Female autonomy generated successful 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, Wageningen University & Research

sandradepleijt@gmail.com, sandra.depleijt@wur.nl

This is the last working paper version before it was accepted by World Development

Abstract: Does higher female autonomy increase human capital formation? To find out, we employ novel data on numeracy as a proxy for human capital and the demographic indicator female age at marriage as a measure for female autonomy for 27 countries and 153 regions in Europe between 1500 and 1900. Our empirical analysis shows that countries and regions with a relatively high level of female autonomy became success cases and pioneers in long-term human capital development. Because women had an advantage in dairy-farming, we approach endogeneity issues by exploiting 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 European Historical Economics Society Conference, the 18th World Economic History Conference, the CEPR/NYU/CAGE Economic History Conference, the WEast conference in Vienna, as well as the seminar participants at the University of Oxford, Vienna University of Economics and Business and Sapienza University of Rome for helpful comments and suggestions. We are grateful to Steve Broadberry, Sarah Ferber, Bishnu Gupta, Markus Lampe, Avner Offer, Deb Oxley, Paul Sharp and Jan Luiten van Zanden for various useful suggestions and support. Anne Abraham, Carina Haller, Tim Holzapfel and Benjamin Kurtz provided excellent assistance. Alexandra de Pleijt has benefitted from funding made available by the Netherlands Organisation for Scientific Research (NWO), project number 446-16-014. Joerg 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 less educated and developed than they could be (Sen 1990; 1999, 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 (Chari et al. 2017). 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. Numeracy is estimated using the age-heaping method that relies on the observation that many persons report an age such as "I am 40" if they do not know their age exactly (Crayen and Baten 2010). What they really mean is "I am around 40". This reporting behavior results in an over-presentation of ages ending in 5 and 0 in many censuses and similar data. The ratio between the other, not preferred (and not rounded) ages, and the round ages – expressed between 0 and 100[1] – was named the "numeracy index" in the literature, because it correlates with numerical abilities and other educational indicators (see Appendix A).

As the setting of the study, we chose Europe in its development process from the 16th to the 19th centuries. We contribute with new empirical evidence in order to provide a macro-view and long-term perspective of this core question of global development. We define the concept "female autonomy" as the capacity for women to achieve high standards of wellbeing and assume a significant and substantive role in decision-making both in the household and the public sphere. We are using the demographic indicator "female age at marriage" (FAM from here) as a proxy indicator for female autonomy. Low FAM is substantially correlated with low female autonomy, as we discuss below (Delprato et al. 2017; Gruber and Szoltysek 2016). 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). However, as we demonstrate, also in Europe, there were vast differences between Eastern and Southeastern regions with FAMs of around 17 or 18, and Denmark with female marriage ages above 29. There is a large literature on South Asia and Africa that supports this (McGavock 2021; Field and Ambrus 2008; Delprato et al. 2017). For example, Sekhri and Debnath (2014) use Indian data to show that a one-year delay in woman's marriage increases the probability that her children will be able to perform higher level cognitive tasks by 3.5 percentage points (see also Sunder 2019 on Uganda).[2]

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 (mostly informal) education and working-skills 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 – of both genders. 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 Northeastern 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 a 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). 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 human capital formation in Northwestern European countries. With respect to economic development, Diebolt and Perrin (2013) argued, theoretically, that gender inequality retarded modern economic growth in many countries (for a critique, especially on endogeneity issues see Dennison and Ogilvie 2014 and 2016; reply: Carmichael et al. 2016).[3] Our study is the first to directly assess the human capital effects of female autonomy in a dynamic historical context. We find evidence that female autonomy was actually decisive for numeracy formation. 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 B). 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 robustness of the relationship, also with respect to causality. 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 dairy farming to cover a high share of total agricultural production (Boehm 1995). Similarly, pasture suitability of the soil (relative to grain suitability) allows dairy to be productive. 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.[4] In our instrumental variable regressions, we show that the relationship between female autonomy and human capital is likely to be causal. Moreover, we gain insights into identification 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 16th 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. They found that persons who received an unexpected income did not increase the schooling of their children, but spent it on other purposes. However, our work will concentrate on the nexus between female autonomy and numeracy, not on the effects of numeracy on income growth – we note this here only to explain the wider context and relevance of our study.

In this study, 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 human capital development. 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 FAM explains almost 50% of the variation in numeracy between 1500 and 1850.

2. Relationship between female autonomy and human capital formation

In the empirical analysis of our study, we use the FAM as a proxy for female autonomy. Low FAM is usually associated with low levels of autonomy – one 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 FAMs of around 17 or 18, and Denmark with female marriage ages above 29 (the latter corresponding with high female autonomy). In addition to the general indicator function, we argue that FAM 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 of both genders, 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). This description of the underlying mechanism is crucial, as it allows understanding development in education until today: Even in societies in which female autonomy is very low – such as in parts of South Asia and the Middle East – it is often the responsibility of women to take care of the basic education of children of both genders. This is partly determined by the organization of the household, in which males typically left the household for work outdoor, while women are staying at home, and encouragement and support for children in their learning at home from the adult side is given mostly by the mother (de Moor and van Zanden 2010). If female autonomy is low, mothers are not educated themselves, and could not collect organization experience on the labor market, their input to their children's learning is expected to be modest in many cases. This mechanism is often observed today, and it is very likely to have been the main mechanism during the early modern development phase of Europe that we study empirically in this article (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 milkmaids, 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 (and skills for numerical proportions for estimating labor duration), which was more challenging outside of the family. Within families, in contrast, rules were often set by the husband. In general, we argue that both female autonomy increased labor market participation and that the demand for female labor might have increased female autonomy. Both factors might go hand in hand, we focus here on the effect on human capital formation.

What were the main underlying determinants to this causal chain? Apart from culturally idiosyncratic factors, geographic and climatic conditions allowing for dairy farming 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 FAM), and this in turn enabled better numeracy for both genders in some of the countries and regions studied here.

3. 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) (see Appendix C for an overview). In Table 1, numeracy estimates range from 36 to almost 100. The number of underlying observations is reported in Table 2 and the values in Table 3. Although later centuries have more observations, even the early ones are not small, and Appendix D explains that they are not regionally selective samples. We apply a high 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). Furthermore, 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). We added a number of cases applying the same standards (Appendix C and D). We include some sources which are not fully compatible with these standards, i.e. court records of witchcraft accusations and mortality registers. These sources are assessed systematically in several robustness tests below.

European historic al demographers provided a rich fundus of FAM data for almost all European countries (Appendix A in Dennison and Ogilvie 2013).[5] We added a set of Portuguese and Central-Eastern European estimates for early periods, which were collected by Mikolaj Szoltysek, a renowned specialist for Central-Eastern 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 et al. 2017). The values and the number of underlying studies are reported in Table 4 and 5.

Table 1 Table 2 Table 3 Table 4 Table 5 here

* Figure 1 here*

The early 16th century saw high rates of numeracy in Central Europe and the Netherlands, whereas Spain had lower values (Figure 1, Table 3). 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.[6]

In order to cross-validate our evidence, we now compare the FAM indicator with other proxies for female autonomy (see also Appendix E). Several indicators have been suggested to approximate female autonomy (Gruber and Szoltysek, 2016). 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.

* Table 6 here*

Gruber and Szołtysek (2016) have studied the correlation between these indicators for Europe -- mainly between the eighteenth and early twentieth century -- using not less than 700,000 underlying observations. They aggregated these on the place and time level. The results in Table 6 show that our FAM 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). 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).[7]

* Figure 2 here*

In Appendix E we also assessed this correlation using more recent and even earlier data. For the period around 1900, we calculated gender inequality in literacy as the difference in literacy between males and females, divided by the level of literacy of males (see Figure 2). For example, in Central Serbia with its very low FAM, literacy for males was 34.2%, the one for females only 7.1, hence the gap was 27.1%. Relative to male literacy level, the gap represents 79.2% of male literacy. In contrast, in Lithuania, a region influenced by neighboring Scandinavia in its high FAM, but governed by the Russian Empire that invested low amounts in schooling, the FAM was almost 25, the literacy of males and females 50.1 and 41.8%, respectively, and the gender inequality of literacy relative to the level only 17.6%. We observe a close negative relationship between the gender inequality of literacy and FAM: The correlation is -0.73 (p=0.000, N=111). If we perform regressions with country fixed effects or including the level of literacy, still a significant coefficient of FAM emerges. This is also not a boundary effect of regions close to 100% literacy having automatically lower gender inequality – even if we remove the most literate quarter of the sample (including only 80% and lower literacy), the correlation coefficient is still -0.60, p=0.000, N=82).

Moreover, some additional early evidence can be gained from archaeological samples from female and male long bones (that are correlated with height). Maravall and Baten (2019) recently developed this technique looking at relative long bone length and relative frequency of enamel hypoplasia values among males and females.[8] In Appendix E we observe that gender equality of nutrition (reflected in femur length of skeletons of both genders) correlates with FAM by country.

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 FAM indicator is available for a substantial number of countries. [9] Especially the earlier centuries are not covered by these other indicators in a representative way. Fortunately, given the correlation between the variables, FAM is a reasonable indicator that reflects various aspects of female autonomy.

* Figure 3 here*

In Figure 3 we map FAM in four periods between the 16th and 19th century. For the 16th century, we observe that the UK had high values. This might be related to the "educational revolution" in England which took place in the 16th and 17th centuries (Stone 1964). In the middle group, we find the Scandinavian and Central European countries, whereas Southern Europe had lower FAM. The South of Europe had lower values relative to the North, which might have foreshadowed the "Little Divergence" (De Pleijt and van Zanden 2016). In the 18th century, Central Europe showed high FAM values, whereas now some of the East-Central European economies enter the picture with relatively low rates. This continues in the 19th century, when especially South-Eastern Europe had low rates. In contrast, Scandinavia and Central Europe, including Switzerland and Austria, had quite high FAM. These are exactly those countries which became the success cases in the Second Industrial Revolution, which took place shortly thereafter in the late 19th and early 20th century.

In the regressions below we include a number of control variables (see appendix F for an overview). 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.

4. Empirical analysis: research design

To determine the importance of female autonomy for human capital formation, we explore the empirical relationship between FAM and numeracy, while controlling for potential confounding factors as discussed above. 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 (France, Germany, UK, Spain, Italy), but already 1600-49 is covered by a substantial number of countries (see Table 2 and 3). However, FAM data is only available by century, hence we will assign the same value for FAM to two half-century numeracy observations below.[10] 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 FAM for 16 countries and Empires (which include several modern countries) – i.e. Bulgaria, Serbia, Hungary (including modern Hungary, Slovakia, parts of Romania, Croatia, Serbia, Kosovo, 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.

* Figure 4 here*

Starting with the early modern period, Figure 4 depicts a strong and positive relationship between FAM 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, Bulgaria, Poland, Belarus, Slovakia had low values; Italy, Spain, Ireland and other countries were in between. 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 sometimes had lower numeracy than expected based on the female autonomy proxy. But in general, these deviations from the regression line were not substantial. The correlation coefficients ranged between 0.61 and 0.79 (p=0.000, see notes to Figure 4).

Of course, the observed relationship between female autonomy and numeracy is not necessarily causal. Higher numeracy and FAM 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 female marriage ages.

We base the construction of our instruments on the studies by Alesina et al. (2013), Voigtlaender and Voth (2013) and a vast related literature. They 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 dairy farming, which was typically associated with a more active role for women.[11] Voigtländer and Voth (2013) similarly suggest that the relative prevalence of dairy farming over grain cultivation might be an important determinant of differences in FAM (Baten et al. 2017). Dairy farming 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 specialize in dairy farming, because dairy products could be more easily consumed (dairy countries: see Lampe and Sharp 2015).[12] Hence, we would expect higher levels of female autonomy in lactose-tolerant populations in Europe.[13]

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.

Could the lactose tolerance data set be affected by regional selectivity? Ingram et al. (2009) used regional evidence on several tolerance measurements per country that allowed to minimize regional selectivity bias. For example, for Finland they obtained lactose tolerance values for seven different regions and populations groups -- and this included Sami from the very far north, Southern Fins, Swedish speaking Fins and several other groups. For Germany, 13 regional groups could be analyzed and for Italy 12. Although not all European countries are equally well documented, the authors of the lactose-tolerance estimates have taken care to avoid regional selectivity wherever possible.

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 dairy 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, Appendix I).

As a potential caveat, we note that lactose tolerance might be the result of a coevolution of culture and genes. Hansen et al. (2015) discussed modern hunter-gatherer societies that were characterized by relatively high gender equality if the gatherer contribution to incomes was strong (especially in Africa). Hansen et al. discuss whether this might have been similar in early societies millennia ago (there is some archaeological evidence, though based on tiny samples), hence a coevolution of culture and genes could be imagined (see Appendix I). If this coevolution of genes and culture would be as old the evolution of lactose tolerance differences, the latter would be not exogenous. At the present stage of archaeological research, we do not know how substantial this effect was, but it is helpful that we can use pasture suitability below as a second source of exogeneity.

However, at the beginning of our study period, 1500, the genetic factor of lactose tolerance was predetermined and dominant. 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 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, 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 arguably 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 5 here*

* Figure 6 here*

Figure 5 and Figure 6 show the variables underlying our instrument which are obviously correlated with FAM. 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 FAM and early industrialization (Komlos 1989). The soil suitability for pasture (relative to grain) is high in Scandinavia, while it is low in eastern and southeastern Europe. In the IV regressions below, we demonstrate that lactose interacted with soil suitability is a strong instrument for FAM. We also assess the exclusion restriction of our instrument by controlling for other potential channels between lactose tolerance and soil suitability for pasture and numeracy. The results, however, show that the effects of lactose tolerance and pasture suitability 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 alone as instrumental variable as lactose tolerance is not available at the regional level. We calculate the ratio between the pasture suitability of a region (=good for dairy farming) over its cereal suitability, using a similar argument for this instrument as for lactose tolerance: it was more economic to specialize in dairy farming there.

