, who defined the threshold between 20 and 50), we also used wild bootstrapped standard errors. In the formula above, μ_c are century fixed effects, X' is the vector of control variables that we described in the previous section, α is a constant, and ε is the error term. φ_c are country fixed effects. In addition, our data on numeracy has been derived from different underlying sources and therefore we control for source type η_s.

Next, we turn to the Two-Stage-Least-Square (TSLS) regressions. In the first stage, age at marriage is instrumented by our indicator for lactose tolerance:

$$MA_{i,c} = \delta_1 LCT_i + X'\gamma + \mu_c + \eta_s + \varphi_r + \nu_{i,c} \,, \qquad (2)$$

where LCT_i is the lactose tolerance of the population in country i. X' is the vector of control variables included in Equation (1), and $\nu_{i,c}$ is the error term.

In Table 4, we provide the LSDV regressions measuring the conditional correlation of age at marriage and numeracy. LSDV specifications are equivalent to panel fixed effects models (we also provide one OLS regression without any dummy variables). We begin with the single variable regression (Col.1), then add country fixed effects (Col.2), time fixed effects (Col. 3), and source fixed effects (Col. 4), using dummy variable specifications. In column (5) we restrict our sample to the countries for which we have information on migration adjusted lactose tolerance, for a later comparison with the IV estimates below. In columns (6) – (9) we add control variables for war, civil war, serfdom, and the "bad control" national income, and in column (10) we combine the control variables war, civil war and serfdom. We observe consistently statistically significant correlations between female autonomy and numeracy. Although the number of clusters is 63 in Column 1, and hence larger than the threshold of 50 which Cameron and Miller (2015) identified as the upper division between small and not small samples, we also calculated the Wild bootstrapped p-values that we observe to be 0.000, indicating clearly statistical significance. Adding time, country and source fixed effects and control variables reduces the coefficient on age at marriage slightly, as expected, but it is still found to be statistically significant and of substantial size.

War, civil war and serfdom can be considered as being exogenous to the relationship studied here (although probably for any variable in macroeconomics and history some endogenous relationship can be constructed). However, for national income the endogenous nature is relatively clear, as it is influenced by numeracy (Hanushek and Woessmann 2020). We still include this "bad control" here, because Carmichael et al. (2016) wondered whether in some specific situations age at marriage declines after a positive income shock. Nevertheless, we still find a strong positive correlation between age at marriage and numeracy whilst controlling for levels of per capita GDP (Col. 9).

In the TSLS regressions of Table 5, we observe strong correlation of our instrument with age at marriage in the first stage results (Panel B). The coefficient on age at marriage in the corresponding second stage is statistically significant, suggesting a positive and causal effect of female autonomy on numeracy in pre-industrial Europe. Although the number of clusters is 63 in Column 1, and hence larger than the threshold of 50 which Cameron and Miller (2015) identified as the upper division between small and not small samples, we also calculated the Wild bootstrapped p-values that we observe to be 0.000, indicating clearly statistical significance. We also executed Montiel Olea Pflueger tests for weak instruments that resulted in a critical value of 23.11 for our number of clusters at the 10% level (Pflueger and Wang 2015).[7] As our F-statistic is 26.1, we find that our instruments are not weak, at least not at the 10% level.

Interestingly, the coefficient on age at marriage in the IV regressions is higher than the coefficient in the OLS regressions (see Col. (5) in Table 4). This can probably be attributed to measurement error in the age at marriage variable.

Is the coefficient of age at marriage economically meaningful? Economists employ different approaches to measure the economic significance of an independent variable. One measure considers the effects of one standard deviation of the explanatory variable. If we multiply the standard deviation of age at marriage (2.56) with its coefficient (6.54 in Col. 2, our preferred model), we obtain 16.74. This is roughly 85% of the standard deviation of the dependent variable numeracy (standard deviation: 19.63). If we use the LSDV coefficient of 4.33 (Col. 3 of Table 4) a standard deviation effect is still 11.09 or 57% of the standard deviation of the dependent variable. Hence the economic importance of this factor is remarkable – it is not only statistically significant. Moreover, in the conclusion below we explain that the difference between age at marriage in NW and SE Europe accounts for a similar numeracy difference as 80% of the differential between Europe and the Global South in 1900 – hence this gap is clearly economically important.

[7] The Montiel Olea Pflueger test is a routine that is robust in situations of heteroscedasticity and autocorrelation, and it provides a cluster–robust weak-instrument test (Pflueger and Wang 2015).

The exclusion restriction is always an issue in IV estimation. Economists are often able to imagine direct effects from almost any potential instrumental variable on the dependent variable, or there might be variables that happen to correlate with omitted variables, which would also affect the exclusion restriction. For example, we could imagine that lactose tolerance might affect numeracy via nutritional benefits from consuming milk, hence the effect would not only run via the female autonomy channel. To assess this potential issue, we include in our regressions average human stature, as height is often used as a proxy for nutritional quality (Baten et al. 2014). [8]

We find that the addition of heights does not render our second stage estimates of the age at marriage coefficient statistically insignificant (Table 6, Col. 1). The coefficient for human stature is not statistically significant, and it does not affect the age at marriage indicator we are mainly interested in: the coefficient of age at marriage barely changes. While we cannot be perfectly sure that the exclusion restriction does not cause problems, we can conclude from this that the effect of lactose tolerance did not run via nutritional benefits from consuming milk.

There may also be autocorrelation effects from the lagged dependent variable on the instrument. To formally test this, we also explore whether the lactose intolerance can be explained by very early human capital values. As early human capital values we are using medieval shares of the so-called birth year known estimates (Keywood and Baten 2020). The idea is that the percentage of rulers, for which the birth year is known, requires a basic numeracy of elite groups in a society. Keywood and Baten found a strong correlation between

[8] Another possibility would be that dairy farming might require a higher numeracy than grain agriculture. However, Galor and Özak (2016) argued that it is especially the grain harvesting regions which required a precise estimation of agricultural outcomes over long periods, planning ahead for months and taking into account harvest information. Hence, while for dairy faming also a high numeracy is required, the same applies to grain agriculture.

this “ruler birth year known” variable and other early educational indicators, such as the number of monastery manuscripts per capita. Hence, it can be shown to serve as a valid indicator for medieval elite numeracy. We find in regressions that it is not significantly influencing our instrument, which supports the validity of this instrument (not shown, available from authors).

Finally, we explore the robustness of our baseline model via the inclusion of additional control variables: (a) population density, (b) political institutions and (c) international trade. We include these controls in a robustness check (rather than in the main regressions, see above), because they might be endogenous (“bad controls”).

(a) One strand of the literature has identified low population density as a cause for low levels of human capital formation (Boucekkine et al. 2007). The idea is that sparse population and the lack of a proper transport system made commuting to schools costlier. We therefore introduce a measure for persons per square kilometre to control for this possibility (Source: Clio-Infra.eu).

(b) Regarding political institutions, North and Weingast (1989), Acemoglu and Robinson (2012) and van Zanden et al. (2012) have argued that the sovereigns had to be constrained in order to protect the property rights of citizens. In democratic systems with strong parliaments, property rights were more secure than in states ruled by absolutist kings. As a consequence, republican systems had lower interest rates at the capital market and this may have translated into faster economic growth and the accumulation of human capital (de Pleijt and van Zanden 2016). On the other hand, more human capital might encourage and enable more political participation, which is why we included it here under “potentially endogenous controls”. To capture this, we use a dummy variable derived from Long and Shleifer (1993) which distinguishes between states governed by “Princes” and those without (absolute) monarchs, the “Republics”. In addition, we also use the activity index of parliaments as a proxy for the quality of political institutions. The index is defined as the number of years a parliament was in session during a century. It varies from zero when no parliament is convened to close to 100 for England after the Glorious Revolution of 1688 (van Zanden et al. 2012).

(c) The empirical analyses of Allen (2003) and Acemoglu et al. (2005) have argued that international trade is a main driver of pre-industrial growth. It may also have been correlated with human capital formation, as literate and numerate societies may have been more likely to engage in international trade and shipping. A first control variable that we use is the log of the volume of Atlantic trade of Acemoglu et al. (2005). A second variable for international trade and shipping that we use is the (per capita) tonnage size of the merchant fleet. De Pleijt and van Zanden (2016) show that this variable is available for a large set of European countries and argue that it captures more general trade flows.

In Table 7 we observe that none of these potential confounders makes the effect of age at marriage insignificant. Both the volume of Atlantic trade and the log of the size of the merchant fleet enter the regression with the expected sign, but only Atlantic trade is statistically significant. Testing for the effect of political institutions, we find that there is a positive association between active parliaments and numeracy, indeed suggesting that the checks and balances on the executive may have been beneficial for numeracy formation. The "Prince" variable has an unexpected sign, but is statistically insignificant. In all regressions the coefficient on age at marriage remains highly significant at the 1% level.

Moreover, we assess the robustness of the IV analysis, using different subsamples. To begin with, our results could be driven by some of the economically most successful countries. Until the early 19th century, the UK and Low Countries developed into a rich part of the continent (Broadberry et al. 2015, van Leeuwen and van Zanden, 2012). Hence, in Table 7 we exclude the United Kingdom (Col.1) or the Netherlands (Col.2).[9]

[9] Excluding both countries at the same time gives very similar results.

Column (3) repeats the exercise excluding Russia, which is another potential extreme case in our sample. It was much poorer than the other countries in our sample, numeracy was relatively low by international standards, and women married very early.

Another potential concern is that part of the evidence on numeracy has been derived from witch and death records. Both sources are likely to yield estimates on numeracy that are biased downwards. We have controlled for this possibility in the regressions in Tables 4 - 6 via the inclusion of dummy variables for the different source types that are used to derive the estimates on numeracy. However, to address this issue head-on, we have estimated the regressions omitting the evidence derived from "death records" (Col. (4) in Table 7) and from "witch trials" (Col. (5)). In column (6) of Table 7 we omit both the evidence from "witch-" and "death records". All the coefficients of age at marriage are still statistically significant and the coefficient changes very little. Finally we regress numeracy also on Protestantism and orthodox confession. This could not be integrated into the LSDV regression framework, as religious confession did not change sufficently after the 16th century (Appendix D).

In sum, we conclude that a wide range of econometric methods suggests that female autonomy is important in determining numeracy formation in early modern Europe.

6. The relationship between female autonomy and human capital in 153 regions

In the second part of the analysis, we turn our attention to the relationship between female autonomy and numeracy formation at the regional level in the 19th century (the sources are documented in Appendix C). More specifically, we regress numeracy by region on age at marriage, using robust standard errors and clustering (see notes to Table 9):

$$N_{r,c} = \alpha + ß_1 \, MA_{r,c} + X'\gamma + \varphi_c + \varepsilon_{r,c} \, , \qquad (3)$$

where $N_{r,c}$ captures numeracy in region r in country c. $MA_{r,c}$ is the main variable of interest: age at marriage in region r in country c. X' is the vector of control variables, which differs slightly from equation (1) as some of the control variables were not available at the regional level. A first control variable that is included is the share of large landowners (the area owned by large landowners who had at least 50 hectares). The share of large landowners is included because Baten and Hippe (2018) recently demonstrated that regional numeracy is strongly influenced by this variable in a large sample of European regions in the 19th century. We again include the share of protestants in order to control for religious effects.

In the first stage, age at marriage is instrumented by relative pasture suitability, which is a measure for how well the soil and climate is suited to pasture and cattle agriculture, relative to grain agriculture. Baten et al. (2017) explore this instrument for eastern European female autonomy and find it highly correlated with age at marriage. It follows a similar logic as lactose tolerance: if regions are more suitable for pasture than for grain, there is a higher likelihood that the region develops a strong cattle agriculture, which is usually associated with higher female autonomy. Hence, we instrument age at marriage in region r and country c ($MA_{r,c}$):

$$MA_{r,c} = \delta_1 PS_r + X'\gamma + \varphi_c + \nu_{i,c} \,, \quad (4)$$

where PS_r is the relative pasture suitability of region r. X' is the vector of control variables included in Equation (3). In the IV regressions, below, we control for heights as pasture suitability may influence numeracy via a nutrition effect (see above). Finally $\nu_{i,c}$ is the error term. In Table 8, we summarize the descriptives for the regions. Numeracy ranges from 21 to 100 percent.