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

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

where $N_{i,c}$ captures numeracy (of both genders) in country i in half-century c. $FAM_{i,c}$ is the main variable of interest: female age at marriage in country i in century c. Our data on numeracy is by country and half-century, whereas FAM is available only for centuries. This means that in the regressions below we assign one FAM value 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 7, Col. 1). Although this is not a “small” number of clusters according to Cameron et al. (2008, who defined the threshold between 20 and 50), we also used wild bootstrapped standard errors that confirmed the usual standard errors. In the formula above, μ_c are century fixed effects, $X'\gamma$ is the vector of control variables that we described in the previous section, α is a constant, and ε is the error term. 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). 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.

* Table 7 here*

Carmichael et al. (2016) argued that in some situations, a high FAM is the result of bad economic conditions. Hence we also used a pre-adjusted regression, which adjusts the FAM variable by the GDP level. This makes our analysis consistent with Carmichael et al. (2016).

Next, we turn to the Two-Stage-Least-Square (TSLS) regressions. In the first stage, FAM is instrumented by lactose tolerance interacted with the relative suitability for pasture, relative to the suitability for grain:

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

where $LCTS_i$ is the lactose tolerance and pasture suitability interaction in country i. $X'\gamma$ is the same vector of control variables included in Equation (1), and $\nu_{i,c}$ is the error term. We use the interaction between relative soil suitability for pasture relative to grain and lactose tolerance because the soil suitability by itself is not a strong instrument (given the aggregation issues at the country level). Nevertheless, it results in very similar coefficient estimates if it is included as the only instrument (and the same applies to lactose tolerance, see results in Appendix G).

5. Panel Results

In Table 7, we provide the LSDV regressions measuring the conditional correlation of FAM 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, and we control for region fixed effects instead of country fixed effects, to allow a comparison with the IV estimates below. In columns (6) – (10) we add control variables for war, civil war, serfdom, latitude and longitude. We preadjust FAM in Column (9) for national income, because Carmichael et al. (2016) wondered whether in some specific situations FAM declines after a positive income shock. Nevertheless, we still find a strong positive correlation between GDP adjusted FAM and numeracy (Col. 9). In column (11) we combine the control variables war, civil war and serfdom. Finally, in column (12) we restrict our sample to the period 1500 - 1749 and in column (13) we do so for 1750 -1849. We observe consistently statistically significant correlations between female autonomy proxied with FAM and numeracy. Adding time, country and source fixed effects and control variables reduces the coefficient on FAM slightly, as expected, but it is still found to be statistically significant and of substantial size.

Schultz et al. (2019) studied the role of medieval church exposure for today's behavioral patterns, we therefore include this variable in a set of additional robustness-checks. Table J1 in Appendix J however shows that the correlation between church exposure and human capital formation is not as clear-cut. In particular, within Europe, Scandinavian countries have a very modest church exposure, because missionary activity took place very late in Scandinavia.

* Table 8 here*

In the TSLS regressions of Table 8, we observe a strong correlation of our instrument with FAM in the first stage results, as the F-statistic is substantially larger than 10 (Panel B). The coefficient on FAM in the corresponding second stage is statistically significant, suggesting a positive and causal effect of female autonomy on numeracy in pre-industrial Europe. The IV coefficient is larger relative to the OLS coefficient. This might be caused by the fact that the endogenous variable in the OLS estimate is measured with a higher degree of noise, as Becker and Woessmann (2009) argued.

Is the coefficient of FAM economically meaningful? We consider the effects of one standard deviation of the explanatory variable. If we multiply the standard deviation of FAM (2.56) with its coefficient (6.79 in Col. 2), we obtain 17.38. This is roughly the size of the standard deviation of the dependent variable numeracy (standard deviation: 17.32). If we use the LSDV coefficient of 4.38 (Col. 3 of Table 7) a standard deviation effect is still 11.2 or 65% 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 FAM 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 economically important. We assess a potential nutritional channel that might violate the exclusion restriction below in section 6 on regional evidence.

In Appendix G 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 regress numeracy also on Protestantism and orthodox confession (Table G1). We observe that none of these potential confounders makes the effect of FAM insignificant. 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 via the inclusion of dummy variables for the different source types. However, to address this issue head-on, we have estimated the regressions omitting the evidence derived from "death records" and from "witch trials" (see Table G2). All the coefficients of FAM are still statistically significant and their size changes very little.

After presenting these results for the 1500-1849 period, we take one step back and ask the general question: why was north-western Europe not already ahead of Italy and the

Mediterranean in 1500? After all, the dairy-farming-driven gender equality mechanism might have been working already quite early. However, recent research showed that the Western Mediterranean, including Italy, was always more advanced economically – from Roman antiquity to the 15th century (Keywood and Baten 2021, see their Figure 8). Hence the strong position of Italy in the 14th and 15th centuries can be well explained by path-dependency. The astonishing development that we had to explain in this study was rather the catching up of the Northwest and the subsequent overtaking of the Mediterranean during the following centuries.[14]

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.

* Figure 7 here*

* Figure 8 here*

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 (Appendix F lists the sources). More specifically, we regress numeracy by region on FAM, using robust standard errors and clustering (see notes to Table 10):

$$N_{r,c} = \alpha + \beta_1 \, FAM_{r,c} + X'\gamma + \varphi_c + \varepsilon_{r,c} \, , \qquad (3)$$

where $N_{r,c}$ captures numeracy in region *r* in country *c*. $FAM_{r,c}$ is the main variable of interest: female age at marriage in region *r* in country *c*. $X'\gamma$ is the vector of control variables, which differs slightly from equation (1) as some of the control variables are not available at the regional level, while others can be assessed. 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 include the share of protestants in order to control for religious effects.

In the first stage, FAM is instrumented by relative pasture suitability, which is a measure for how well the soil and climate is suited to dairy farming, relative to grain agriculture.[15] Baten et al. (2017) explore this instrument for eastern European female autonomy and find it highly correlated with FAM. 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 dairy farming, which is usually associated with higher female autonomy. Hence, we instrument FAM in region r and country c ($FAM_{r,c}$):

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

where PS_r is the relative pasture suitability of region r. $X'\gamma$ is the vector of control variables included in Equation (3). We control for heights as pasture suitability may influence numeracy via a nutrition effect. Finally $\nu_{i,c}$ is the error term. In Table 9, we summarize the descriptives for the regions, where numeracy ranges from 21 to 100 percent.

* Table 9 here*

* Table 10 here*

In Table 10, we analyze the effect of regional FAM, 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, such as culture or measurement concepts. Finally, column (3) combines the country fixed effects specifications with additional control variables. All three models indicate that FAM has a statistically significant impact on numeracy, even in the subnational dataset.

* Table 11 here*

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 11, we use Bulgaria and Hungary and find a statistically significant effect of FAM 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).

For the regional numeracy regressions, we had to concentrate on Southern and Eastern Europe for the 19th century regional data, because numeracy was already close to 100% in Central and Northwestern Europe. Hence, there was not enough numeracy variation to be analyzed for the UK, for example. However, we can provide literacy evidence on the UK for 1840, when the UK still had strong deficits of literacy (that were already absent in Scandinavia and in other countries early on, see Col 5 in Table 11). We observe that FAM also explains the variation of literacy in the UK, which is an important result, as this is one of the pioneering countries of early development.

We show a scatter diagram of all the regions (see Figure 7) and observe that the residual numeracy and the residual FAM 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 FAM. 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 FAM and low residual numeracy. There is still quite a bit of variation on both sides of the regression line. One of the largest outliers is 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.

* Table 12 here*

The results in Table 12 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 FAM or other variables. The results show a strong relationship between FAM 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 a strong instrument. Female autonomy is a stronger explanatory variable than Protestantism, partly because even within the Catholic countries the FAM has explanatory power (Belgium, Slovenia, Czech Lands have higher values, for example).

In all four instrumental variable regression specifications, we observe a statistically significant effect of FAM 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 FAM or is there a separate direct effect on numeracy via nutrition?

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 pasture suitability might affect numeracy via nutritional benefits from consuming dairy products such as milk, hence the effect would not only run via the female autonomy channel. To assess this potential issue, we include in our regional regressions below average human stature, as height is often used as a proxy for nutritional quality (Baten et al. 2014).[16]

We find that the addition of heights does not render our second stage estimates of the FAM coefficient statistically insignificant. The coefficient for human stature is not statistically significant, and it does not affect the FAM indicator we are mainly interested in: the coefficient of FAM 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 pasture suitability did not run via nutritional benefits from consuming milk.

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 (proxied with FAM)? 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 (FAM) 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 ($ß_{full}$).[17] $ß_{restr}$-$ß_{full}$ 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.[18] If control variables only remove a small part of the FAM coefficient, then unobservables would need to have a very strong effect to completely eliminate the impact of FAM – under the assumption of roughly proportional selection on observables and unobservables. We assess 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 FAM.

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 FAM is totally eliminated, while taking into account the R-squared's movements as well. As rule of thumb, an Oster test above 1 suggests the selection of observables does not invalidate results (see Table 13). 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.

* Table 13 here*

* Table 14 here*

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 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 spatially adjusted standard errors, a standard procedure for cross-sectional data. For panel data, in which temporal autocorrelation might also play a role, Hsiang (2010) developed a specific method (see notes to Table 14 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 FAM 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.

7. Conclusion

Our empirical results suggest that countries and regions with more female autonomy became (or remained) success cases 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 female 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 female age at marriage corresponded with 4 numeracy units. Is this economically relevant? The distance between Russia and the Netherlands represents about 8 years of female 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 female 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 and the geological factor of the relative soil suitability for pasture. These factors increased the possibility and demands to spezialize in dairy farming. Dairy farming is an important agricultural activity allowing women to participate in income generation, since it required less 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, the genetic and soil-suitability-related factors influence long-term differences in agricultural and overall economic specialization, implying a set of gender-specific institutions.

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.

We can also note that our results contribute to a heated debate between de Moor and van Zanden (2010) and Dennison and Ogilvie (2016) – our findings support the de Moor and van Zanden view, because the implications of girl power are clearly visible in our empirical results (see also Le Bris and Tallec 2021). In sum, we argue that the female autonomy factor is a crucial one for human capital formation. It plays a particularly important role for the development of numerical skills, which are the ones that matter most for economic growth in the modernizing world after 1500 (Hanushek and Woessmann 2012, 2020). We conclude that the success cases of numeracy development were the ones with high female autonomy.

References

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.

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.

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 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., and van Zanden, J.L. (2008). Book production and the onset of modern economic growth. *Journal of Economic* Growth, *13*(3), 217-35.

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 Campestrini, 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.

Becker, A. (2019). On the economic origins of restricting women's promiscuity. *CESifo Working Papers* No. 7770.

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*, 1455–1495

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

Bolt, J., and van Zanden, J.L. (2014). The Maddison project: collaborative research on historical national accounts. *Economic History Review*, *67*(3), 627-51.

Boserup E. (1970). *Women's role in economic development*. George Allen & Unwin.

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

Cameron, A., Gelbach, J. & Miller, L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414–427.

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.

Chari, A.V., Heath, R, Maertens, A., Fatima, F. (2017). The causal effect of maternal age at marriage on child wellbeing: Evidence from India. *Journal of Development Economics*, *127*, 42–55.

[dataset] Clio Infra Data. www.clio-infra.eu. Last accessed February 20, 2022.

Cook, C.J. (2014). The role of lactase persistence in precolonial development. *Journal of Economic Growth*, *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.

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.

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.

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.

Delprato, M., Akyeampong, K., and Dunne, M. (2017). Intergenerational Education Effects of Early Marriage in Sub-Saharan Africa. *World Development*, *91*, 173 – 192.

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.

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.

Field and Ambrus (2008). Early marriage, age of menarche, and female schooling attainment in Bangladesh. *Journal of Political Economy*, *116*(5), 881-930.

Flatz, G. (1995). The genetic polymophism of intestinal lactase activity in adult aumans. In Scriver, C.R. et al. (eds.). *The metabolic and molecular bases of inherited disease*, Seventh Edition, Mc Graw-Hill.

Galor, O. and Özak, O. (2016). The agricultural origins of time preference, *American Economic Review*, *106*(10), 3064-3103.

Gruber, S., and Szoltysek, M. (2016). The patriarchy index: A comparative study of power relations across historic Europe. *History of the Family*, *21*(2), 133-174.

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

Hansen, C.W., Jensen, P.S., Skovsgaard, C.V. (2015). Modern gender roles and agricultural history: the Neolithic inheritance. *Journal of Economic Growth*, *20*(4), 365-404.

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. Elsevier https://doi.org/10.1016/B978-0-12-815391-8.00014-8.

Hsiang, S. (2010). Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of sciences*, *107*(35), 15367-15372.

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.

Keywood, T. and Baten, J. (2021). Elite violence and elite numeracy in Europe from 500 to 1900 CE: Roots of the divergence, *Cliometrica*, *15*, 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. *Annales de démographie historique* 1977, 271-284.

Kula, W. (1976). *An economic theory of the feudal system: Towards a model of the Polish economy, 1500-1800*. N.L.B. Publisher.

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.

Le Bris, D., and Tallec, R. (2021). The European Marriage Pattern and its positive consequences: Montesquieu-Volvestre, 1660-1789. *MPRA Working Paper* No. 105324.