In Table 9, we analyze the effect of regional age at marriage, our proxy for female autonomy, on regional numeracy. The parsimonious basic model in column (1) indicates that there is a statistically significant impact in all European regions for which data is available. We add country fixed effects in order to control for unobserved country heterogeneity, e.g. culture or measurement concepts might be important. Finally, column (3) combines the country fixed effects specifications with additional control variables. All three models indicate that age at marriage has a statistically significant impact on numeracy, even in the subnational dataset.

In an additional regression, we analyze the regional determinants of numeracy separately; either by country or by two countries combined if the sample sizes are not large enough. In column (1) of Table 10, we use Bulgaria and Hungary and find a statistically significant effect of age at marriage on numeracy for 63 observations. In column (2), which displays the results for Serbia, we similarly observe a positive effect even after just including one country. The same applies to a combination of Spain and Italy (Col. 3), as well as to Russia (Col. 4).

We show a scatter diagram of all the regions (see figure 11) and observe that the residual numeracy and the residual age at marriage after controlling for country fixed effects and other control variables yields an upward sloping regression line. Most importantly, the effect is not driven by single outliers. For example, in the Croatian-Italian province of Fiume, we have both a higher residual numeracy and a higher residual age at marriage. The same applies to the north-east Russian province of Vologda and the Italian province of Lazio. In contrast, in central Russian Smolensk and in the Italian region of Basilicata, we have a low residual age at marriage and a low residual numeracy. There is still quite a bit of variation on both sides of the regression line. One of the largest outliers might be St Petersburg, which is obviously determined by the fact that, as the capital of the empire, there is a high level of numeracy. Adding this variable would increase the R-squared.

The results in Table 11 show the IV regressions of all European regions that we can include. As an instrument, we use the relative suitability of pasture because this geographic suitability variable is not likely to be influenced by age at marriage or other variables. The results show a strong relationship between age at marriage and numeracy whilst controlling for the share of landowners (Col. 1) and religion and heights (Col. 2 and 3). Column (4) includes country fixed effects. The F-Statistic indicates that we generally have relatively strong instruments.

In all four instrumental variable regression specifications, we observe a statistically significant effect of age at marriage in the second stage. This is unaffected by whether or not we control for land inequality – which has a negative effect of varying size according to these estimates – for religion, or for country fixed effects. We also add a height variable here to assess the exclusion restriction: Does the relative suitability of pasture influence numeracy only via age at marriage or is there a separate direct effect on numeracy via nutrition? Including the height variable (Col. 2 to 4), which is the indicator of nutritional quality and health, helps to disentangle and to control this potential alternative channel.

A potential issue could also be that cattle prevalence may be related to numeracy via wealth effects. After all, cattle represent a substantial wealth item. However, the effects are unlikely considering the geographic landscape in Europe. The early modern highest cattle densities are not found in the rich regions, which tended to be very urbanized, but in mountainous regions. Due to the rugged terrain, no possibility of grain agriculture existed in these regions. Hence, dairy farming was the best alternative.

Clearly, other variables that are unobserved might matter as well, posing the potential issue of omitted variable bias. Altonji, Elder and Taber (2005) have suggested a method to estimate the selection on unobservables relative to the selection on observables, and Oster (2017) has refined it to include information about the R-squared. We can apply it for our regional sample as we have sufficient observations (rather than in the panel above, which has other strengths). Recently, this method has been applied in a variety of empirical frameworks, including Nunn and Wantchekon (2011) on the long-run effects of slavery. The basic question is 'How large does the effect of unobservables have to be in order to eliminate the effect of the main explanatory variable', in our case, female autonomy? In most multiple regressions, the coefficient of the main explanatory variable declines as more (observable) control variables are introduced. Hence, the Altonji-Elder-Taber ratio (AET) compares the size of the coefficient of interest (female autonomy) in a restricted regression including only a constant (and, in our case, the fixed effects) $ß_{restr}$ to the coefficient of a regression with a variety of controls ($ß_{full}$).[10] $ß_{restr}$-$ß_{full}$ is the denominator of the AET-ratio and $ß_{full}$ is the numerator, because the larger it is, the stronger is the effect of the variable of interest (female autonomy).[11] If control variables only remove a small part of the female autonomy coefficient, then unobservables would need to have a very strong effect to completely eliminate the impact of female autonomy – under the assumption of roughly proportional selection on observables and on unobservables. We assess for two models, once including land inequality and once not including it. We observe that omitted variables would need to be between 2 and 11 times larger than the observed variables to remove the effect of female autonomy.

Oster (2017) refined this method by taking account of the explanatory share of the two regressions as well. Hence, we use her method to report the relative degree of selection on unobservables such that the effect of age at marriage is totally eliminated, while taking into account the R-squared's movements as well. As rule of thumb, an Oster test above 1 suggests that the observables are "sufficiently relevant" (see table 12). This is the case for both models. In all specifications, country fixed effects are included as controls. We thus conclude that the AET ratio and Oster tests indicate a low probability of identification issues caused by omitted variable bias.

Spatial autocorrelation may be a concern, which we address in the following. Similar to temporal autocorrelation, in which the previous period might have an impact on the current

---

[10] The regression equation for $ß_{full}$ is as follows: $abcc_{ih} = \alpha + ß_{full*}\ agemarr_{ic} + Y_c + S_j + C_i + ß_{control} - X_{ih} + \varepsilon_{ih}$, where same definitions apply as for equation (1).

[11] The regression equation for $ß_{restr}$ is as follows: $abcc_{ih} = \alpha + ß_{restr*}\ agemarr_{ic} + Y_c + S_j + C_i + \varepsilon_{it}$

behaviour of a variable (independent of the explanatory variables), the behaviour in a region might be influenced by the behaviour in one of the adjacent regions (again independent of explanatory variables). An econometric technique to take this into account is to calculate Conley standard errors, a standard procedure for cross-sectional data (Conley 2008). For panel data, in which temporal autocorrelation might also play a role, Hsiang (2010) developed a specific method (see notes to Table 12, for details). We applied this method using plausible bandwidths for distances of 250, 500, or 750 kilometres. We find that spatial autocorrelation does not make our general results invalid, as all three bandwidths result in standard errors for age at marriage that imply statistical significance at the 5% level and the results are remarkably similar for regions of the 19th Century and the panel of countries for 1500-1850.

6 Conclusion

Our empirical results suggest that economies with more female autonomy became (or remained) superstars in numeracy development. The female part of the population needed to contribute to overall human capital formation and prosperity, otherwise the competition with other economies was lost. Institutions that excluded women from developing human capital – such as being married early, and hence, often dropping out of the independent, skill demanding economic activities – prevented many economies from being successful in human history. We have shown this for the long-term development of European economies. We find that the indicator age at marriage is a reasonable proxy for female autonomy in European development. It predicts numeracy formation in a variety of regressions and graphical analyses. Typically, one year of age at marriage corresponded with 4 numeracy units. Is this economically relevant? The distance between Russia and the Netherlands represents about 8 years of age at marriage. Multiplied with a coefficient of 4, this would correspond with 32 numeracy units – a substantial value, given that the difference between high numeracy values in Europe around 1900 and low values in South Asia and Africa were only about 40%

(Tollnek and Baten 2017). In other words, the numeracy differential between areas of high and low values of age at marriage in the early modern period equals 80% of the difference between the less developed world and Europe in 1900.

We gain exogenous variation of female autonomy by studying the genetic factor of lactose tolerance. This factor increased the possibility and demands to perform dairy farming. Dairy farming is an important agricultural activity allowing women to participate in income generation, since it was less demanding concerning upper-body strength than grain production. Moreover, it demanded specific skills, such as caring for the health of the cattle, and guaranteeing a minimum value of hygiene to the animals, which was culturally often associated with female attitudes. Hence, this genetic factor influences long-term differences in agricultural and overall economic specialization, implying a set of gender-specific institutions. A second identification strategy for a regional sample is based on soil and climatic suitability for pasture.

In the instrumental variable regressions, we find that the impact of female autonomy on human capital formation is probably causal. We also intensively discuss the issue of the exclusion restriction, since one could imagine that nutrition would be an omitted factor here, which it turns out not to be.

The larger regional sample also allows to access identification issues by using AET and Oster ratio tests. In this assessment, we arrive at the results that identification issues based on omitted variables would probably not have such a large role as to eliminate the whole female autonomy effect on numeracy formation.

In sum, we argue that the female autonomy factor is a crucial one. It plays a particularly important role for the development of numerical skills, which are the ones that matter most for economic growth (Hanushek and Woessmann 2012, 2020). The superstars of numeracy development were the ones with high gender equality.

References

A'Hearn, B., Crayen, D., and Baten, J., 2009. Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital. Journal of Economic History 68(3), 783-808.

Acemoglu, D., Cantoni, D. Johnson, and S. Robinson, J.A. 2011. The Consequences of Radical Reform: The French Revolution. *American Economic Review 101 (December 2011): 3286–3307*

Acemoglu D., and Robinson, J.A., 2012. Why Nations Fail: The Origins of Power, Prosperity and Poverty. New York: Crown Publishing Group.

Acemoglu, D., Johnson, S., and Robinson, J.A., 2005. The Rise of Europe: Atlantic Trade, Institutional Change and Growth. American Economic Review 95(3), 546-79.

Alesina, A., Giuliano, P. and Nunn, N., 2013. On the Origins of Gender Roles: Women and the Plough. Quarterly Journal of Economics 128(2), 469-530.

Allen, R.C., 2003. Progress and Poverty in Early Modern Europe. Economic History Review 56(3), 403-43.

Altonji, J.G., Elder, T.E., and Taber, C.R., 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. Journal of Political Economy 113(1), 151-84.

Arellano, M., and Bond, S., 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. The Review of Economic Studies 58(2), 277-97.

Baten, J., 2009. Protein Supply and Nutritional Status in Nineteenth Century Bavaria, Prussia and France, Economics and Human Biology 7(2), 165-180.

Baten, J., 2013. Were Women Accused of Witchcraft the Wise Men of the Middle Ages and Early Modern Period? Conference Contribution to the Human Capital and History Workshop Tübingen.

Baten, J. 2016. A History of the Global Economy. Cambridge: CUP.

Baten, J., and Blum, M., 2014a. Human height since 1820. In: van Zanden, J.L. (eds.), How was Life? Global Well-being since 1820. Paris: OECD Publishing, 117-37.

Baten, J., and Blum, M., 2014b. Why are You Tall while Others are Short? Agricultural Production and Other Proximate Determinants of Global Heights. European Review of Economic History 18(2), 144-65.

Baten, J., Crayen, D., and Voth, H-J., 2014. Numeracy and the Impact of High Food Prices in Industrializing Britain, 1780-1850. Review of Economics and Statistics 96(3), 418-30.

Baten, J. and Hippe, R., 2018. Geography, Land Inequality and Regional Numeracy in Europe in Historical Perspective, Journal of Economic Growth 23-1 (2018), 79-109.

Baten, J., and Juif, D., 2013. A Story of Large Land -Owners and Math Skills: Inequality and Human Capital Formation in Long-Run Development, 1820-2000. Journal of Comparative Economics 42(2), 375-401.

Baten, J., and Mumme, C., 2013. Does inequality Lead to Civil Wars? A Global Long-Term Study Using Anthropometric Indicators (1816–1999). European Journal of Political Economy 32, 56-79.

Baten, J., Szoltysek, M., and Camestrini, M., 2017. “Girl Power” in Eastern Europe? The Human Capital Development of Central-Eastern and Eastern Europe in the Seventeenth to Nineteenth Century and its Determinants. European Review of Economic History 21(1), 29-63.

Baten, J., and van Zanden, J.L., 2008. Book Production and the Onset of Modern Economic Growth. Journal of Economic Growth 13(3), 217-35.

Becker, S., and Woessmann, L., 2009. Was Weber Wrong? A Human Capital Theory of Protestant Economic History. Quarterly Journal of Economics 124(2), 531-96.

Bils, M., & Klenow, P. J. (2000). Does schooling cause growth? The American Economic Review, 90(5), 1160-1183.