Maravall, L., and Baten, J. (2019). Valkyries: Was gender equality high in the Scandinavian periphery since Viking times? Evidence from enamel hypoplasia and height ratios. *Economics and Human Biology*, *34*, 181-193.

McGavock, T. (2021). Here waits the bride? The effect of Ethiopia's child marriage law. *Journal of Development Economics*, *149*, Article 102580.

Millward, R. (1982). An economic analysis of the organization of serfdom in Eastern Europe. *Journal of Economic History*, *42*, 513-48.

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*, *1*(18), DOI: 10.1080/07350015.2016.1227711

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 food 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* (pp. 121-52). Habelt Publisher.

Schultz, D., Bahrami-Rad, J.F. Beauchamp, J.P., Henrich, J. (2019). The church, intensive kinship, and global psychological cariation. *Science*, *366*, Article eaau5141.

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.

Sekhri and Debnath (2014). Intergenerational consequences of early age marriages of girls: effect on children's human capital. *The Journal of Development Studies*, *50*(12), 1670-1686.

Sen, A. (1990). More than 100 million women sre missing. *New York Review of Books, 37*, 20.

Sen, A. (1999). Development as freedom. Oxford University Press.

Stone, L. (1964). The educational revolution in England, 1560 – 1640. *Past & Present*, *28*(1), 41-80.

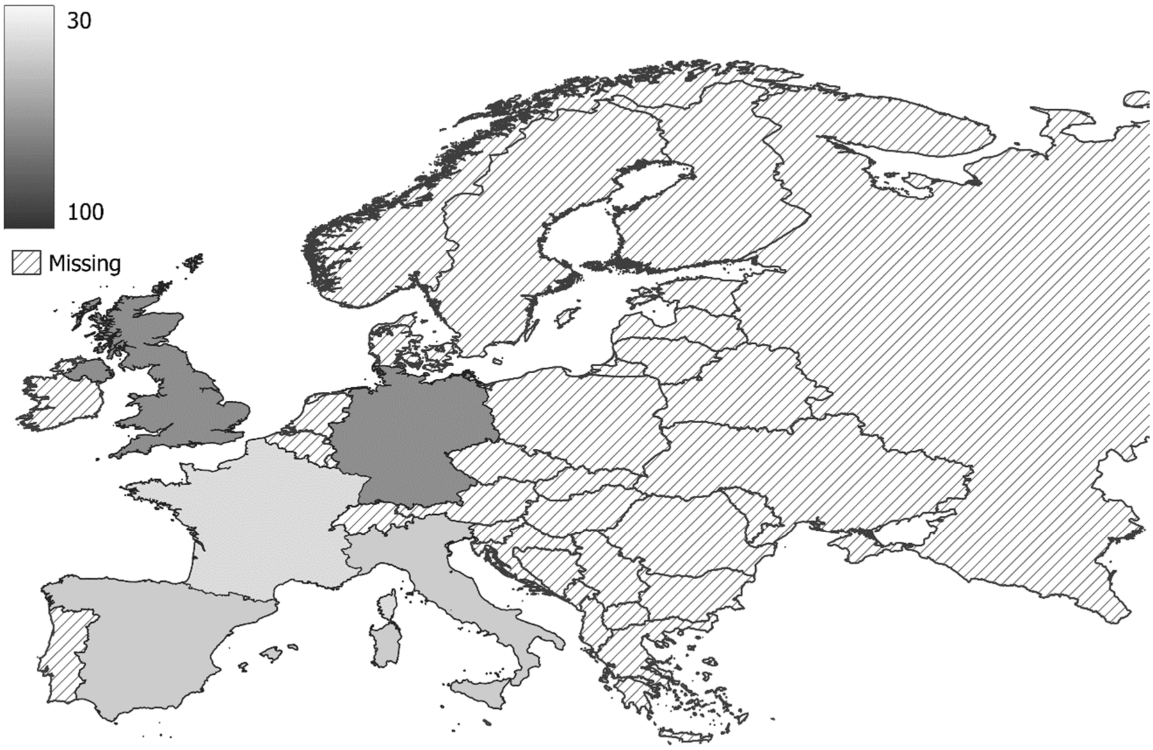
Sunder, N. (2019). Marriage age, social status and intergenerational effects in Uganda, *Demography*, *56*(6), 2123–46.

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.

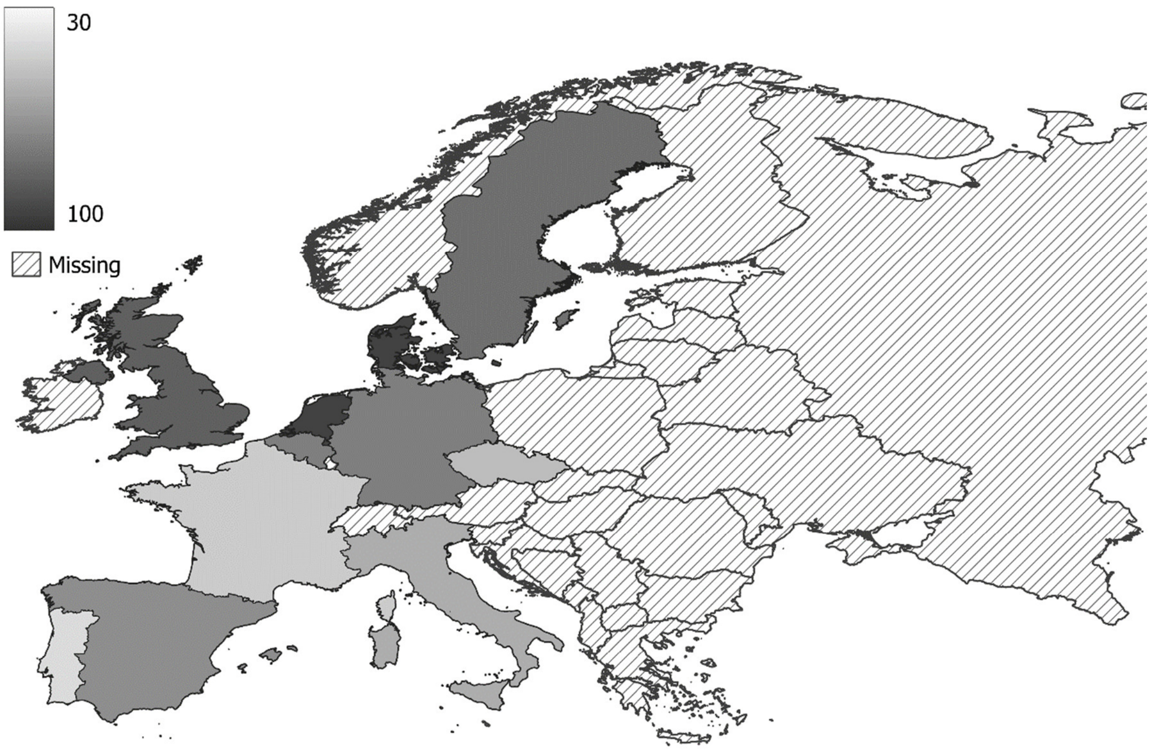
United Nations (2005). *Progress towards the Millennium Development Goals, 1990–2005*. Secretary-General's Millennium Development Goals Report. United Nations Publications.

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*, FAO Publications.

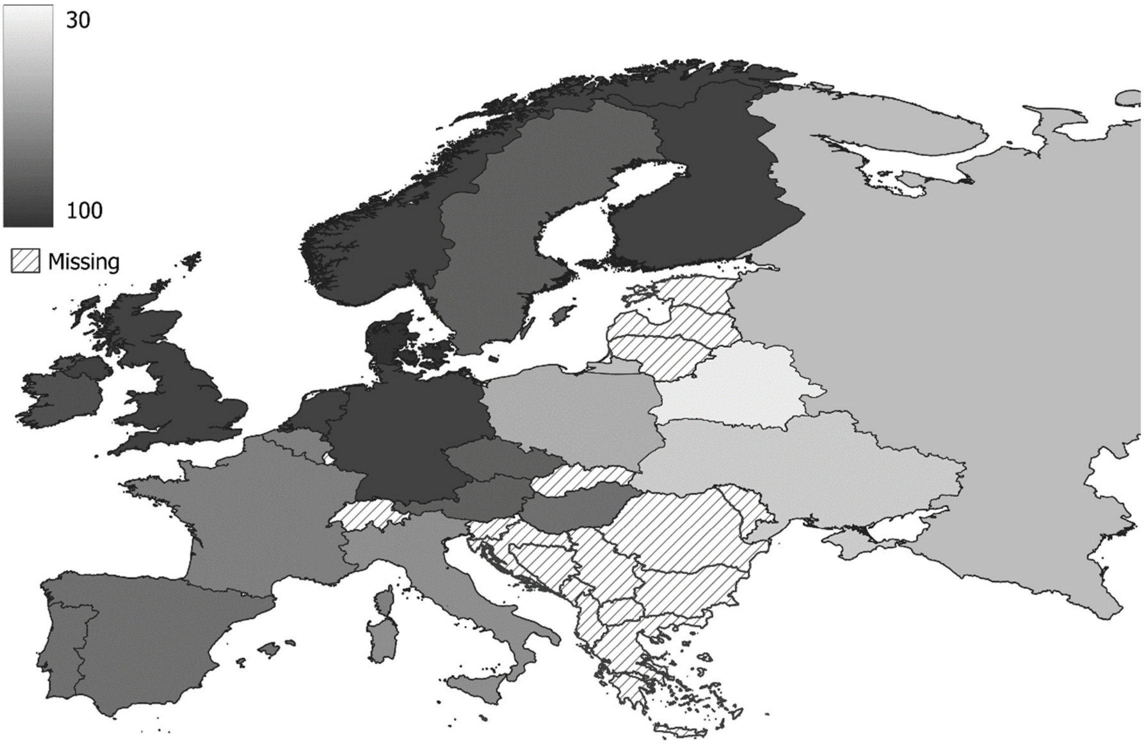
Voigtländer, N., and Voth, H-J. (2013). How the west 'invented' fertility restriction. *American Economic Review*, *103*(6), 2227-2264.



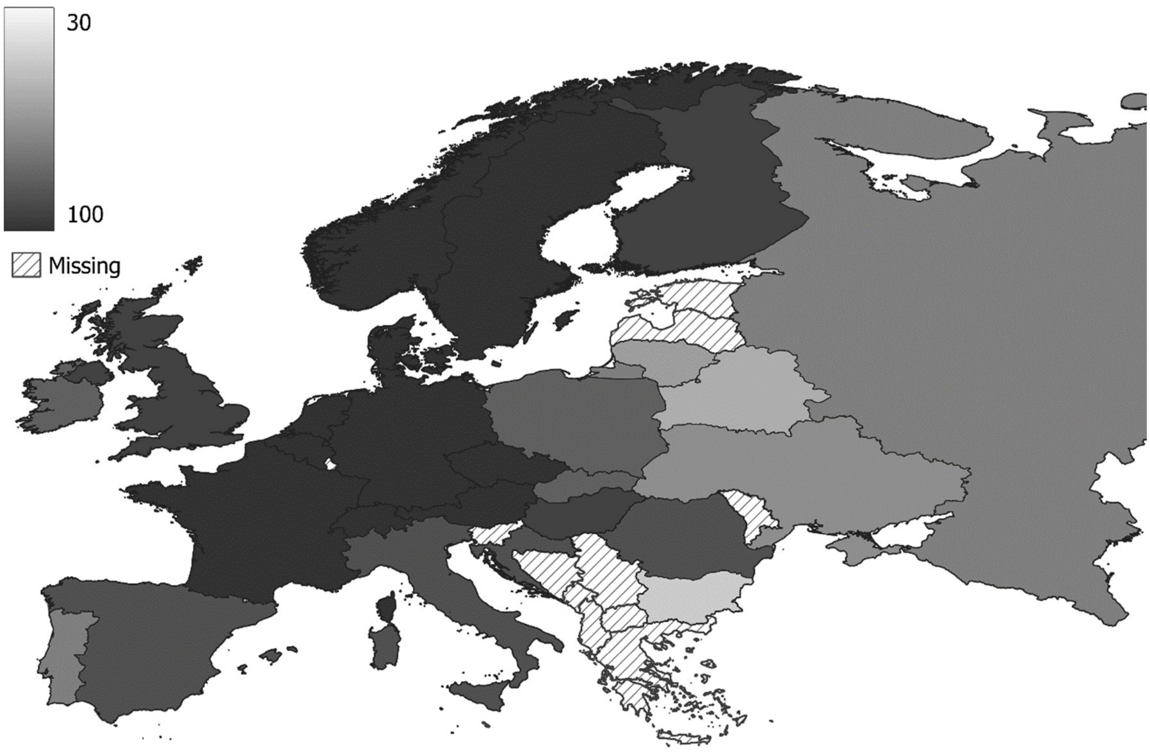
Panel (a): 1500-49



Panel (b): 1600-49



Panel (c): 1700-49



Panel (d): 1800-49.

Figure 1. ABCC Index in Europe, 1500-49 - 1800-49.
Notes and sources: See Appendix C.