Bleakley, H., Ferrie, J., 2016. Shocking behavior: Random wealth in Antebellum Georgia and human capital across generations. Quarterly Journal of Economics (2016), 1455–1495

Boehm, Max (1995). Bayerns Agrarproduktion 1800-1870. St. Katharinen.

Boserup E. (1970). *Women's Role in Economic Development*. London: George Allen & Unwin.

Bolt, J., and van Zanden, J.L., 2014. The Maddison Project: Collaborative Research on Historical National Accounts. Economic History Review 67(3), 627-51.

Botão Rego, A., Amorim, M.N., Ferreira, A., Santos, C., Silva, M.N., Simas Guerreiro, M.J., 2016. Variation in age at marriage in Portugal from the seventeenth century to the early twentieth century. Revista de Demografía Histórica 34-2, 129-159

Broadberry, S.N., Campbell, B., Klein, A., Overton, M. and van Leeuwen, B. 2015. British Economic Growth, 1270-1870. Cambridge: Cambridge University Press.

Boucekkine, R., de la Croix, D. and Peeters, D. (2007). “Early literacy achievements, population density, and the transition to modern growth”, *Journal of the European Economic Association* 5, 183-226.

Carmichael S. (2011), “Marriage and power: Age at first marriage and spousal age gap in lesser developed countries”, History of the Family, Vol. 16, pp. 416-436.

Carmichael, S., de Pleijt, A.M., van Zanden, J.L., and de Moor, T., 2016. The European Marriage Pattern and its Measurement. Journal of Economic History 76(1), 196-204.

Carruthers, B. and Espeland, W.N., 1991. Accounting for Rationality: Double-Entry Book-Keeping and the Rhetoric of Economic Rationality. American Journal of Sociology. 97, no. 1 (1991): 31–96.

Cook, C.J., 2014. The role of lactase persistence in precolonial development, Journal of Economic Growth (2014) 19:369–406.

Crayen, D., and Baten, J., 2010. Global Trends in Numeracy 1820-1949 and Its Implications for Long-Term Growth. Explorations in Economic History 47(1), 82-99.

Deaton, A., 1989, Looking for Boy-Girl Discrimination in Household Expenditure Data, World Bank Economic Review 3(1). 1-15.

Deaton, A, 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. Washington, DC: The World Bank.

De Moor, T., and van Zanden, J.L. 2010. Girl Power: The European Marriage Pattern and Labour Markets in the North Sea Region in the Late Medieval and Early Modern Period. Economic History Review 63(1), 1-33.

Dennison, T., and Ogilvie, S., 2013. Does the European Marriage Pattern Explain Economic Growth. CESifo Group Munich Working Paper, no. 4244.

Dennison, T., and Ogilvie, S., 2014. Does the European marriage pattern explain economic growth? The Journal of Economic History 74(3), 651-93.

Dennison, T., and Ogilvie, S., 2016. Institutions, Demography and Economic Growth. The Journal of Economic History 76(1), 205-17.

De Pleijt, A.M. 2018. Human Capital Formation in the Long Run: Evidence from Average Years of Schooling in England, 1300-1900. Cliometrica 12(1), 99-126.

De Pleijt, A.M., and van Zanden, J.L., 2016. Accounting for the “Little Divergence”. What Drove Economic Growth in Preindustrial Europe, 1300-1800? European Review of Economic History 20(4), 387-409.

De Pleijt, A.M., Nuvolari, A., and Weisdorf, J.L., 2018. Human Capital Formation during the First Industrial Revolution: Evidence from the Use of Steam Engines. Journal of the European Economic Association, forthcoming.

De Pleijt, A.M. and Weisdorf, J.L., 2017. Human Capital Formation from Occupations: The ‘Deskilling Hypothesis’ Revisited. Cliometrica 11(1), 1-30.

De Pleijt, A.M., van Zanden, J.L., and Carmichael, S.G., 2016b. Gender Relations and Economic Development: Hypotheses about the Reversal of Fortune in EurAsia. Centre for Global Economic History (CGEH) Working Paper Series, no. 79

Diebolt, C. and Perrin, F., 2013. From Stagnation to Sustained Growth: The Role of Female Empowerment. American Economic Review: Papers and Proceedings 103, 545-49.

Duflo, E., 2012. Women Empowerment and economic Development. Journal of Economic Literature, 50(4). 1051-1079.

Durlauf, S.N., Johnson, P.A. and Temple, J.R.W. 2005: Growth Econometrics, in Aghion, P., Durlauf, S.N., Handbook of Economic Growth, vol 1, 555-677.

Flatz, G., 1995: The Genetic Polymophism of Intestinal Lactase Activity in Adult Humans; in: Scriver, C.R. et al. (eds.): The Metabolic and Molecular Bases of Inherited Desease, Seventh Edition, Mc Graw-Hill, New York.

Galor, O. and Özak, O., 2016 "The Agricultural Origins of Time Preference", *American Economic Review* 106 (10) , 3064-3103.

Gennaioli, N. and Voth, H.J., 2015. State Capacity and Military Conflict. The Review of Economic Studies 4-1, 1409-1448.

Goldthwaite, R.A., 1972. Schools and Teachers of Commercial Arithmetic in Renaissance Florence. Journal of European Economic History 1, 418- 33.

Graff, H.J., 1987. The Legacies of Literacy: Continuities and Contradictions in Western Culture and Society. Indiana University Press, Bloomington.

Gruber, S. and Szoltysek, M., 2014. Th Patriarchy Index: A Comparative Study of Power Relations across Historic Europe. MPIDR WORKING PAPER WP 2014-007

Gruen, C. and S. Klasen, 2008. Growth, inequality, and welfare: Comparisons across space and time. Oxford Economic Papers 60: 212-236.

Hanushek, E.A., and Woessmann, L., 2012. Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation. Journal of Economic Growth 17(4), 267-321.

Hanushek, E.A. and Woessmann, L. 2020 Education, knowledge capital, and economic growth, in Bradley, S., Green, C. (eds.), The Economics of Education, Second Edition, Amsterdam: Elsevier https://doi.org/10.1016/B978-0-12-815391-8.00014-8

Ingram, C.J.E., Mulcare, C.A., Itan, Y., Thomas, M.G., Swallow, D.M., 2009. Lactose Digestion and the Evolutionary Genetics of Lactose Persistence. Human Genetics 124, 579–91

Long, J. B. de, and Shleifer, A., 1993. Princes and Merchants: European City Growth Before the Industrial Revolution. Journal of Law and Economics 30(2), 671-702.

Kelly, M., Mokyr, J. and Ó Gráda, C., 2013. Precocious Albion: A New Interpretation of the British Industrial Revolution. UCD Centre for Economic Research Working Paper Series, no 13/11.

Keywood, T. and Baten, J., 2020. Elite Violence and Elite Numeracy in Europe from 500 to 1900 CE: Roots of the Divergence", Cliometrica 15, pp. 319–389.

Khanna, R., Kumar, A., Vaghela, J. F., and Puliyel, J. M., 2003. Community Based Retrospective Study of Sex in Infant Mortality in India," British Journal of Medicine 327. 126–129

Klasen, S. and Lamanna, F., 2009. The Impact of Gender Inequality in education and employment on economic growth: new evidence for a panel of countries. Feminist Economics 15(3),91–132.

Komlos, John (1989). Nutrition and Economic Development in the Eighteenth-Century Habsburg Monarchy: An Anthropometric History. Princeton.

Krishnan P., 1977. Age at marriage in a nineteenth century Indian parish. In: Annales de démographie historique 1977, pp. 271-284;

Kula, W., 1976. An Economic Theory of the Feudal System: Towards a Model of the Polish Economy, 1500-1800. London: N.L.B.

Lampe, M. and Sharp, P., 2015. Just add milk: a productivity analysis of the revolutionary changes in nineteenth-century Danish dairying. The Economic History Review 68 (4), 1132-1153.

Millward, R., 1982. An Economic Analysis of the Organization of Serfdom in Eastern Europe. Journal of Economic History 42, 513-48.

North, D.C., and Weingast B.R., 1989. Constitutions and Commitment: Evolution of Institutions Governing Public Choice in Seventeenth-Century England. Journal of Economic History 49(4), 803-32.

Nunn, N., and Wantchekon, L., 2011. The Slave Trade and the Origins of Mistrust in Africa. American Economic Review 101(7), 3221-52.

Oster, E., 2017. Unobservable Selection and Coefficient Stability: Theory and Evidence, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2016.1227711

Perrin F, and M. Benaim (2018), "Regional Patterns of Economic Development: France during the Industrialization", in: C. Diebolt, S. Carmichael, S. Dilli, A. Rijpma, C. Störmer (Eds.): Cliometrics of the Family: Global Patterns and Their Impact on Divergin Development, Studies in Economic History, Editions Springer.

Plötz, J., 2013. Human Capital Trends in Central America and the Caribbean from the 18th to the 20th Century. Bachelor Thesis. University of Tuebingen.

Reis, J., 2005. Economic Growth, Human Capital Formation and Consumption in Western Europe before 1800., in Allen, R.C., Bengtsson, T., Dribe, M., (eds). Living Standards in the Past: New Perspectives on Well-Being in Asia and Europe. Oxford University Press, Oxford, 195-225.

Rose, E. 1999. Consumption Smoothing and Excess Female Mortality in Rural India, Review of Economics and Statistics 81(1). 41-49.

Rosenstock, E., Groß, M., Hujić, A. and Scheibner, A., 2015. Back to Good Shape: Biological Standard of Living in the Copper and Bronze Ages and the Possible Role of Food., in Kneisel, J., Dal Corso, M., Kirleis, W., Scholz, H., Taylor N., and Tiedtke, V. (eds.). The third food revolution? Setting the bronze age table: common trends in economic and subsistence strategies in Bronze Age Europe. Habelt, Bonn, pp. 121-52.

Sandberg, L.G., 1979. The Case of the Impoverished Sophisticate: Human Capital and Swedish Economic Growth before World War I. The Journal of Economic History 39(1), 225-41.

Schumpeter, J., 1950. Capitalism, Socialism and Democracy. Harper Torchbooks, New York.

Sen, A., 1990. More than 100 Million Women Are Missing. New York Review of books. 37 (20).

Tollnek, F., and Baten, J. 2017. Farmers at the Heart of the "Human Capital Revolution"? Decomposing the Numeracy Increase in Early Modern Europe, *Economic History Review 70-3*, 779-809.

United Nations, 2005. *Progress towards the Millennium Development Goals, 1990–2005*. Secretary-General's Millennium Development Goals Report. New York and Geneva: United Nations Publications.

Van Zanden, J.L., and van Leeuwen, B., 2012. Persistent but not Consistent: The Growth of National Income in Holland, 1347-1807. Explorations in Economic History 49(2), 119-130.

Van Zanden, J.L., Buringh, E., and Bosker, M., 2012. The Rise and Decline of European Parliaments, 1188-1789. Economic History Review 65(3), 835-61.

Voigtländer, N., and Voth, H-J., 2013. How the West 'Invented' Fertility Restriction. American Economic Review 103(6), 2227-64.

Weber, M. (1930/1976). The Protestant Ethic and the Spirit of Capitalism. Scribner, New York.

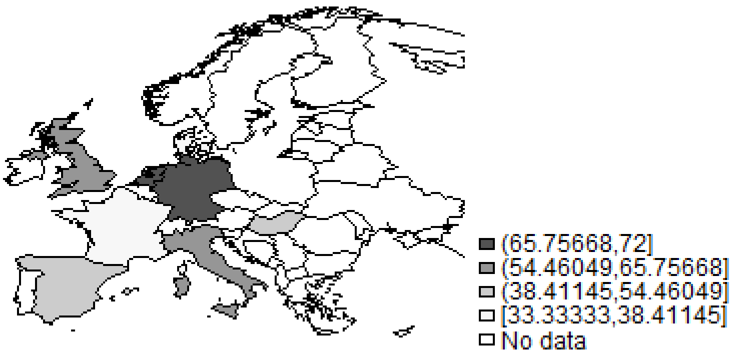


Figure 1. ABCC Index in Europe, 1500-49
Notes and sources: See text. Dark countries: high numeracy; light grey: low numeracy; white countries: no data.