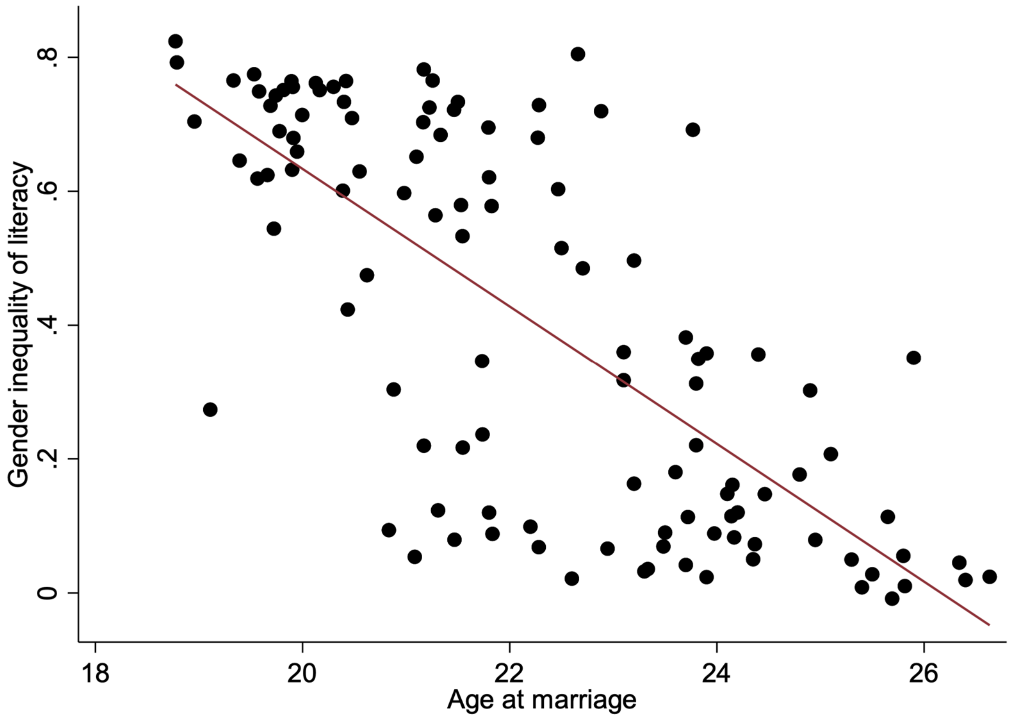
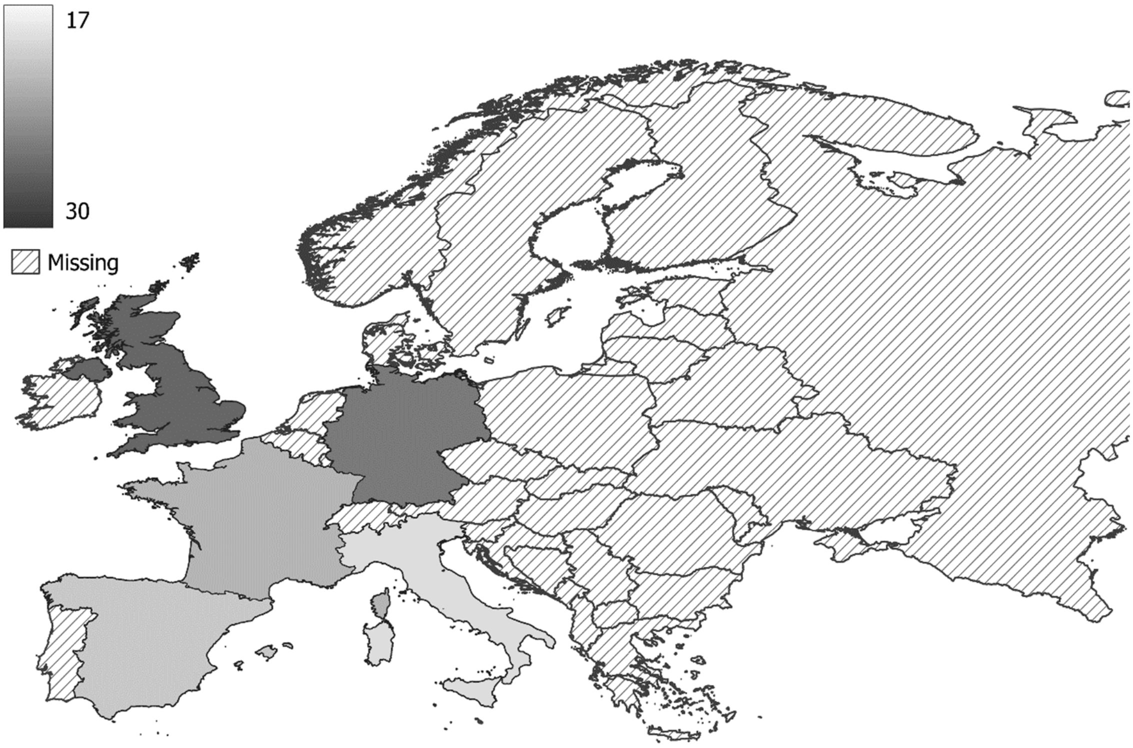
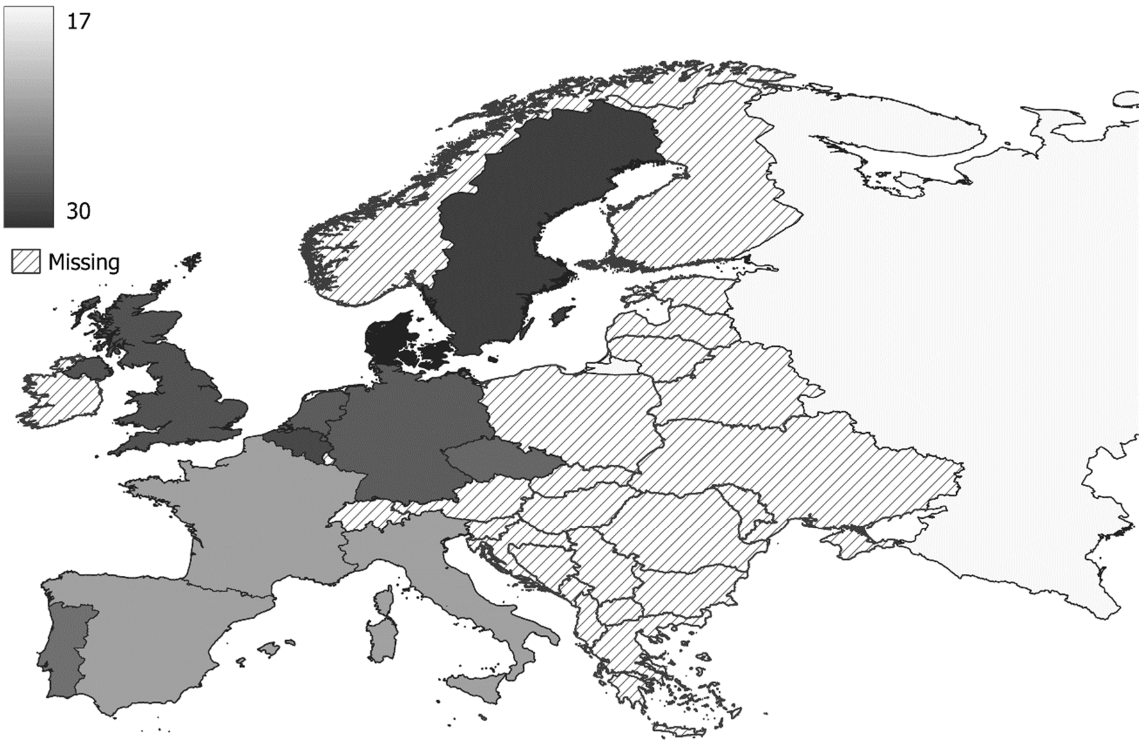


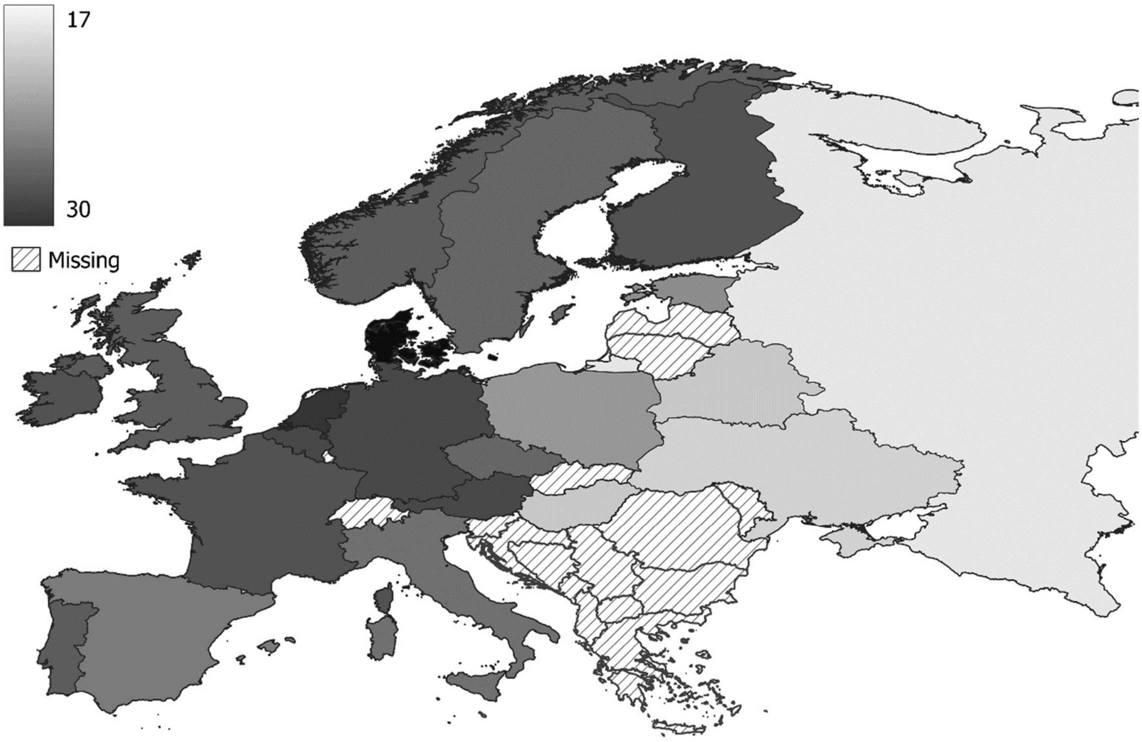
Figure 2. Female age at marriage and gender inequality of literacy around 1900
Notes: As Northwestern and Central Europe had already literacy rates close to 100% in 1900, it is more informative to include the regions of South and East Europe. However, Austria and France also had some regions with "sufficiently" low literacy still. Other regions included are the ones of the modern countries of Armenia, Azerbaidshan, Bulgaria, Belarus, Croatia, Estonia, Georgia, Hungary, Italy, Lithuania, Latvia, Moldavia, Poland, Romania, Russia, Serbia and Montenegro, Slovakia, Ukraine.