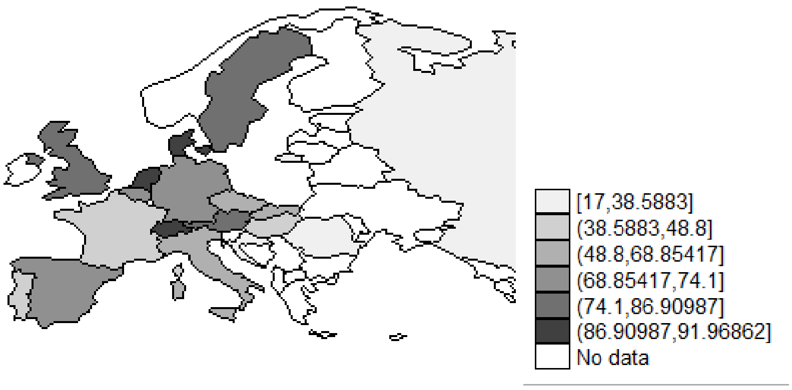


Figure 2. ABCC Index in Europe, 1600-49
Notes and sources: See text. Dark countries: high numeracy; light grey: low numeracy; white countries: no data.

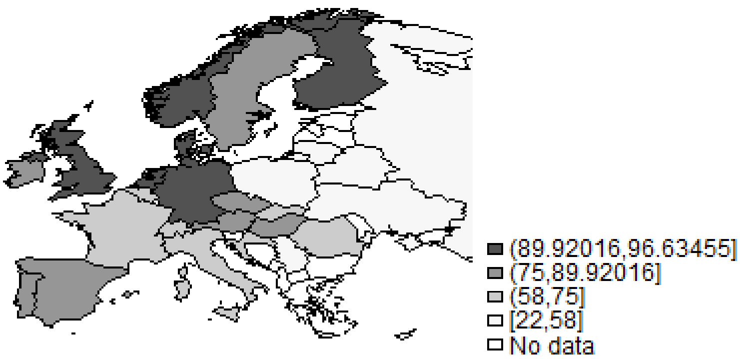


Figure 3. ABCC Index in Europe, 1700-49
Notes and sources: See text. Dark countries: high numeracy; light grey: low numeracy; white countries: no data.

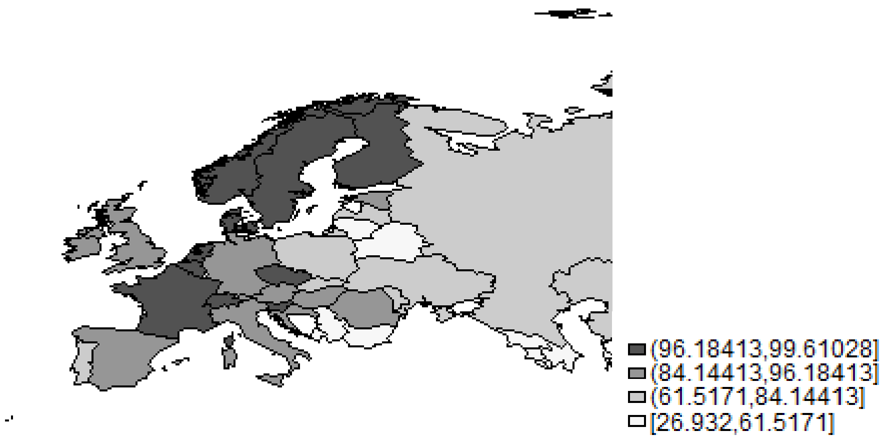


Figure 4. ABCC Index in Europe, 1800-49
Notes and sources: See text. Dark countries: high numeracy; light grey: low numeracy; white countries: no data.

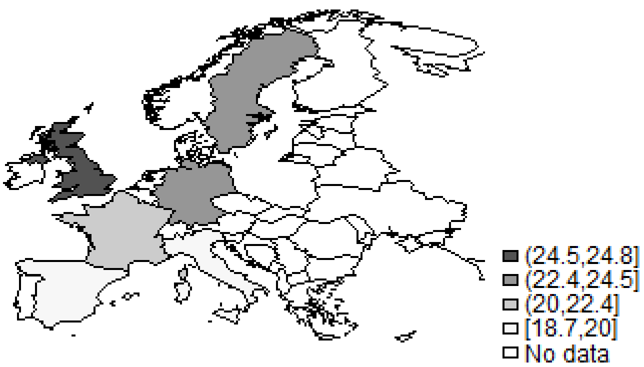


Figure 5. Average age at marriage in Europe, 1500-49
Notes and sources: See text. Dark countries: high average age at marriage; light grey: low average age at marriage; white countries: no data.

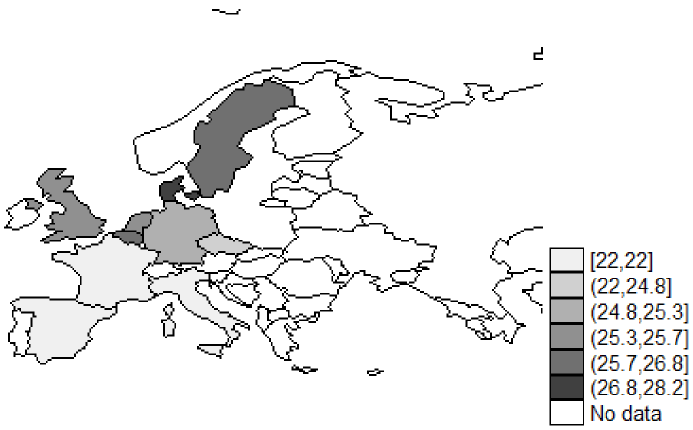


Figure 6. Average age at marriage in Europe, 1600-49
Notes and sources: See text. Dark countries: high average age at marriage; light grey: low average age at marriage; white countries: no data.

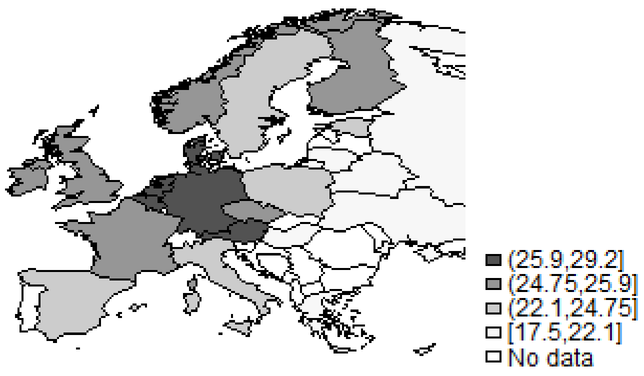


Figure 7. Average age at marriage in Europe, 1700-49
Notes and sources: See text. Dark countries: high average age at marriage; light grey: low average age at marriage; white countries: no data.

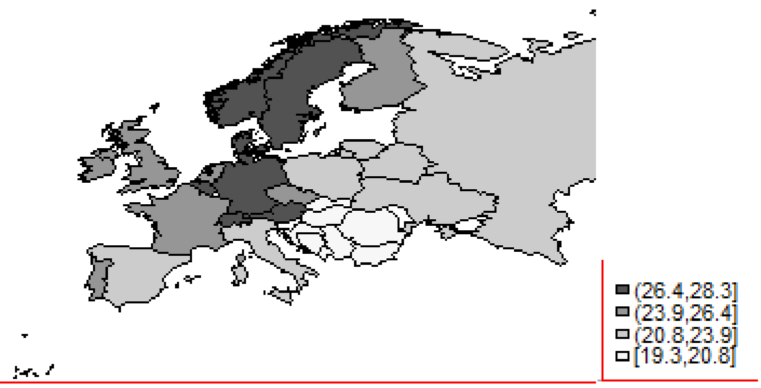
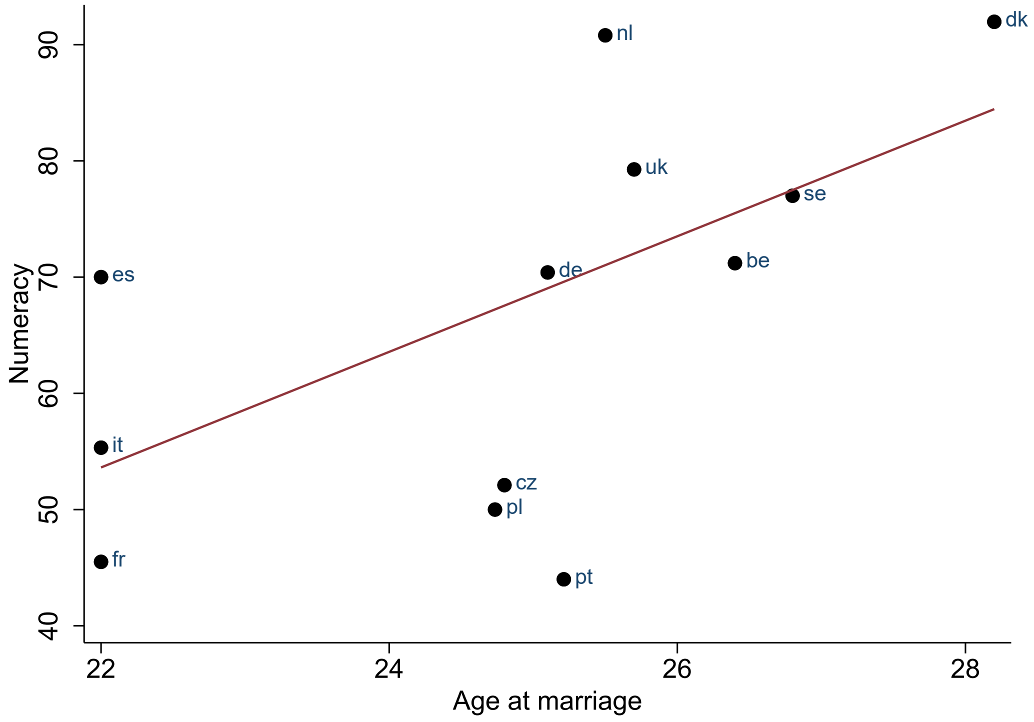
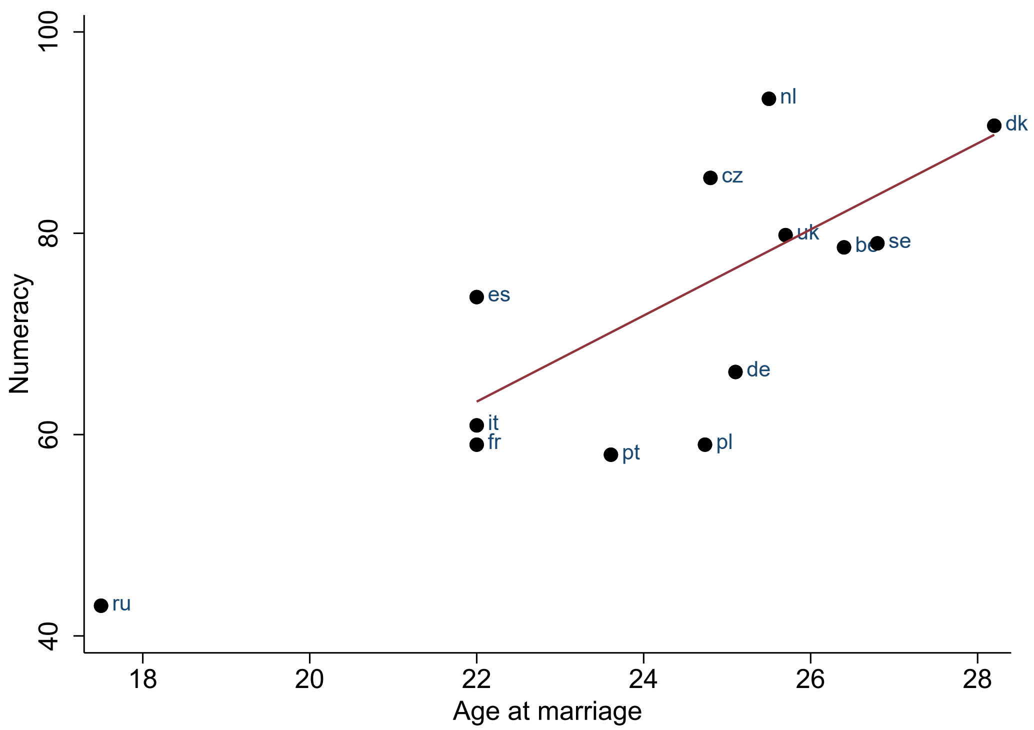


Figure 8. Average age at marriage in Europe, 1800-49
Notes and sources: See text. Dark countries: high average age at marriage; light grey: low average age at marriage; white countries: no data.

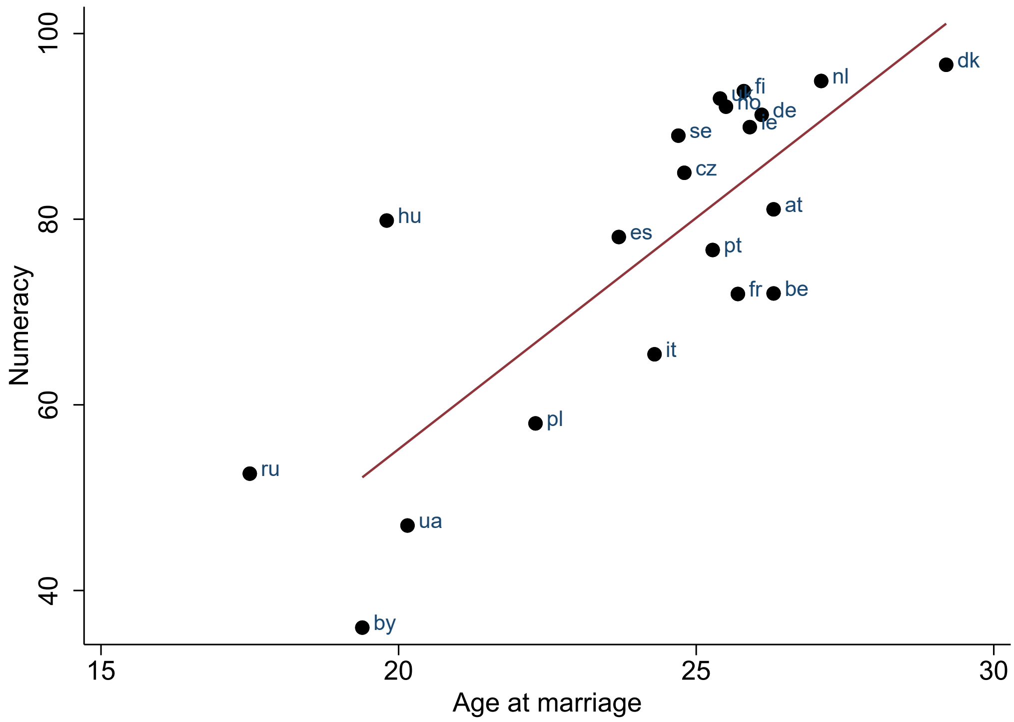
Panel A: 1600-49



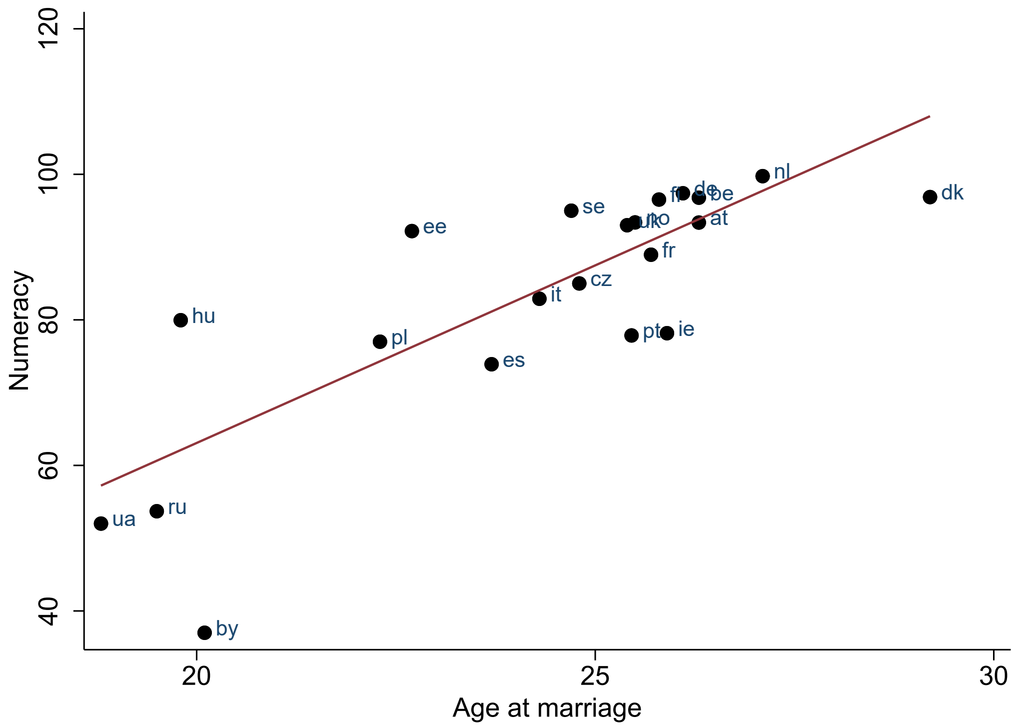
Panel B: 1650-99



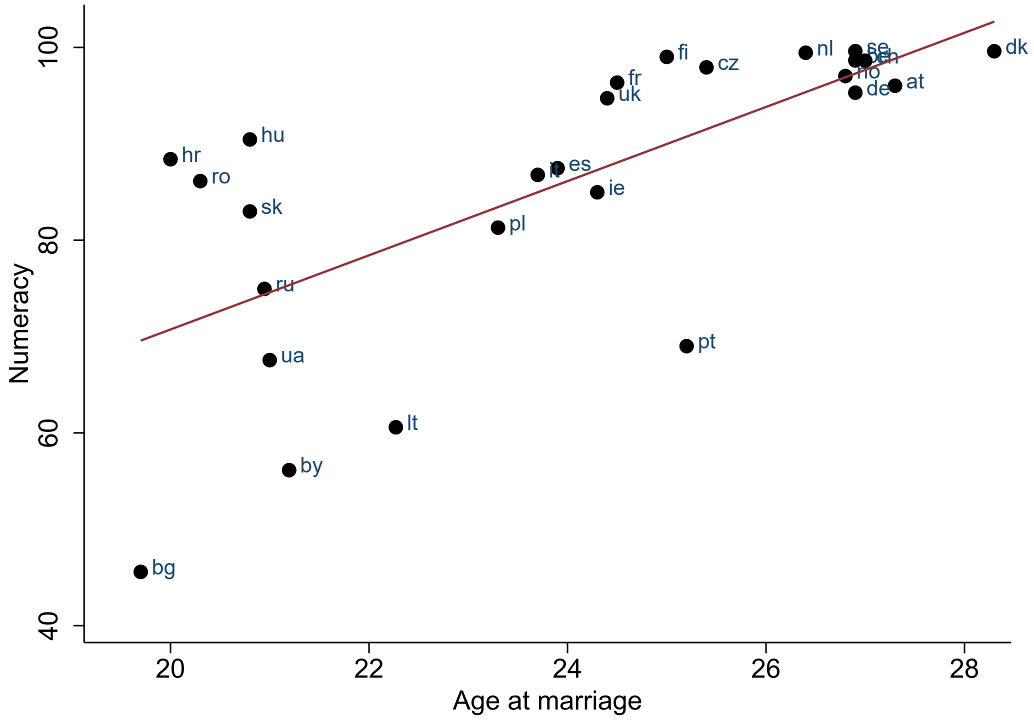
Panel C: 1700-49



Panel D: 1750-99



Panel E: 1800-1849



Country codes:

at	Austria	fr	France	pt	Portugal
by	Belarus	de	Germany	ro	Romania
be	Belgium	hu	Hungary	ru	Russia
bg	Bulgaria	ie	Ireland	sk	Slovakia
hr	Croatia	it	Italy	es	Spain
cz	Czech Republic	lt	Lithuania	se	Sweden
dk	Denmark	nl	Netherlands	ch	Switzerland
ee	Estonia	no	Norway	ua	Ukraine
fi	Finland	pl	Poland	uk	United Kingdom

Figure 9. Average age at marriage and numeracy in 1700

Notes: The figure shows the relationship between average age at marriage and numeracy in the half centuries by country. *Sources*: See text.

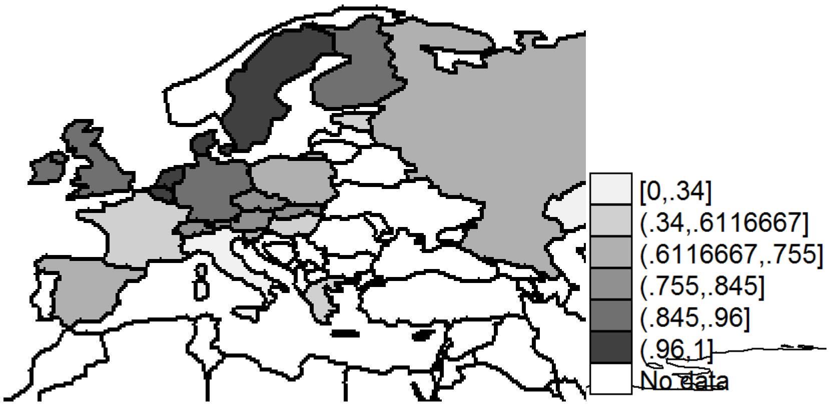


Figure 10. Lactose tolerance in Europe (dark: high; white: missing value)
Notes and sources: Baten and Blum (2014a, 2014b). Dark countries: high lactose tolerance; light grey: low lactose tolerance; white countries: no data.

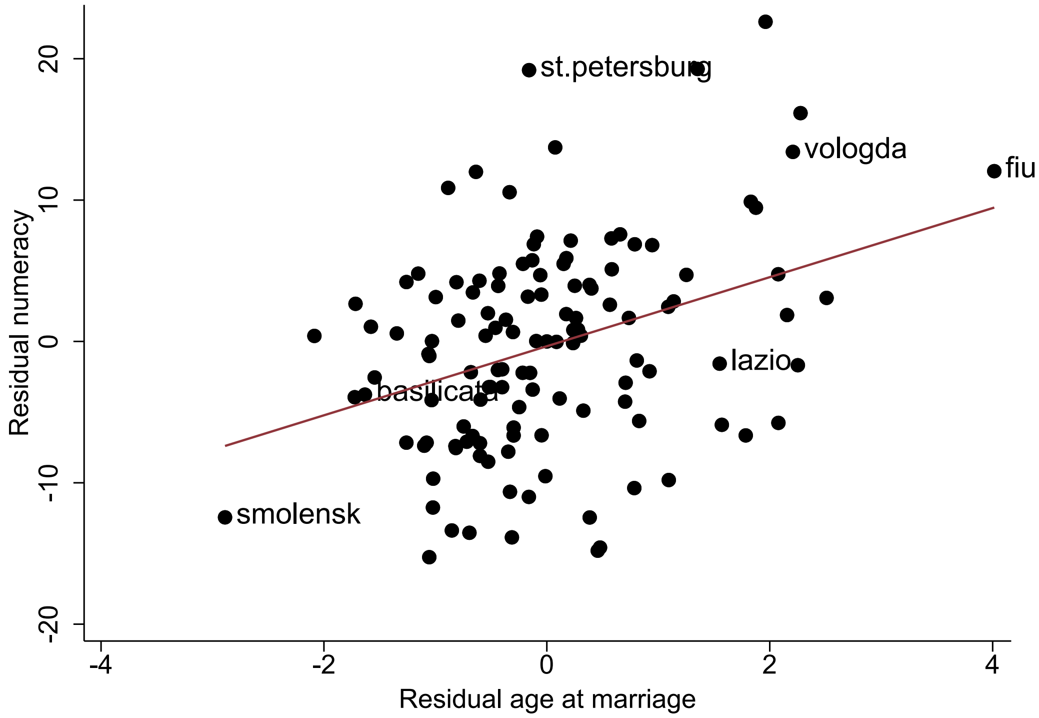


Figure 11. Partial scatter-plot of the regions
Notes and sources: see text.