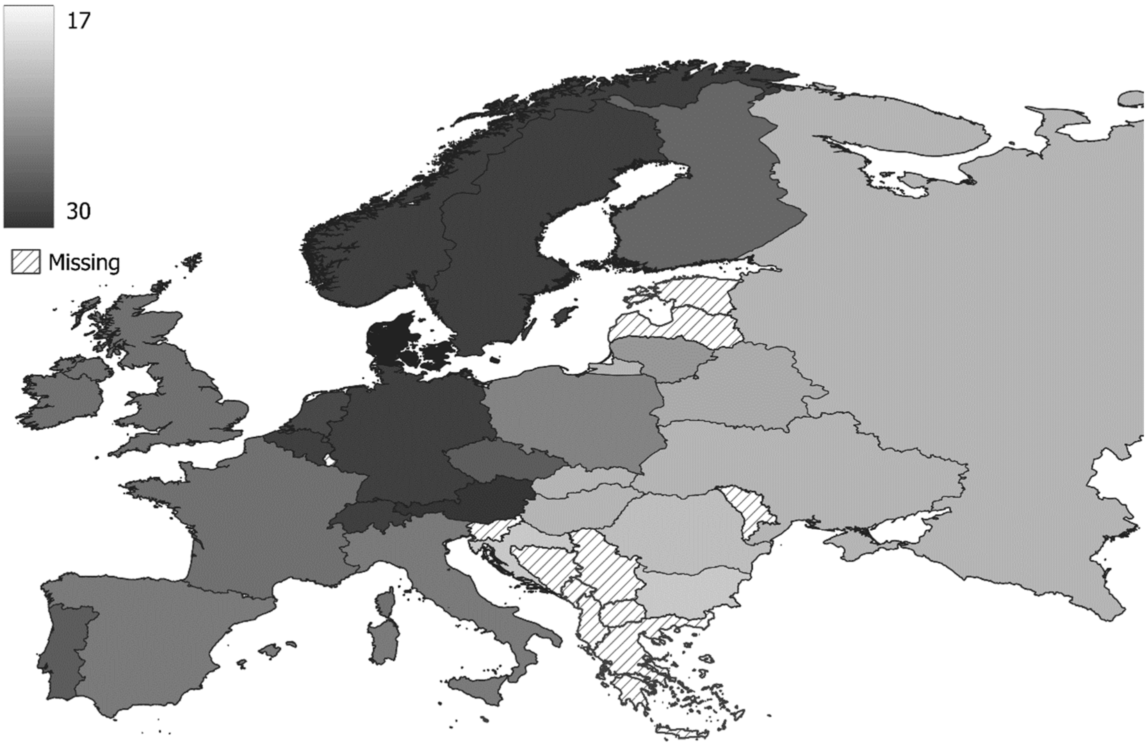
Panel (a): 1500-99



Panel (b): 1600-99

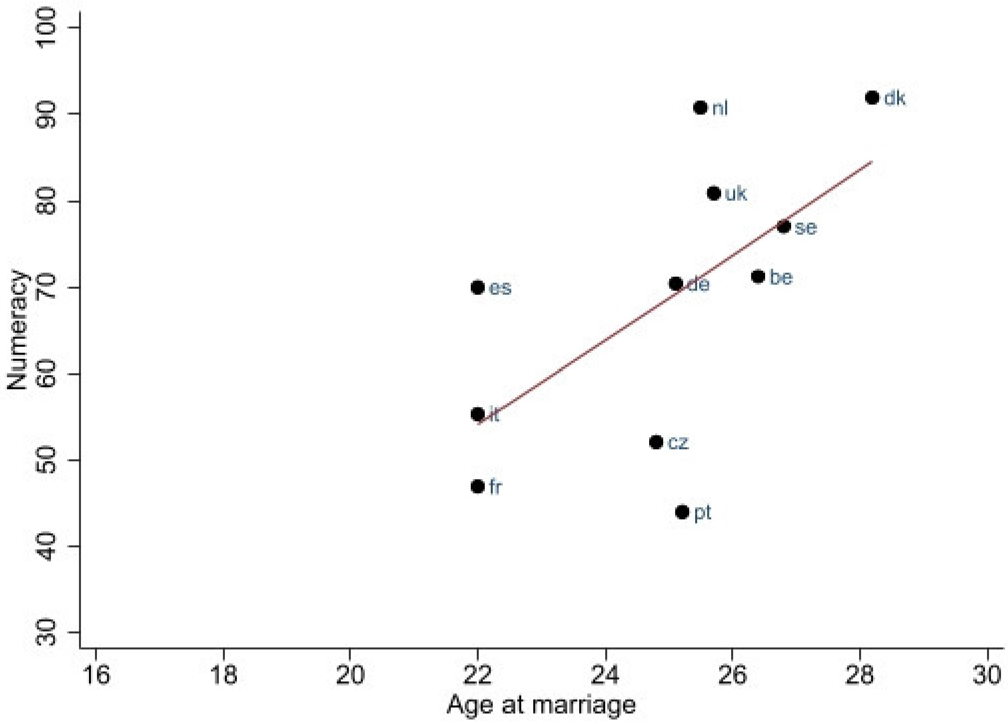


Panel (c): 1700-99

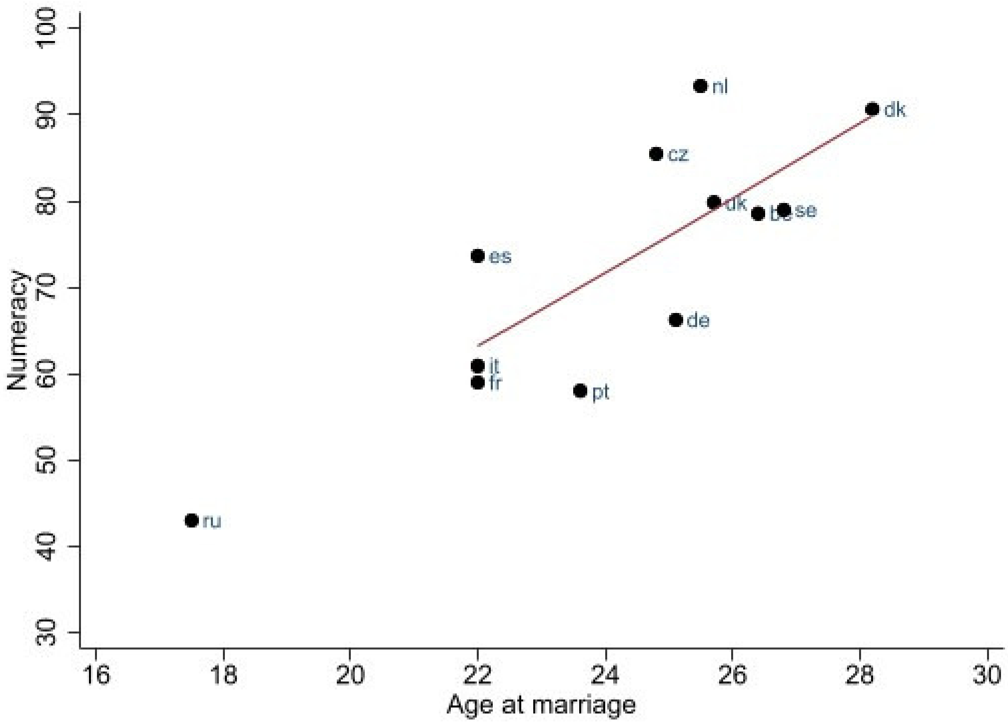


Panel (d): 1800-99

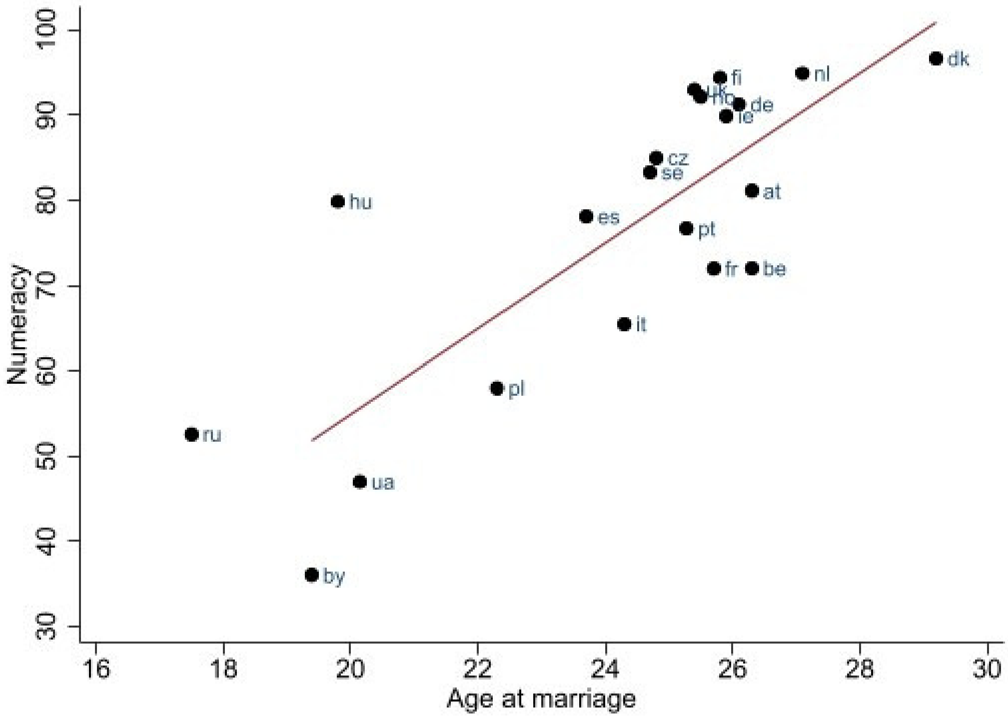
Figure 3. Female age at marriage in Europe, 1500-99 - 1800-99.
Notes and sources: See Appendix F.



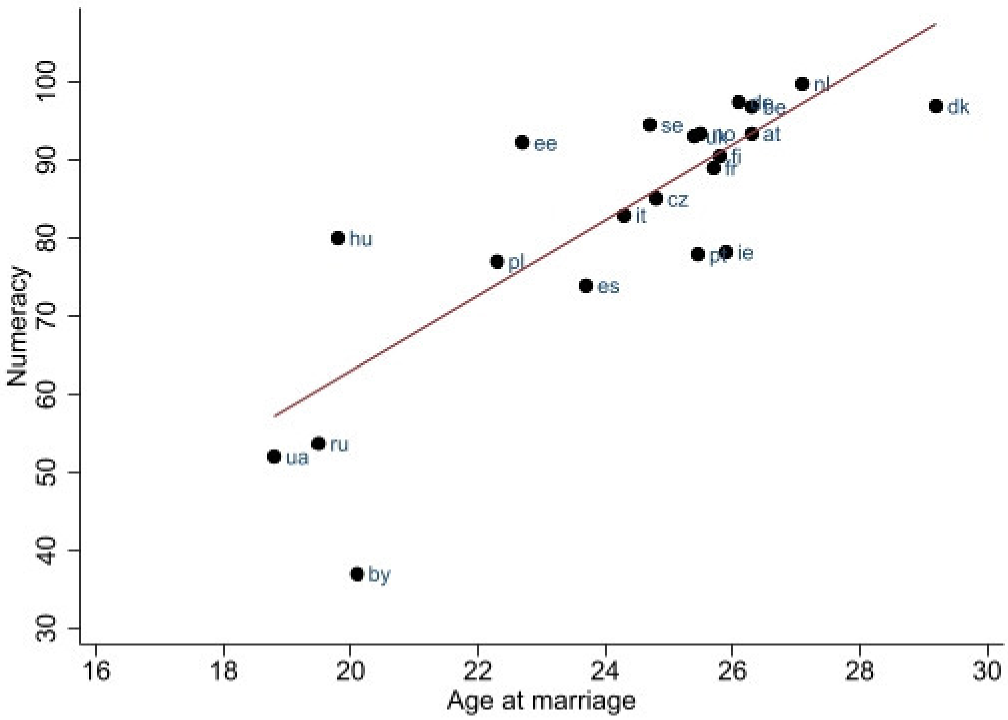
Panel (a): 1600-49 (numeracy) and 1600-99 (female age at marriage)



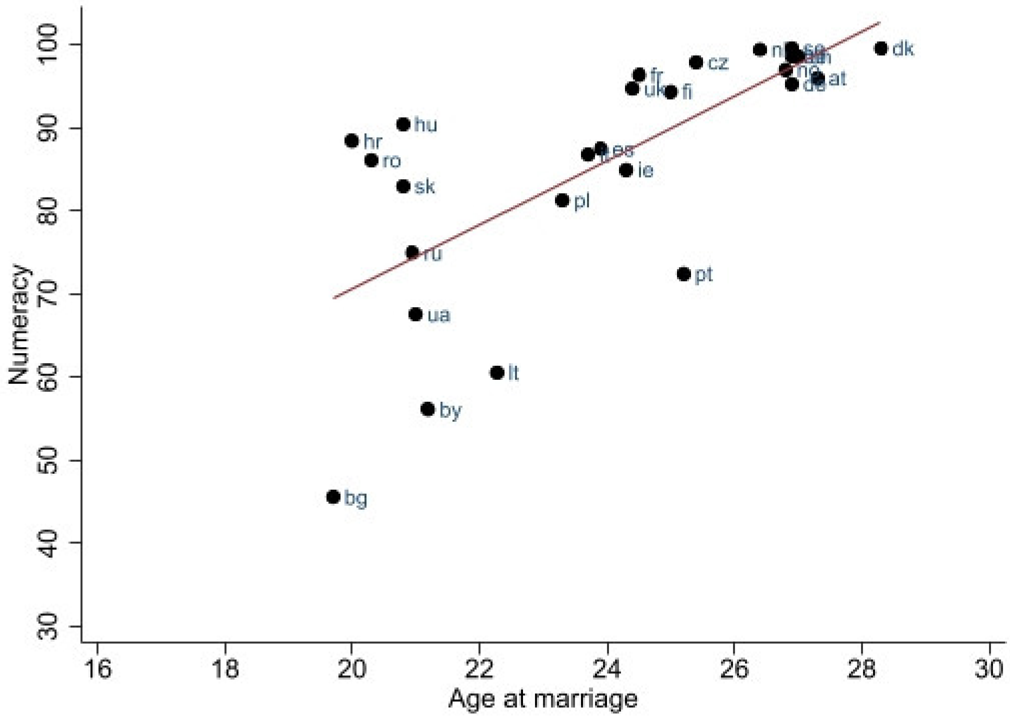
Panel (b): 1650-99 (numeracy) and 1600-99 (female age at marriage)



Panel (c): 1700-49 (numeracy) and 1700-99 (female age at marriage)



Panel (d): 1750-99 (numeracy) and 1700-99 (female age at marriage)



Panel (e): 1800-1849 (numeracy) and 1800-99 (female age at marriage)

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 4. Female age at marriage and numeracy: unconditional scatter diagrams.

Notes: The figure shows the relationship between female age at marriage by century and numeracy in the half centuries, by country. Correlation coefficients are 1600-49 (num) and 1600-99 (fem. Age at marr.): 0.6099, p=0.0463, n=11; 1700-49: 0.7885, p= 0.0000, n=20; 1800-49: 0.6972, p=0.0001, n=26. *Sources*: See text.

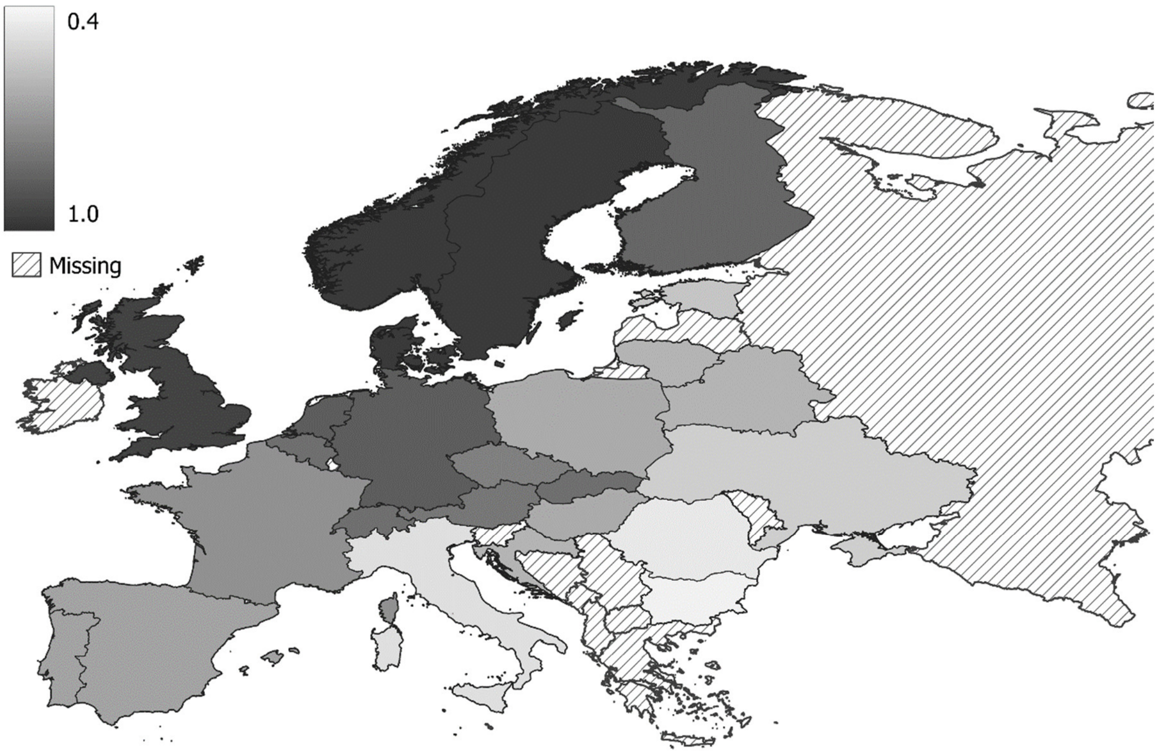


Figure 5. Lactose tolerance in Europe (darker: higher)
Notes and sources: See appendix F.

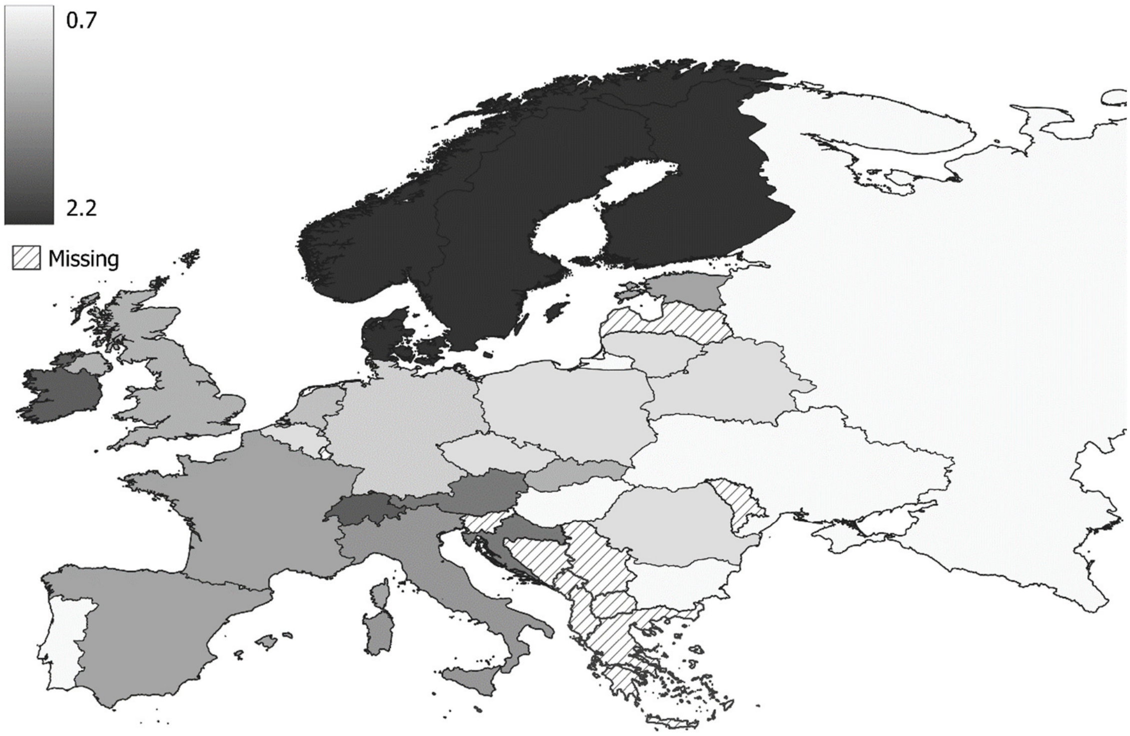


Figure 6. Pasture soil suitability, relative to grain soil suitability in Europe (darker: higher)
Notes and sources: See appendix F.

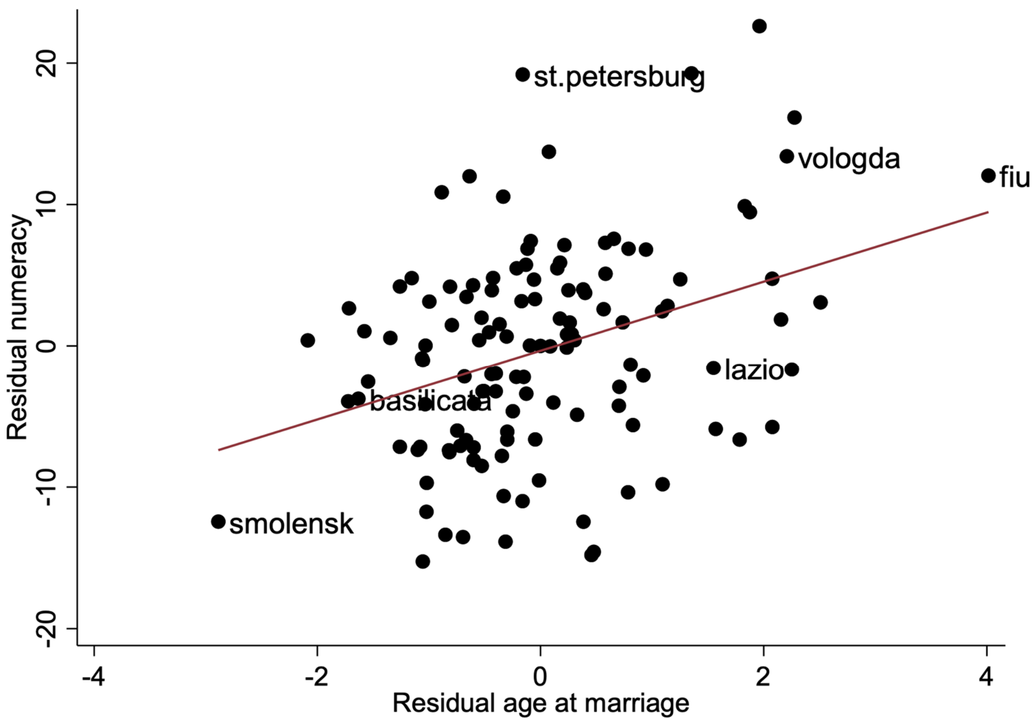


Figure 7. Partial scatter-plot of the regions
Notes and sources: see text. Conditional scattergram.

Variable	N	Mean	Standard deviation	Minimum	Maximum
Numeracy	88	78.67	17.32	36.00	99.75
Female 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
Parliaments	60	2.82	1.47	0	4.62
Lactose tolerance	88	0.75	0.15	0.43	0.96
Pasture suitability	88	1.32	0.46	0.70	2.16

Table 1. Descriptives of panel data set

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

Country	1500-49	1600-49	1650-99	1700-49	1750-99	1800-49
at				9946	2584	2819419
be		179	1060	28046	3214138	6898055
bg						659426
by				5381	12395	149622
ch						208887
cz		336	454	416	631	4076630
de	187	2494	1008	23779	53318	2141111
dk		6596	21229	7096	18366	1515977
ee					124	
es	1494	411	2637	12907	7512	6624016
fi				5090	1270	600
fr	256	701	13345	34015	19849	9024418
hr						670516
hu				14007	3577	2523902
ie				521	382	2456490
it	1577	8550	4076	20045	8350	7786423
lt						61295
nl		1049	1272	3699	1654	4818966

no				118220	329337	686311
pl				5704	4447	230911
pt		2839	2024	752	545	4881
ro						3333561
ru			5973	119	584	433719
se		7805	28912	5660	1917	2358645
sk						2317129
ua				313	3480	126938
uk	9035	2030	4271	66640	158236	1855199

Table 2: Numeracy data: the number of underlying observations.
Sources: See Appendix C. ISO-2-country abbreviations: see Figure 4

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	90	94
France	42	47	59	72	89	96
Germany	68	70	66	91	97	95
Hungary				80	80	90
Ireland				90	78	85
Italy	49	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	72
Romania						86

Russia			43	53	54	75
Slovakia						83
Spain	49	70	74	78	74	87
Sweden		77	79	83	94	100
Switzerland						99
Ukraine				47	52	68
United Kingdom	68	81	80	93	93	95

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

Country	1500-99	1600-99	1700-99	1800-99
at			26.3	27.3
be		26.4	26.3	26.9
bg				19.7
by			19.7	21.2
ch				27.0
cz		24.8	24.8	25.4
de	24.0	25.1	26.1	26.9
dk		28.2	29.2	28.3
ee			22.7	
es	20.0	22.0	23.7	23.9
fi			25.8	25.0
fr	20.8	22.0	25.7	24.5
hr				20.0
hu			19.8	20.8
ie			25.9	24.3
it	18.7	22.0	24.3	23.7
lt				22.3
nl		25.5	27.1	26.4

no			25.5	26.8
pl			22.3	23.3
pt		24.4	25.4	25.2
ro				20.3
ru	17.5		18.5	20.9
se		26.8	24.7	26.9
sk				20.8
ua			19.5	21.0
uk	24.8	25.7	25.4	24.4

Table 4. Female age at marriage estimates

Sources: See Appendix F. ISO-2-country abbreviations: see Figure 4. As these are based on published studies, they are averages of averages.

	1500-99	1600-99	1700-99	1800-99
at			8	10
be		10	52	26
bg				5
by			1	1
ch				2
cz		4	34	62
de	5	40	103	159
dk		1	5	12
ee			1	
es	1	1	5	4
fi			2	1
fr	3	1	13	37
hr				1
hu			41	45
ie			1	38
it	8	31	72	4
lt				1
nl		2	11	12

no			1	5
pl			7	2
pt		1	8	8
ro				3
ru			5	36
se		16	2	9
sk				2
ua			1	1
uk	3	66	110	48

Table 5: Female age at marriage data: the number of underlying studies.

Notes: Each study is typically based on several hundred to sometimes several thousand individuals. See also Data Appendix F. ISO-2-country abbreviations: see Figure 4. See our Appendix N, for an estimate of the underlying number of cases of Dennison and Ogilvie's female age at marriage studies.

	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 6. Correlation analysis of female autonomy indicators by Gruber and Szołtysek (2016).

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 (2016). *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 female 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).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
							The dependent variable is numeracy						
Age at marriage	4.511***	6.557***	4.378***	3.796***	4.555***	3.929***	4.151***	3.809***	3.653***	3.851***	4.051***	4.186***	4.565***
	(0.570)	(1.075)	(0.554)	(0.879)	(0.747)	(1.144)	(1.172)	(0.868)	(1.377)	(0.923)	(1.334)	(0.966)	(1.104)
War						-2.684					-2.510	-3.393	6.499
						(6.906)					(7.203)	(8.741)	(4.237)
Civil war							-3.285				-3.020	-9.047	7.128
							(8.601)				(9.686)	(19.07)	(4.886)
Serfdom								-0.270			0.0847	23.12***	-12.89**
								(2.391)			(2.890)	(7.533)	(6.007)
Latitude										0.419			
										(0.265)			
Longitude										-0.0896			
										(0.0945)			
First and last third	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	Yes	No	No	No	No	No	No	Yes	Yes
Source FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Century FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Country FE	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No

Constant	-30.1**	-83.9***	-43.1***	-23.7	-39.5**	-26.9	-34.0	-24.0	-64.1	-42.5	-29.9	-46.6***	-26.6
	(14.45)	(29.06)	(12.49)	(21.58)	(15.53)	(29.17)	(28.06)	(21.29)	(50.95)	(26.97)	(34.13)	(15.46)	(27.09)
Observations	91	91	91	91	84	85	85	91	71	85	85	41	44
R-squared	0.520	0.785	0.666	0.904	0.672	0.913	0.912	0.904	0.812	0.902	0.913	0.771	0.625

Table 7. 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. 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 and pasture soil suitability, and uses Region FE instead of country FE (regions are Eastern and Western Europe). In columns (6) to (8) we add control variables for war, civil war, serfdom. In column (9), we pre-adjust female age at marriage by first regressing it on national income, and using the residuals in the regression. Since GDP per capita is not given for all countries, the sample in column (9) is smaller. In column (11), latitude and longitude are added. In column (11), the three control variables war, civil war and serfdom are combined. In Column (12) and (13), we only use the half centuries until 1700 (incl.), and thereafter, respectively. We also calculated Wild bootstrapped p-values that we observe to be 0.000, indicating clearly statistical significance (Cameron et al. 2008). The variable female age at marriage is statistically significant in all specifications. We added a dummy variable for the cases in which a numeracy decade was located in the first third of the half century, and if a decade was located in the last third of the half century. *Sources*: see text.