Variable	N	Mean	Standard deviation	Minimum	Maximum
Numeracy	88	78.79	17.54	35.16	99.75
Age at marriage	88	24.29	2.56	18.70	29.20
War	78	0.29	0.31	0	1
Civil war	78	0.07	0.12	0	0.48
Serfdom	88	0.11	0.32	0	1
Log population density	88	-3.49	1.70	-6.45	3.19
Atlantic trade	70	0.59	0.86	0	2.05
Size fleet	66	6.07	4.89	0	12.25
Prince	46	0.72	0.46	0	1
Parliaments	60	2.82	1.47	0	4.62
Lactose tolerance	88	0.75	0.15	0.43	0.96
Height	46	165.33	2.12	161.66	170.48

Table 1. Descriptives of panel data set

Sources: see text. *Notes*: The table shows the observations for which there is evidence on numeracy, age at marriage and lactose tolerance at the same time.

	Female hhh	Married 15-19	Older wives	Female 20-34 non-kin
Female hhh	1			
Married 15-19	-0.40**	1		
Older wives	0.52**	-0.73**	1	
Female 20-34 non-kin	0.35**	-0.76**	0.69**	1

Table 2. Correlation analysis of female autonomy indicators by Gruber and Szołtysek (2014).

Notes: These correlations are by European place and period, based on an underlying sample of 700,000 Europeans. Coverage: mostly 18th-early 20th century. *Source*: Gruber and Szołtysek (2014). *Definitions*: 'Female hhh': proportion of all female household heads among all adults (20+ years); 'Married 15-19': proportion of ever-married women in the age group 15-19 years (this is almost perfectly negatively correlated with age at marriage, corr=0.91, p=0.00); 'Older wives': Proportion of all the wives who are older than their husbands; 'Female 20-34 non-kin': proportion of women aged 20-34 years who live as non-kin, usually as lodgers or servants (i.e. outside the home/control of her husband or her husband's relatives).

Country	1500	1600	1650	1700	1750	1800
Austria				81	93	96
Belarus				36	37	56
Belgium		71	79	72	97	99
Bulgaria						46
Croatia						88
Czech Republic		52	86	85	85	98
Denmark		92	91	97	97	100
Estonia					92	
Finland				94	97	99
France	35	46	59	72	89	96
Germany	68	70	66	91	97	95
Hungary				80	80	90
Ireland				90	78	85
Italy	64	55	61	65	83	87
Lithuania						61
Netherlands		91	93	95	100	99
Norway				92	93	97
Poland				58	77	81
Portugal		44	58	77	78	69
Romania						86
Russia			43	53	54	75
Slovakia						83
Spain	49	70	74	78	74	87
Sweden		77	79	89	95	100
Switzerland						99
Ukraine				47	52	68
United Kingdom	60	79	80	93	93	95

Table 3. Numeracy estimates for European countries at selected points in time
Sources: see text. *Notes*: “1500” refers to 1500-1549, “1600” to 1600-49, etc.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Incl. cases	All	All	All	All	Lact.	All	All	All	All	All
Age at marriage	4.389***	6.319***	4.332***	3.347***	3.140**	3.273**	3.395**	3.372***	3.600**	3.408**
	(0.583)	(1.419)	(0.566)	(1.030)	(1.341)	(1.258)	(1.388)	(1.029)	(1.595)	(1.487)
p-value Wild	0.0000	0.0000	0.0000	0.0110	0.0521	0.0541	0.0561	0.0090	0.0780	0.0791
War						-0.284				-0.0410
						(6.104)				(6.465)
Civil war							-2.760			-2.694
							(9.372)			(10.15)
Serfdom								-0.493		-0.338
								(2.232)		(2.401)
GDP (ln.)									8.840	
									(8.861)	
Time FE	N	N	Y	Y	Y	Y	Y	Y	Y	Y
Country FE	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Source FE	N	N	N	Y	Y	Y	Y	Y	Y	Y
Observations	93	93	93	93	84	87	87	93	73	87
R-sq.	0.461	0.761	0.624	0.897	0.886	0.905	0.905	0.897	0.880	0.905

Table 4. Least Square Dummy variable (LSDV) regressions of numeracy

Notes: *** indicates significance at the 1% level; ** at the 5% level; and * at the 10% level. The number in brackets corresponds to the robust standard errors, clustered at the country-century level, 63 clusters in column 1. *Sources*: see text. Column (1) shows the results of the OLS regression without restrictions or additional explanatory variables. In column (2), country fixed effects are introduced. In columns (3) and (4) we introduce time and source fixed effects. Column (5) restricts the analysis to the countries for which we have information on migration adjusted lactose tolerance. In columns (6) to (9) we add control variables for war, civil war, serfdom, and national income to perform robustness checks. Since GDP per capita is not given for all countries, the sample in column (9) is smaller. In column (10), the three control variables war, civil war and serfdom are combined. We calculated Wild bootstrapped p-values that we observe to be 0.000, indicating clearly statistical significance (Cameron and Miller 2015). The variable age at marriage is statistically significant in all specifications.

	(1)	(2)
Second Stage		
Age at marriage	6.260***	6.544***
	(0.889)	(1.005)
Wild test p-value	0.0000	0.0000
Controls	No	Yes
Time FE	Yes	Yes
Source FE	Yes	Yes
Region FE	Yes	Yes
First Stage		
Lactose tolerance	9.538***	8.783***
	(1.867)	-1.73
F-Stat	26.109	25.764
Montiel-OP-crit.value 10%	23.109	23.109
Observations	84	78
R-squared	0.618	0.626
Endogeneity	0.0230	0.0541

Table 5. 2SLS regressions measuring the effect of average age at marriage on numeracy

Notes: *** indicates significance at the 1% level; ** at the 5% level; and * at the 10% level. The number in brackets corresponds to the robust standard errors, clustered at the country-century level, 57 clusters. We calculated Wild bootstrapped p-values that we observe to be 0.000, indicating clearly statistical significance (Cameron and Miller 2015). As the intrument lactose tolerance is time-invariant, we cannot control for country FE, and use region FE instead. We use lactose tolerance as the instrument. The lower part of this table shows the first stage of the 2SLS estimation. The upper part of this table shows the second stage, including the controls explained in Table 4. In both models, we include time effects (using century dummy variables), source fixed effects (using source-specific dummy variables), and regional fixed effects (using world region dummy variables). In model 2, we include additionally all the control variables of Table 4, Col. 10: war, civil war and serfdom. We also executed Montiel Olea Pflueger tests for weak instruments that resulted in a critical value of 23.11 for our number of clusters at the 10% level. The Montiel Olea Pflueger test is a routine that is robust in situations of heteroscedasticity and autocorrelation, and it provides a cluster–robust weak-instrument test (Pflueger and Wang 2015). *Sources*: see text.

	(1)	(2)	(3)	(4)	(5)	(6)
Second Stage						
Age at marriage	5.217***	4.507***	4.959***	6.509***	5.737***	5.217***
	(1.247)	(0.737)	(0.858)	(1.448)	(1.006)	(1.247)
Height	1.510					
	(0.939)					
Atlantic trade (ln)		3.125***				
		(1.204)				
Size of merchant fleet (ln)			0.0936			
			(0.242)			
Prince				0.232		
				(3.514)		
Parliamentary activity (ln)					2.525**	
					(1.000)	
Pop. Density (ln)						-0.722
						(1.348)
Controls included	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y	Y
Source FE	Y	Y	Y	Y	Y	Y
Constant	-268.6*	-31.16***	-27.02***	-28.33***	-26.18***	-268.56**
	(149.1)	(7.401)	(6.964)	(7.960)	(7.510)	(149.12)
First Stage						
Lactose tol.	9.402***	9.871***	9.373***	8.932***	6.830***	7.843***
	(2.506)	(1.651)	(1.325)	(1.700)	(1.315)	(1.331)
Constant	47.83**	14.39***	12.06***	12.71***	14.38***	13.88***
	(20.91)	(1.140)	(1.076)	(1.281)	(1.201)	(1.085)
Observations	78	64	60	40	54	40
R-squared	0.791	0.792	0.809	0.928	0.918	0.802
F-stat	14.08	50.05	27.62	26.98	34.75	35.74

Table 6. 2SLS regressions including potentially endogenous controls

Notes: *** indicates significance at the 1% level; ** at the 5% level; and * at the 10% level. The number in brackets corresponds to the robust standard errors, clustered at the country-century level. As the intrument lactose tolerance is time-invariant, we cannot control for country FE, and use region FE instead. *Sources*: see text. In addition to the specifications explained in Table 5, we use height in column (1), and the log of the volume of Atlantic trade of Acemoglu et al. (2005) as control variable in column (2); we add the log of the (per capita) tonnage size of the merchant fleet in column (3). We add the "Prince" variable of the Long and Shleifer (1993) column (4) and use the log of the activity index of European Parliaments of van Zanden et al. (2012) in column (5) to test for the effect of political institutions; finally, log population density in Column (6), as this reflects the demographic behaviour to a certain extent.

Excluded:	(1) UK	(2) NL	(3) RU	(4) Death	(5) Witches	(6) Both
Second Stage						
Age at marriage	6.066***	6.575***	6.544***	6.697***	6.544***	6.697***
	(0.902)	(1.022)	(1.005)	(1.283)	(1.005)	(1.283)
Controls incl.	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
Source FE	Y	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y	Y
First Stage						
Lactose tolerance	9.668***	8.671***	8.783***	7.498***	8.783***	7.498***
	(1.892)	(1.733)	(1.730)	(1.757)	(1.719)	(1.744)
F-Stat	26.11	25.05	25.7657	18.22	26.095	18.49
Observations	72	73	78	67	77	66
R-squared	0.645	0.611	0.626	0.654	0.593	0.618

Table 7. Robustness-checks of the baseline regressions.

Notes: *** indicates significance at the 1% level; ** at the 5% level; and * at the 10% level. The number in brackets corresponds to robust standard errors, clustered at the country-century level (53 clusters in model 3). As the instrument lactose tolerance is time-invariant, we cannot control for country FE, and use region FE instead. *Sources*: see text. In the table above we excluded several groups in order to make sure the inclusion of these in the baseline results did not divert the results. In columns (1) and (2) we excluded the data from the UK and Netherlands. In column (3) we excluded Russia because it was much poorer than the other countries in the sample. We have re-estimated the regressions by omitting death records in column (4), witch trials in column (5), and either of them in column (6).

	N	Mean	Standard Deviation	Min.	Max
Numeracy	184	76.92	17.07	21.38	100.51
Age at marriage	184	21.67	2.11	18.00	27.27
Land ineq	184	0.40	0.19	0.00	0.74
Protestantism	135	13.81	20.35	0	89.87
Serfdom	125	0.08	0.27	0	1
Population density (ln)	125	23.35	14.22	0.19	80.68
Height	77	162.12	1.34	158.00	164.80
Pasture suitabilitity	131	0.10	0.04	0.04	0.24

Table 8. Descriptives of variables used for European regions during the 19th century

Notes: We report here descriptives for numeracy and other variables for 184 European regions during the 19th century. However, the evidence on age at marriage sometimes requires that we assign the same value to two regions, as age at marriage was reported on a slightly more regionally aggregated level. Hence, the regressions below require clustering at the aggregated level resulting in 153 regions with independent age at marriage and numeracy values. *Sources*: see text.

	(1)	(2)	(3)
Age at marriage	4.973***	2.646***	1.900***
	(0.663)	(0.741)	(0.616)
Share large landowners			-10.72**
			(4.941)
Protestantism			0.153***
			(0.0517)
Country FE	N	Y	Y
Constant	-32.12**	-3.292	19.37
	(15.00)	(15.61)	(18.02)
Observations	153	153	120
R-squared	0.305	0.776	0.817

Table 9. Cross-sectional regressions of regional numeracy during the 19th century

Notes: *** indicates significance at the 1% level; ** at the 5% level; and * at the 10% level. The number in brackets corresponds to robust standard errors, clustered at the regional level. *Sources*: see text. We estimate OLS to assess the conditional correlation between numeracy (which is the dependent variable) in column (1) using age at marriage as explanatory variable for all European countries with available data. In order to control for unobserved country heterogeneity (e.g culture, measurement concept, etc), we add country fixed effects in column (2). In column (3), we confirm the impact of age at marriage by adding two additional explanatory variables (area share of large landowners and Protestantism).