	(1)	(2)
Second stage:	The dependent variable is numeracy	
Female age at marriage	6.378***	6.787***
	(1.357)	(1.467)
First and last third	Yes	Yes
Controls	No	Yes
Regional Fixed Effects	Yes	Yes
Source type	Yes	Yes
Century FE	Yes	Yes
Constant	-66.45***	-73.39***
	(22.04)	(25.74)
First stage:	The dependent variable is female age at marriage	
Lactose tolerance x pasture suitability	2.154***	2.111***
	(0.550)	(0.539)

Constant	14.91***	15.45***
	(1.024)	(1.101)
Observations	84	78
R-squared	0.683	0.740
F-stat	15.31	15.3%

Table 8. 2SLS regressions measuring the effect of female 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 also Wild bootstrapped p-values that we observe to be 0.000, indicating clearly statistical significance (Cameron et al. 2008, not shown). We use lactose tolerance interacted with pasture suitability as the instrument. As the instrument lactose tolerance is time-invariant, we cannot control for country FE, and use region FE instead. The reason for this would be perfect correlation between the time invariant instrument and the country fixed effects. 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. Please note that a time-varying variable can be instrumented by a time-invariant variable. This issue was discussed on the Stata discussion forum. The conclusion of all senior econometricians was that while the dynamic part of the panel variation is obviously not captured by this procedure, the cross-sectional variation is appropriately instrumented and hence the overall procedure is valid. Finally, we added a dummy variable for the cases in which a numeracy decade was located in the first third of the half century, and if a decade was located in the last third of the half century. *Sources*: see text.

	N	Mean	Standard Deviation	Min.	Max
Numeracy	184	76.92	17.07	21.38	100.51
Female 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	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 9. Descriptives of variables used for European regions during the 19th century

Notes: We report descriptives for numeracy and other variables for 184 European regions during the 19th century. However, the evidence on female age at marriage sometimes requires that we assign the same value to two regions, as female 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 female age at marriage and numeracy values. *Sources*: see text.

	(1)	(2)	(3)
		The dependent variable is numeracy	
Female 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	No	Yes	Yes
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 10. 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. We estimate OLS to assess the conditional correlation between numeracy (which is the dependent variable) in column (1) using female 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 female age at marriage by adding two additional explanatory variables (area share of large landowners and Protestantism). *Sources*: see text.

	(1)	(2)	(3)	(4)	(5)
Country included:	Bulgar. & Hungary	Serbia	Spain & Italy	Russia	UK
Dependent variable:	Numeracy	Numeracy	Numeracy	Numeracy	Literacy
Female age at marriage	2.695**	3.838**	0.971**	2.015*	6.310***
	(1.142)	(1.776)	(0.463)	(1.104)	(2.228)
Constant	27.42	3.066	69.96***	20.69	-103.6*
	(25.84)	(33.32)	(11.40)	(22.93)	(57.22)
Observations	63	15	30	76	37
R-squared	0.077	0.200	0.114	0.058	0.202

Table 11. 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. We report OLS estimates of female 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. *Sources*: see text.

	(1)	(2)	(3)	(4)
Second stage:	The dependent variable is numeracy			
Female age at marriage	7.209***	7.358***	7.257***	7.345***
	(2.681)	(1.409)	(1.322)	(2.311)
Share large landowners	-26.48***	-61.60***	-63.31***	-22.16**
	(8.996)	(13.00)	(11.80)	(9.173)
Heights		-2.408*	-2.741**	-0.145
		(1.464)	(1.171)	(0.964)
Protestantism	0.287**	-0.0774		0.143
	(0.141)	(0.115)		(0.0975)
Country FE	No	No	No	Yes
Constant	-70.98	324.7	381.4**	-89.46
	(57.81)	(248.3)	(190.8)	(150.3)
First stage:	The dependent variable is female age at marriage			
Pasture suitability	11.32***	17.73***	19.95***	6.615**

	(1)	(2)	(3)	(4)
	(3.084)	(4.257)	(4.669)	(2.916)
Constant	19.52***	44.74***	-9.763	21.54***
	(0.619)	(9.967)	(6.549)	(7.215)
Observations	156	89	89	89
R-squared	0.563	0.724	0.641	0.888
F-stat	15.14	18.72	20.26	12.07

Table 12. 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. We report the IV regressions of the effect of regional female 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 assess the validity of the exclusion restriction, the control variable "share of large owners" is included in all 4 Columns, "Protestantism" in Column (1), (2) and (4). 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, 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. For columns 2-4 in table 11, only 57 regions could be included because the combination of regional heights data and the other variables was only available for these regions that were situated in Italy, Spain and Russia (in contrast, for Hungary and the whole Habsburg Empire, such disaggregated height data is not available, nor for the smaller Southeastern European countries). In Column 1, 119 observations are left because of missing values in the three right-hand side variables protestant, land inequality and the instrument soil quality. *Sources*: see text.

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 13. 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: female 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 female 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 female age at marriage, relative to the joint effect of the observables, to account for the full size of the female 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

Female age at marriage	3.426
Spatial std. error, 250 km	(0.080)***
Spatial std. error, 500 km	(0.096)***
Spatial std. error, 750 km	(0.100)***
Observations	91
Controls included	YES

Panel B: for the regional data set

Female age at marriage	3.926
Spatial std. error, 250 km	(0.103)***
Spatial std. error, 500 km	(0.103)***
Spatial std. error, 750 km	(0.103)***
Observations	120
Controls included	YES

Table 14: Assessing spatial autocorrelation

Notes: *** indicates significance at the 1% level; ** at the 5% level; and * at the 10% level. Standard errors are corrected for cross-sectional spatial dependence and (in panel A) for panel-serial correlation. In panel A, we regress numeracy on female 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 spatially adjusted standard errors using the distance of 250, 500 and 750 km (Hsiang 2010). In panel B, we regress numeracy on female age at marriage, land inequality, protestant share and serfdom.

(Online) appendices

Appendix A: 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.

In our paper 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.

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)

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

Apart from reasons of data availability, 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 et al. (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 economics that was used by hundreds of studies (see Tollnek and Baten 2017 and Baten et a. 2022 for reviews).

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.

Appendix B: Representativeness of our sample for the countries of Europe

To study the representativity of our sample, we carefully analyzed our sample of 27 countries relative to all sizeable European countries. We only lack six countries among the European countries that were situated to the west of the Uralian mountains and the Caucasus and had 500.000 or more inhabitants in 1990 (see notes to Table B1). To assess the representativeness, we compared the urbanization ratio in 1850 of our 27 sample countries with the six countries on which no data is available (The urbanization ratio is often taken as an indicator of early development level). Among these 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 (see Table B1). 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

Table B1. Urbanization and GDP per capita in Europe.

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 rank correlation of the population size over time, except for France whose population size ranked declined slightly in the 19th century, see clio-infra.eu). Countries not included in main sample, but listed here, are Albania, Greece, Latvia, Moldova, North Macedonia and Slovenia.

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

All estimates and their sources are reported on 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 half century for which the estimate was reported, and the abbreviated source). At the end of this document, abbreviated source are referenced. One technical paper with a long list of archival sources is available (Baten 2021, see https://uni-tuebingen.de/en/18732).

Census data:

at	1700	Tollnek and Baten (2017)
at	1750	Tollnek and Baten (2017)

be	1600	Baten (2021)
be	1650	Baten (2021)
bg	1800	Baten and Hippe (2018)
by	1700	Baten et al. (2017)
by	1750	Baten et al. (2017)
cs	1700	Baten (2021)
cs	1750	Baten (2021)
cz	1800	Baten and Hippe (2018)
de	1600	Baten (2021)
de	1650	Tollnek and Baten (2017)
de	1700	Tollnek and Baten (2017)
de	1750	Tollnek and Baten (2017)
dk	1700	Tollnek and Baten (2017)
dk	1750	Tollnek and Baten (2017)
ee	1750	Baten et al. (2017)
es	1650	Tollnek and Baten (2017)
es	1700	Tollnek and Baten (2017)
hr	1800	Baten (2021)
hu	1600	Baten (2021)
hu	1650	Baten (2021)
hu	1800	Baten and Hippe (2018)
it	1650	Tollnek and Baten (2017)
it	1700	Tollnek and Baten (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	Baten (2021)
ro	1650	Baten (2021)
ro	1700	Baten (2021)
ro	1750	Baten (2021)
ru	1600	Baten et al. (2017)
si	1800	Baten and Hippe (2018)
sk	1600	Baten (2021)
sk	1650	Baten (2021)
ua	1650	Baten (2021)

ua	1700	Baten et al. (2017)
ua	1750	Baten et al. (2017)

Court records (women and men accused as witches)

de	1500	Baten (2021)
fr	1500	Baten (2021)
uk	1500	Baten (2021)

Death register data

cz	1600	Familysearch.org, last accessed Nov 4th, 2021
cz	1650	Familysearch.org, last accessed Nov 4th, 2021
dk	1600	Familysearch.org, last accessed Nov 4th, 2021
dk	1650	Familysearch.org, last accessed Nov 4th, 2021
es	1750	Familysearch.org, last accessed Nov 4th, 2021
fr	1600	Baten (2021)
fr	1650	Familysearch.org, last accessed Nov 4th, 2021
ie	1650	Familysearch.org, last accessed Nov 4th, 2021
it	1500	Baten (2021)
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.org, last accessed Nov 4th, 2021
se	1600	Familysearch.org, last accessed Nov 4th, 2021
se	1600	Familysearch.org, last accessed Nov 4th, 2021
se	1700	Familysearch.org, last accessed Nov 4th, 2021
se	1750	Familysearch.org, last accessed Nov 4th, 2021
uk	1650	Familysearch.org, last accessed Nov 4th, 2021

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 and men 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, they might have been less numerate than women who were not accused of witchcraft, as our source fixed effects show negative values. In the regressions we control for this 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 (see appendix G).

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 (we used 50 : 50 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 three 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.

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 three witchcraft court records and the 18 death record based observations need a special treatment (described above).

Appendix D: Strategies used to achieve a high degree of regional representativity of numeracy data

Regional composition could potentially create a bias, because we do not have surviving sources for all cities and villages. To come to terms with regional selectivity issues, we have employed a specialized methodology to which we will now briefly turn. We do this on the basis of country examples:

Germany, Austria, Switzerland, Southern Italy, and Spain (18th C):

Perfect representativeness is probably impossible for historical samples. But we need to consider whether the dataset has systematic biases that could distort the results in a significant way, and we should seek to minimize potential measurement errors. The unequal availability of sources naturally leads to a stronger focus on certain regions in some of the countries. Tollnek and Baten (2017) assessed the representativeness of data on Germany, Austria, Switzerland, Southern Italy, and Spain during the 18th C. They did not observe major selectivities between regions, but they did find a certain oversampling of urban places. Hence, their estimates required appropriate weights that would reduce the impact of urban and increase the impact of rural places on the final estimate. one motivation was to make sure that the urban and rural sample components are appropriately weighted. We used their strategy (and estimation output) in our study accordingly. In doing so, the samples are weighted to represent both urban and rural parts of the countries in a representative way.

Eastern and East-Central Europe:

Baten, Szołtysek and Campestrini (2017) have employed an adjustment strategy in their research on human capital development in central-eastern and eastern Europe, looking at five countries in the 17th to 19th century. To control for regional selectivity, they have studied deviations from the national mean. For example, for Russia, five regions could be documented for different periods, and the inclusion of Moscow in the five documented regions suggests a possible upward bias. To find out about representativity of the regions, Baten et al. (2017) compared regional ABCC values with national averages found for Russia for the 1820s. For that specific period, the analysis shows that four out of five regions had ABCC values that were 15-25 percent above the Russian average, which implies a bias. Since they know the national average for the 1820s, they could adjust the values for Moscow and the other places accordingly. For instance, if Moscow had a numeracy rate of 83, which was 19 percent above the national level in the 1820s, they adjusted the earlier values for Moscow proportionally downwards to increase representativity of the data. A potential remaining issue is that the bias may have changed over time. A robustness-check performed by Baten et al. (2017) was to scale the deviation by the overall level of numeracy instead of using constant differences derived from a given year. The benefit of this second procedure is that it considers the national trends of the countries and that it makes the adjustments sensitive to possible differences in trends. The results of this second approach were very similar to the first one, thereby confirming expectations.

Portugal and Spain during the 16th and 17th centuries:

Juif, Baten and Péres-Artes (2020) studied the regional representativeness of Spanish and Portuguese mortality and inquisition data, mostly for the 16th and 17th centuries. For Spain and Portugal they have verified that the data are representative by comparing the regional coverage of the samples with the regional distributions of the actual populations in the censuses of the 18th and 19th century. In doing so, they asked how regionally representative the dataset is for the whole population. The Spanish source is quite representative by region: the larger units of the country have similar population shares in the sample and in the census. The same is true for Portugal.

Taking these three strategies together ((1) urban/rural weights, (2) adjustment of smaller regional samples for regional bias using later census evidence for the whole country, and (3) a systematic analysis of regional shares for large samples), we can reduce potential regional unrepresentativeness to become a problem that is not substantial.

Appendix E: Does female age at marriage correlate with other gender equality indicators?

In this paper we use female age at marriage as an indicator for female autonomy. The rationale behind this is that women who married young were usually not able to collect much experience on the labor market (they could not gain basic organizational skills, for example) and pass that on to their off-spring. Hence it might be possible to use the female age at marriage as an indicator for female autonomy. In Table 6 in the main text we have already shown that female age at marriage correlates with other indicators of female autonomy during the early modern period (based on the study by Gruber and Szoltysek 2016). In this appendix, we provide additional details of our assessment using two different alternative sources of gender equality that allow to compare the corresponding female age at marriage values.

Can we measure female status relative to male status in early modern Europe using female age at marriage? On a global scale, this indicator seems justified considering more recent evidence in Subsaharan Africa during the 20th century, for example. Female age at marriage was the highest in Southern Africa and the lowest in the Western Sahel zone, reflecting similar educational patterns -- for example, if we consider the relative education for females and males (Baten et al. 2021). However, one could imagine that within Europe the overall variation of female age at marriage might have been low, or perhaps the correlation with relative female status might not be strong. How can we assess this? One possibility is to use a recent data set prepared by Hippe and Perrin (2017) which reflects the gender equality of literacy around 1900 in the regions on the territory of 20 European countries of today (for the results: see Figure 2 in main text).