	(1) Bulgaria & Hungary	(2) Serbia	(3) Spain & Italy	(4) Russia
Age at marriage	2.695**	3.838**	0.971**	2.015*
	(1.142)	(1.776)	(0.463)	(1.104)
Constant	27.42	3.066	69.96***	20.69
	(25.84)	(33.32)	(11.40)	(22.93)
Observations	63	15	30	76
R-squared	0.077	0.200	0.114	0.058

Table 10. OLS regressions of regions, by country (or combinations of two countries).

Notes: *** indicates significance at the 1% level; ** at the 5% level; and * at the 10% level. The number in brackets corresponds to the robust standard errors. *Sources*: see text. We report OLS estimates of age at marriage as a potential determinant of regional numeracy, separately either by a single country, or by two countries combined if the sample sizes are not large enough.

	(1)	(2)	(3)	(4)
Second stage				
Age at marriage	7.209***	7.358***	7.257***	7.345***
	(2.681)	(1.409)	(1.322)	(2.311)
Share large	-26.48***	-61.60***	-63.31***	-22.16**
landowners	(8.996)	(13.00)	(11.80)	(9.173)
Protestantism		-2.408*	-2.741**	-0.145
		(1.464)	(1.171)	(0.964)
Heights		-61.60***	-63.31***	-22.16**
		(13.00)	(11.80)	(9.173)
Country FE	N	N	N	Y
Constant	-70.98	324.7	381.4**	-89.46
	(57.81)	(248.3)	(190.8)	(150.3)
First stage				
Pasture Suitability	12.59***	28.43***	30.49***	10.55***
	(3.236)	(6.571)	(6.775)	(3.037)
Constant	19.79***	38.52	0.607	7.152
	(0.619)	(26.42)	(37.70)	(19.56)
F-stat. (1st stage)	15.14	18.72	20.26	12.07
Endogeneity (p-value)	0.17	0.09	0.09	0.01
Observations	119	57	57	57
R-squared	0.180	0.465	0.467	0.771

Table 11. IV regressions of all European regions.

Notes: *** indicates significance at the 1% level; ** at the 5% level; and * at the 10% level. The number in brackets corresponds to the robust standard errors, clustered at the regional level. *Sources*: see text. We report the IV regressions of the effect of regional age at marriage for all European regions. We use the relative suitability of pasture as an instrument. In columns (2)-(4) we add heights to access exclusion restriction. Column (4) contains country fixed effects. The IV is constructed based on pasture and cereal suitability data, provided by FAO and IIASA (2007). *Suitability of global land area for rainfed production of cereals (intermediate level of inputs) (FGGD)*, online, last accessed 5 December 2012, dataset downloadable at http://www.fao.org:80/geonetwork/srv/en/resources.get?id=14077&fname=cereal_int.zip&access=private, see also documentation at http://www.fao.org/geonetwork/srv/en/metadata.show?id=14077). For more details, see Van Velthuizen, V., Huddelston, B., Fischer, G., Salvatore, M., Ataman, E., Nachtergaele, F., et al. (2007). *Mapping biophysical factors that influence agricultural production and rural vulnerability,* Rome: FAO.

Controls included in the ratio	AET ratio	Oster tests
Full model	2.28	1.14
Robustness check: model without land inequality	10.97	1.32

Table 12. Altonji-Elder-Taber ratios and Oster tests: omitted variable bias?

Notes: We calculate Altonji–Elder–Taber and Oster ratios to assess potential omitted variable bias. Under the assumption that selectivity from observables and unobservables are proportional, we can estimate that the effect of unobservables needs to be at least two to eleven times stronger than the one of observables to eliminate the coefficient of main interest (here: age at marriage). Observable control variables are land inequality, protestant share and serfdom. As a robustness check, land inequality was not included in the second line. For the estimation, we included fixed effects and used the areg function, as our dataset for the least square dummy variable estimate contains many categorical variables. As the control variables removed a modest part of the size of the age at marriage coefficient, the unobservables would need to have a very strong effect to completely eliminate the impact of female autonomy – under the assumption of roughly proportional selection on observables and on the unobservable variables.
The Oster delta reflects how strongly correlated the unobservables would have to be with age at marriage, relative to the joint effect of the observables, to account for the full size of the age at marriage coefficient. Given that the Oster delta is larger than the critical value of |1|, it is unlikely that unobservables would be much more related to numeracy than the observable controls. We thus conclude that both AET ratios as well as the Oster tests indicate a low probability of identification issues caused by omitted variable bias.

Panel A: for the panel data set

Age at marriage	3.422
Spatial std. error, 250 km	(0.083)***
Spatial std. error, 500 km	(0.101)***
Spatial std. error, 750 km	(0.108)***
Observations	93
Controls included	YES

Panel B: for the regional data set

Age at marriage	3.926
Spatial std. error, 250 km	(0.099)***
Spatial std. error, 500 km	(0.097)***
Spatial std. error, 750 km	(0.0981)***
Observations	120
Controls included	YES

Table 13: Assessing spatial autocorrelation

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are corrected for cross-sectional spatial dependence and (in panel A) for panel-serial correlation. In panel A, we regress numeracy on age at marriage (an indicator for female autonomy) and the three control variables war, civil war and serfdom. The unit of observation is country and half century. We calculate Conley standard errors using the distance of 250, 500 and 750 km. In panel B, we regress numeracy on age at marriage, land inequality, protestant share and serfdom.

Appendix A: Representativeness of our sample

We analyzed our sample of 27 countries relative to all sizeable European countries. We actually only lack six countries among the European countries that were situated to the west of the Uralian mountains and the Caucasus (see notes to Table A.1). To assess the representativeness, we compare the urbanization ratio in 1850 of our 27 samples countries with the 6 countries on which no data is available (The urbanization ratio is often taken as an indicator of early development level). Among this six, only one – Moldova – had a much lower urbanization ratio in 1800. Latvia and Greece had urbanization ratios close to the centre of the European distribution, and Albania, Macedonia and Slovenia might have been in the upper third of the distribution. According to this criterion, no substantial bias can be expected. A potential alternative criterion, GDP per capita in 1850, is only available for a modest number of countries, and does not allow substantial insights into this issue. Among these, Albania is the poorest and Greece in the lower half of the distribution. However, the urbanization ratio is available for a large number of countries and suggests that our sample is not substantially biased.

Country	co	GDP	urbanization ratio	year
Moldova	md		0.027	1800
Czech Republic	cz		0.035	1800
Finland	fi	911	0.041	1850
Belarus	by		0.042	1800
Russia	ru		0.042	1800
Ukraine	ua		0.044	1800
Sweden	se	1076	0.063	1850
Slovakia	sk		0.064	1800
Romania	ro	931	0.074	1850
Lithuania	lt		0.074	1800
Estonia	ee		0.075	1800
Norway	no	956	0.077	1850
Latvia	lv		0.078	1800
Poland	pl	946	0.091	1850
Ireland	ie	1775	0.100	1850
Greece	gr	1008	0.110	1850
Germany	de	1428	0.116	1850
Denmark	dk	2181	0.118	1850
Switzerland	ch	2339	0.137	1850
Bulgaria	bg	840	0.142	1850
Austria	at	1650	0.158	1850
France	fr	1597	0.163	1850
Hungary	hu	1092	0.170	1850
Albania	al	446	0.220	1850

Croatia	hr		0.220	AL
Macedonia	mk		0.220	AL
Serbia	cs		0.220	AL
Slovenia	si		0.220	AL
Belgium	be	1847	0.251	1850
Italy	it	1481	0.251	1850
Spain	es	1079	0.256	1850
Portugal	pt	923	0.290	1850
United Kingdom	uk	2330	0.303	1850
Netherlands	nl	2355	0.340	1850

Notes: Source for urbanization and GDP: clio-infra.eu. For urbanization, we took 1850, or 1800, if the former was not available. For former Yugoslav countries, the value for Albania is included. For GDP, we took 1870 if 1850 was not available. As criterion for “sizeable countries, we used those with more than 500,000 inhabitants in 1990, admittedly an anachronistic criterion, but many studies observed a high correlation of the population, except for France decling a bit in relative population size in the 19th century, see clio-infra.eu).

Appendix B: Sources for numeracy estimates in our panel data set

All estimates and their sources are reported in the Clio-Infra.eu page (and the working paper referenced by it), except for the following additions (we always report a country-two-letter ISO code, followed by the beginning of the halfc century for which the estimate was reported, and the abbreviated source). At the end of this document, abbreviated source are referenced.

Census data:

at	1700	Tollnek and Baten (2013WP, 2017)
at	1750	Tollnek and Baten (2013WP, 2017)
be	1600	Schroelkamp (2010)
be	1650	Schroelkamp (2010)
bg	1800	Baten and Hippe (2018)
by	1700	Baten et al. 2017
by	1750	Baten et al. 2017
cs	1700	Benyus (2009)
cs	1750	Benyus (2009)
cz	1800	Baten and Hippe (2018)
de	1600	Luginsland (2015)
de	1650	Tollnek and Baten (2013WP, 2017)
de	1700	Tollnek and Baten (2013WP, 2017)
de	1750	Tollnek and Baten (2013WP, 2017)
dk	1700	Tollnek and Baten (2013WP, 2017)
dk	1750	Tollnek and Baten (2013WP, 2017)
ee	1750	Baten et al. 2017
es	1650	Tollnek and Baten (2013WP, 2017)
es	1700	Tollnek and Baten (2013WP, 2017)
hr	1800	Benyus (2009)
hu	1600	Benyus (2009)
hu	1650	Benyus (2009)
hu	1800	Baten and Hippe (2018)
it	1650	Tollnek and Baten (2013WP, 2017)

it	1700	Tollnek and Baten (2013WP, 2017)
lt	1700	Baten et al. 2017
lt	1750	Baten et al. 2017
pl	1600	Baten et al., 2017
pl	1650	Baten et al., 2017
pl	1700	Baten et al., 2017
pl	1750	Baten et al. 2017
ro	1600	Benyus (2009)
ro	1650	Benyus (2009)
ro	1700	Benyus (2009)
ro	1750	Benyus (2009)
ru	1600	Baten et al. 2017
si	1800	Baten and Hippe (2018)
sk	1600	Benyus (2009)
sk	1650	Benyus (2009)
ua	1650	Benyus (2009)
ua	1700	Baten et al. 2017
ua	1750	Baten et al. 2017

Court records (women and men accused as witches)

de	1500	Zillner (2014)
fr	1500	Zillner (2014)
uk	1500	Zillner (2014)

Death register data

cz	1600	Familysearch
cz	1650	Familysearch
dk	1600	Familysearch
dk	1650	Familysearch
es	1750	Familysearch
fr	1600	Lucas (2010)

fr	1650	Familysearch
ie	1650	Familysearch
it	1500	Campestrini (2015)
nl	1600	de Moor and van Zanden (2006)
nl	1650	de Moor and van Zanden (2006)
nl	1700	de Moor and van Zanden (2006)
pt	1800	Familysearch
se	1600	Familysearch
se	1600	Familysearch
se	1700	Familysearch
se	1750	Familysearch
uk	1650	Familysearch

Two types of sources were added that do not adhere to the clio-infra numeracy evaluation standards of the absence of social selectivities: death records and court records of women who were accused of witchcraft. They were included in the full sample, as they fill important gaps. Women accused of being witches are portrayed by popular history books as being the "wise women of the Middle Ages" (and the early modern period, when most were actually accused). However, a preliminary analysis by Baten (2013) showed that they were less numerate than women who were not accused of witchcraft. In the regressions below we control their bias with a dummy variable strategy, and in robustness analyses we exclude these observations altogether showing that the results did not substantially change by this exclusion.

The witchcraft sources were not affected by gender or regional composition bias. For example, jointly with women who represented two thirds of the accused, also one third of males was asked for their age in front of witchcraft courts. This allows us to remove sex-related bias by applying weights. Regional bias is not a major problem for this source because a wide range of territories and societies accused men and women of witchcraft, and we could apply weights.

Actually, only two out of the 77 observations derive from this source. Hence, their influence is limited anyways.

The other source that we added to fill some important gaps and which is probably biased are death records. As dead people do not respond to questions about age, earlier studies about age-heaping and numeracy found that usually the closest relatives – mostly widows or widowers – were asked about the age of the deceased. Sometimes also priests had asked the deceased before his or her death because it was the priest's responsibility to enter the age and other personal characteristics to the church registers. However, for example, if a foreigner arrived and died, the priest might still have entered the personal data into the registers and took an estimated rounded value for his or her age. In earlier studies, the variation of death records by country and period was closely correlated with the corresponding variation of numeracy based on (mostly unbiased) census records (Plötz 2013). However, there was a negative dummy coefficient of death-record-based numeracy in pooled regressions, which can be adjusted accordingly. We use the same adjustment strategy to cope with this bias for the ten estimates that we estimated based on death records.

Clearly almost no empirical observation in economic studies is without a modest selectivity, but most of the biases of our observations can be identified as being small. However, the two witchcraft court records and the ten death record based observations need a special treatment (described above).

Additional references for Appendix B

Baten, J., 2013. "Were witches the 'Wise Women' of the Early Modern Societies?", Working Paper Univ. Tuebingen.

Baten, J., Szołtysek, M. and Campestrini, M., 2017. "'Girl Power' in Eastern Europe? The Human Capital Development of Central-Eastern and Eastern Europe in the Seventeenth to Nineteenth Centuries and Its Determinants", *European Review of Economic History 21-1* (2017), 29-63.

Benyus, R., 2009. Numeracy in Hungary and the Hungarian Empire. Unpubl. Master thesis, Univ. Tübingen.

Campestrini, M., 2015. Numeracy in Italy. Unpubl. Master thesis, Univ. Tübingen.

De Moor, T. and van Zanden, J. L., 2006.*Vrouwen en de geboorte van het kapitalisme in West-Europa,* Amsterdam.

Familysearch: Familysearch.org/search: a collection of historical genealogical information

Lucas, V., 2010. Ungleichheit und politische Partizipation in Frankreich und Dänemark Unpubl. Master thesis, Univ. Tübingen.

Juif, D. and Baten, J., 2013. "On the Human Capital if Inca Indios before and after the Spanish Conquest. Was there a 'Pre-colonial Legacy'?", *Explorations in Economic History,* 2013, 50-2, 227-241.

Luginsland, A., 2015. "Numeracy in Germany 1600-1800" Unpublished BA thesis, Univ. Tuebingen.

Schroelkamp, K., 2010. Numeracy in Belgium. Unpubl. Master thesis, Univ. Tübingen.

Tollnek, F. and Baten, J., 2017. "Farmers at the Heart of the "Human Capital Revolution"? Decomposing the Numeracy Increase in Early Modern Europe", *Economic History Review 70-3* (2017), pp. 779-809.

Tollnek, F. and Baten, J., 2013. "Farmers at the Heart of the "Human Capital Revolution"? Extended Working Paper version of Tollnek and Baten 2017.

Zillner, M., 2014. Determinanten der frühen Humankapitalbildung: Hexenprozessprotokolle als Quelle Unpubl. Master thesis, Univ. Tübingen.

Appendix C: Sources for regional analysis

Our source for age at marriage estimates in European regions is the evidence provided by the Princeton fertility project, which is online accessible via https://opr.princeton.edu/archive/pefp/ and https://opr.princeton.edu/archive/pefp/russia.aspx (Coale and Watkins 1986, Coale, Anderson, and Harm 1976). If we had the choice between several census years for which evidence on marital status at the various ages was available, we included the earliest, so that the marriage age was estimated preferably for an earlier period than the numeracy evidence (or at least roughly in the same time period, see Table B.1). This strategy worked for most countries, but not for Italy (earliest age at marriage was 1911). However, for Italy we could study later educational data, and the same correlation with age at marriage was observable. While the strategy of obtaining the explanatory for an earlier or at least contemporaneous period is sensible, it is not of overwhelming importance, because the regional differences of marriage ages were quite constant throughout the 19th and early 20th century. We studied this for regions of countries for which several years were available, and the correlations were always extremely high (0.89, p=0.00 for the Italian regions in 1911 and 1921, for example; similary for France, Spain etc.). Similarly, the interregional differences of numeracy are highly stable between census years (See below, section "Intertemporal stability of regional numeracy difference"*).*

Year	Country	Age at marr.	N
1905	Bulgaria	20.7	4
1900	Serbia	18.8	15
1787	Spain	25.9	14
1880	Hungary (Habsb.East)	21.9	59
1911	Italy	23.6	16
1897	Russ. Empire	20.9	76

Table C.1. Age marriage evidence for regions
Note: We report the number of regions for which marriage age is available based on those for which also numeracy is available.

For numeracy evidence on European regions, Baten and Hippe (2018) provide a new and large dataset. During the late 19th century, many countries performed censuses that reported also the population by individual ages on a regional level. This allows to estimate the numeracy by regions. Although regional differences are important, they also remain extremely persistent over time. Overall, numeracy and education in general improve, but the lagging regions of 1850 are the same as those in 1900, and also lag in literacy and schooling in 1930.

Country	Numeracy (year)	Later education data (year)
Bulgaria	1893	
Serbia	1895	
Spain	1900	1930
Italy	1871	1930
Hungary et al. (Habsburg Empire-East)	1869	
Russian Empire	1897	1910-1919

Table C.2. Census years of data for regional numeracy
Source: Adapted and modified from Baten and Hippe 2018. Note: For Italy and Spain the 1930 data refer to literacy, for the Russian Empire the birth decade of the 1910-1919 refers to numeracy.

We also control for land inequality, which has been estimated by Baten and Hippe (2018), as well as population density and protestant religion. They use data from population and agricultural censuses from European countries in the 19th and 20th centuries, and define a large agricultural land holding as extending more than 50 hectares (although the 100 hectare threshold yields similar results). They actually test all size categories for their impact on numeracy. The obvious assumption would be that the largest land owners are the driving force here. However, looking closer at the political economy of the regions, this is less clear because the largest land owners were mostly active in national politics, whereas the aristocracy of more modest standing and wealth (including those who had only 50 hectares) were active in regional and communal politics. “Kartoffeladel” (‘potato nobility’) was the term in Central Eastern Europe for nobility that had to rely on modestly sized estates, and often demonstrated their identification with the nobility group by emphasizing conservative, anti-educational social values even more than the better-endowed parts of the nobility. Moreover, the nobility that had declined to estate sizes of 50 to 100 hectares had the greatest difficulty in affording additional taxes and was, hence, extremely opposed to primary schooling (see Wagner 2005 on these issues). Empirically, they are the first who have really assessed different size categories of large land owners, and find that there is still a negative contribution of those landowners between 50 and 100 hectares, restricting spending taxes on schooling.

Our evidence on serfdom comes from Baten, Szoltysek and Campestrini (2017), using the same definitions as the panel data set above.

The relative pasture suitability evidence used in the IV regressions is based on raster data, with a resolution of 5 arc-minutes, provided by the Food and Agriculture Organization (FAO) and related organisations, which generated this evidence in their project on *Suitability of global land area* (see also Hijmans et al. 2005). It is modern data, requiring the assumption that interregional differences were broadly similar over time, which is frequently made in the relevant literature (See the discussion in Baten and Hippe 2018).

Regional height is provided in a number of studies, such as Coll and Quiroga (1992) for Spain ca. 1900, and Martínez-Carrión, Cámara and Pérez-Castroviejo (2016) on Spain in 1858; Baten, Szoltysek and Campestrini (2017) provide height estimates for the regions of the Russian Empire. And Ahearn, Peracchi and Vecchi do the same for Italy.

Intertemporal stability of regional numeracy difference

The aim of this part of the Appendix of the project is to assess the persistence of differences in numeracy and literacy rates across regions over time. Increases in numeracy or literacy rates over time are unproblematic for the cross-sectional approach, as long as the relative differences between regions remain quite similar. The regions and countries considered are Italy, Russia and Spain.

The approach is simple and relies on considering the correlation between an early census and a later one. We assume that numeracy and literacy are two proxies for human capital (admittedly, some regions might be slightly more advanced in numeracy and others more advanced in literacy. For example, highly urbanised areas sometimes have an advantage in basic literacy, Baten et al., 2014). For Italy, the ABCC-values used in the main article were used (based on the census in 1871), and they are compared to literacy rates from 1930 (also analysed in the main text, see section 6). For Spain, the ABCC-values refer to a census in 1900 and are compared to literacy rates from 1930 (again, both cross-sections analysed in the text). For Russia, both early and late ABCC-values are compared, obtained from censuses in 1897 and 1959 (but considering only the 1910s birth cohort that was not as affected by changes that occurred during the Soviet era), respectively.

The correlation for Italy and Spain between the regional early numeracy and later numeracy or literacy rates in Italy and Spain is quite close (Table B.3).

	Italy	Spain	Russia
Correlation coefficient	0.83 (p=0.00)	0.69 (p=0.00)	0.32 (p=0.09)

Table C.3. Correlation between early numeracy and later numeracy or literacy rates

Additional references for Appendix C

Ahearn, B., Peracchi, F. and Vecchi, G., 2009. Height and the Normal Distribution: Evidence from Italian Military Data. Demography, Volume 46-Number 1, February 2009: 1–25

Annuaire Statistique de la France (various years)

Baten, J., Crayen, D., and Voth, H-J., 2014. Numeracy and the Impact of High Food Prices in Industrializing Britain, 1780-1850. Review of Economics and Statistics 96(3), 418-30.

Baten, J., Szoltysek, M., and Camestrini, M., 2017. “Girl Power” in Eastern Europe? The Human Capital Development of Central-Eastern and Eastern Europe in the Seventeenth to Nineteenth Century and its Determinants. European Review of Economic History 21(1), 29-63.

Coale, A. J. and Watkins, S. C. (eds.), 1986. The Decline of Fertility in Europe: the Revised Proceedings of a Conference on the Princeton European Fertility Project. Princeton University Press.

Coale, A. J.; Anderson, B. and Harm, E., 1979. Human Fertility in Russia since the 19th Century. Princeton, NJ, Princeton University Press.

Coll, S. and Quiroga G.,1992. Height and the standard of living in 20th century Spain: A preliminary report. Working Paper Univ. Cantabria.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, *25*, 1965–1978.

Martínez-Carrión, J.M., Cámara, A.D. and Pérez-Castroviejo, P.M., 2016. Parámetros antropométricos de los reclutas españoles antes de la transición nutricional. Nutricion hospitalaria: organo oficial de la Sociedad Espanola de Nutricion Parenteral y Enteral 33-6, 1-10.

Wagner, P., 2005. Bauern, Junker und Beamte. Lokale Herrschaft und Partizipation im Ostelbien des 19. Jahrhunderts. Goettingen: Wallstein Publisher.

Appendix D: Regressions including time invariant religion variables

As religion did not change much after the 16th century (and we have no evidence before the 16th century), assessing religious differences was not possible in the framework using country fixed effects above. In this appendix we perform an additional robustness-check by including dummy variables for mostly Protestant and Orthodox counties. As Table D1 shows, controlling for religion does not affect the significance of our age at marriage variable.

	(1)	(2)
Age at marriage	3.186***	2.340***
	(0.825)	(0.658)
Protestant	7.210**	7.801**
	(3.160)	(2.946)
Orthodox	-20.74***	-21.37***
	(5.804)	(5.367)
Time FE	N	Y
Region FE	Y	Y
Source FE	Y	Y
Observations	93	93
R-squared	0.693	0.774

Table D.1. Regressions including time invariant religion variables

Notes: *** indicates significance at the 1% level; ** at the 5% level; and * at the 10% level. The number in brackets correspond to robust standard errors, clustered at the country-century level, 63 clusters. As the variable “protestant religion” is mostly time-invariant, we cannot control for country FE, and use region FE instead. *Sources*: see text. In model 1, we do not include time effects, but source fixed effects (using source-specific dummy variables), and regional fixed effects (using world region dummy variables). In model 2, we do include time effects, source fixed effects, and regional fixed effects.