We calculated the gender inequality of literacy as the difference in literacy between males and females, divided by the level of literacy of males. For example, in Central Serbia with its very low FAM, the gender gap represents 79.2% of male literacy. [19] In contrast, in Lithuania, a region influenced by neighboring Scandinavian in its high FAM, but governed by the Russian Empire that invested low amounts in schooling, the FAM was almost 25, the gender inequality of literacy relative to the male level only 17.6%.[20] Using the simple gender gap, or dividing by the average level of both genders, yields qualitatively very similar results.

Of course, we need to take into account which other factors might have determined literacy by gender and total literacy around 1900. For example, the Habsburg school policy aimed at providing schooling in Non-German speaking regions of the Empire in order to dampen Anti-Habsburg (or Anti-German) political views. This reduced -- as a side effect -- the "room" for gender inequalities of literacy (because both females and males had literacy not far from 100 percent). Almost all observations to the lower left of the regression line in Figure 2 are therefore Non-German speaking regions of the Habsburg Empire. Similar "artificial compression" of gender inequality took place in the large city of Paris where almost all females were literate (or reported to be). On the other hand, the Russian Empire invested not as much in its rural regions. All the regions to the upper right of the regression line were Russian regions that were specialized in dairy farming (such as Vologda) which resulted in a notable contribution of women to family income, resulting in a relatively high FAM and relatively high female autonomy, but not as much female literacy. The correlation is -0.73 (p=0.000, N=111).

A second additional possibility is to study the correlation between FAM and the relative height of females relative to males. Schwekendiek and Baten (2019) have systematically assessed the relative height of females – compared to males – for 20th century populations and established this as an indicator of gender equality of nutrition and health. Similarly for the medieval period, Maravall and Baten (2019) have studied the gender equality of nutrition and the relative height of females using archaeological samples from early medieval Europe. They found a strong correlation between the relative height and the relative occurrence of enamel hypoplasia which is a completely independent indicator for relative malnutrition of girls. The correlation with relative height suggests that both are informative about gender equality. We calculate the relative female height based on archaeological femur lengths for 28 European samples based on 100-300 underlying skeletons on individuals (Hence, altogether about 3000 individual heights could be included). This study is based on the European History of Health Dataset (Steckel et al. 2019) and the dataset collected by Koepke and Baten (2005, 2008). We assigned the FAM by country to these heights samples. We have to admit that the combination of this archaeological sample with country-specific FAM values results in a substantial potential for measurement error. Nevertheless, as Figure E2 shows, we observe a substantial correlation between the two indicators. This gives some tentative support to the idea that both indicators (FAM and female height) reflect gender equality during the early modern and medieval period.

In sum, therefore, we conclude that we observe a close correlation between FAM and other indicators of gender equality and female autonomy. In all robustness-checks (Table 6, and Figure E1), we observe a correlation between FAM and various indicators of gender equality, which underlines the informativeness of FAM for the study of gender equality in early periods in Europe.

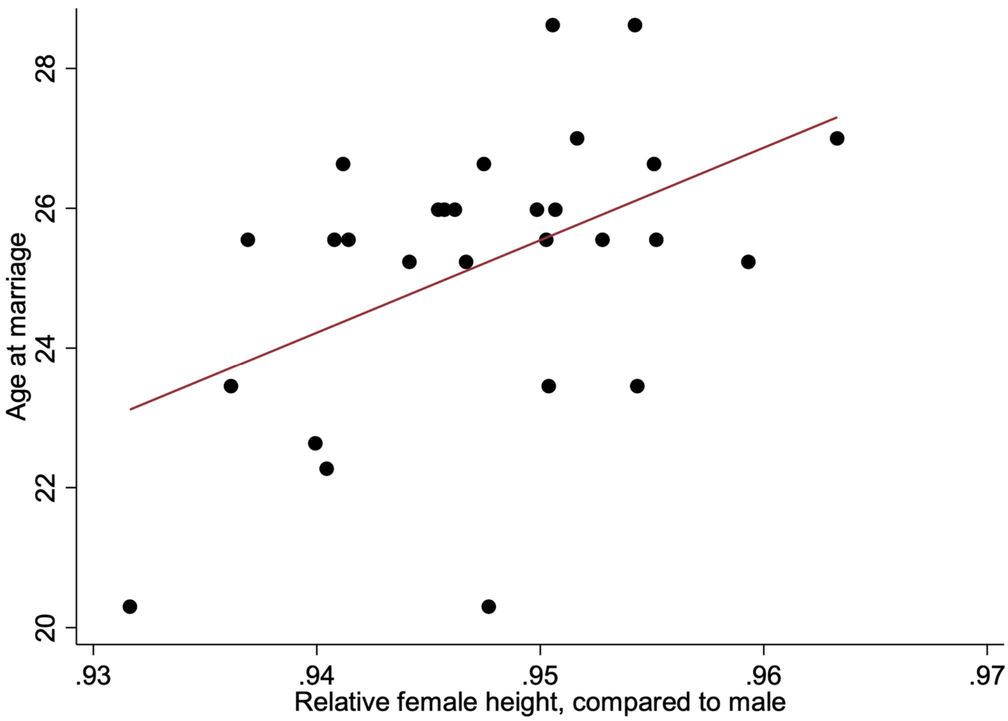


Figure E1 The correlation between relative female height and age at marriage.

Sources: see text.

Appendix F: Sources

This appendix lists all sources that are used for the empirical analysis as conducted in sections 4, 5 and 6 of our paper.

Panel dataset

Female age at first marriage: The data on women's age at first marriage is mostly derived from the study by Dennison and Ogilvie (2013, see their Table 1). The observations are based on a metastudy of 175 publications in historical demography (listed in their Appendix A). A final version of the paper was published as Dennison and Ogilvie (2014, hence the 2013 title is not "grey" literature). It should be mentioned here that Dennison and Ogilvie report the number of studies and not the underlying number of observations in their Table 1, which we used in our analysis for Table 7. We have collected a sample based on their original data to find out how many underlying observations each study combines. Their studies typically include between 500 and 1000 individuals whose marriage data was used to calculate the female age at first marriage, sometimes several thousand individuals are underlying the estimates (see Appendix N). They compute averages of averages, as the original ages were average to the place-century level, and this average was averaged to report country-century level data.

For Italy in the seventeenth century we average the north and south using population weights. For the sixteenth century, as only evidence on the south is available, we assume the same north-south-difference as in the seventeenth century and calculate an adjusted value. The same procedures are followed for France. In the sixteenth century we take the average of the central and north. No downward adjustment for the south is necessary, as obviously during the seventeenth century the south had even higher female age at marriage values than central France.

While we use mainly Dennison and Ogilvie's estimates for female age at marriage, for the Russian empire we used the Princeton estimates of age at marriage in order to estimate the values for today's Belarus and Lithuania (Coale and Watkins 1986, Coale et al. 1976; see sources for our regional regressions below). For Portugal we added evidence from Botao Rego et al. (2016), which provide us data from the seventeenth to the nineteenth century on the southern regions of Beira and Evora, the central region of Lisbon and the northern regions ranging from Viana do Castelo in the northwest to Bragança in the northeast end Viseu and Aveiro in the southern part of the north. In sum, they report data on seven regions of the north and ten regions all together.

Numeracy: For numeracy, please see appendix C.

War and civil war: The data on interstate wars and civil and religious wars have been processed by Baten and Mumme (2013), originally coming from the Polity IV project. www.polityIV.org, last accessed Nov. 4th 2021.

Second Serfdom: To control for the effects stemming from the second serfdom, we have included a dummy identifying those countries that had more than 30 per cent serfs with severe labour obligations. The data has been collected by Baten et al. (2017) based on various statistical and historical sources.

Population density: Clio-infra.eu, last accessed Nov 1st, 2021.

Atlantic trade: The log of the volume of Atlantic Trade has been derived from Acemoglu et al. (2005, Appendix I). It concerns annual average voyages equivalent for ships of 400 deadweight tons.

Merchant fleet: Estimates of the growth of the European merchant fleet between 1500 and 1800 are taken from de Pleijt and van Zanden (2016, their Table 1). It concerns the log of the per capita size of the fleet in thousand ton units.

Parliaments: We have used the activity index of the various Parliaments (defined as the number of years they were in session during a century) as a proxy for political institutions. The data was originally collected by van Zanden et al. (2012) and varies from 0 (for absolutist states) to close to 100 for post-Glorious Revolution England and the Dutch Republic.

Lactose tolerance: As explained in the text, the sources are Ingram et al. (2009), Flatz (1995), Baten and Blum (2014), Cook (2014)

Relative pasture suitability: 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 dataset

Female age at first marriage: Our source for female 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 et al. 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 F1). This strategy worked for most countries, but not for Italy (earliest female age at marriage was 1911). However, for Italy we could study later educational data, and the same correlation with female age at marriage was observable. While the strategy of obtaining the explanatory variable 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; similarly for Spain, etc.). Similarly, the interregional differences of numeracy are highly stable between census years (Baten and Hippe 2018, their Appendix*)*.

Year	**Country**	**Age at marr.**	**N**
1905	Bulgaria	20.7	4
1900	Serbia	18.8	15
1887	Spain	23.9	14
1880	Hungary (Habsb.East)	21.9	59
1911	Italy	23.6	16
1897	Russ. Empire	20.9	76

Table F1. Evidence on female age at marriage for the regions

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

Numeracy: For numeracy evidence on European regions, Baten and Hippe (2018) provide a 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 (Baten and Hippe 2018).

Land inequality and protestantism: We also control for land inequality, which has been estimated by Baten and Hippe (2018), as well as 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, Baten and Hippe (2018) 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.

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

Relative pasture suitability: 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).

Height: Regional height is provided in a number of studies, such Martínez-Carrión et al. (2016) on Spain in 1858; Baten, Szoltysek and Campestrini (2017) provide height estimates for the regions of the Russian Empire. A'Hearn at al.(2009) do the same for Italy.

Appendix G: Robustness tests

In this appendix 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), because they might be endogenous ("bad controls", or rather "proxy controls", see Appendix K).

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).

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 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 was convened to close to 100 for England after the Glorious Revolution of 1688 (van Zanden et al. 2012).

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 G1 we observe that none of these potential confounders makes the effect of female 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. In all regressions the coefficient on female 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 G2 we exclude the United Kingdom (Col.1) or the Netherlands (Col.2).[21] Columns (3) and (4) repeat the exercise excluding Russia and Denmark, which are other potential extreme cases in our sample. Russia was much poorer than the other countries in our sample, numeracy was relatively low by international standards, and women married very early. Similarly, Denmark had one of the highest values of numeracy and female age at marriage in the early modern period.

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 section 4 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. (5) in Table G2) and from "witch trials" (Col. (6)). All the coefficients of female age at marriage are still statistically significant and the coefficient changes very little.

Next, 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 sufficiently after the 16th century. We therefore perform an additional robustness-check by including dummy variables for mostly Protestant and Orthodox countries. As Table G3 shows, controlling for religion does not affect the significance of our female age at marriage variable.

We present reduced form regressions in Table G4. If the IVs would be insignificant at this stage already and if we would not see an effect in the reduced form, it would probably not be there (Angrist and Krueger, 2001). Hence it is good to know that we do observe it.

Finally, we also present IV regressions using both IV components of our interacted instrument separately. Lactose tolerance is a very strong instrument by itself and the results are consistent with the interacted IV. In contrast soil suitability is a weak instrument, but the coefficients for female age at marriage are similar to the ones using a combined IV, or using only lactose tolerance: the coefficient is large and statistically significant, using soil suitability alone. We keep the soil suitability in our interaction term because it makes the choice of the instrument consistent with the regional IVs.

	(1)	(2)	(3)	(4)
Second stage:	The dependent variable is numeracy			
Female age at marriage	5.192***	3.970***	4.615***	6.476***
	(0.947)	(0.854)	(1.001)	(1.131)
Log Atlantic trade	4.458***			
	(1.024)			
Log size of merchant fleet		0.0454		
		(0.214)		
Log parliamentary activity			1.893*	
			(0.968)	

Log population density				-0.477
				(1.358)
Controls	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Source type	Yes	Yes	Yes	Yes
Century FE	Yes	Yes	Yes	Yes
Constant	-38.37**	-25.71	-39.99**	-70.49***
	(17.66)	(16.74)	(19.75)	(22.36)
First stage:	The dependent variable is female age at marriage			
Lactose tolerance x pasture suitability	2.654***	2.671***	2.613***	3.905***
	(0.543)	(0.482)	(0.513)	(0.488)
Constant	16.34***	16.42***	16.29***	20.64***

	(1.166)	(1.341)	(1.261)	(1.526)
Observations	54	60	54	78
R-squared	0.750	0.792	0.841	0.658
F-stat	23.90	30.67	25.97	63.99

Table G1. 2SLS regressions including proxy 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 instrument lactose tolerance is time-invariant, we cannot control for country FE, and use region FE instead. In addition to the specifications explained in Table 5, we use the log of the volume of Atlantic trade of Acemoglu et al. (2005) as control variable in column (1); we add the log of the (per capita) tonnage size of the merchant fleet in column (2). We add the log of the activity index of European Parliaments of van Zanden et al. (2012) in column (3) to test for the effect of political institutions; finally, log population density in Column (4), as this reflects the demographic behaviour to a certain extent. We added a dummy variable for the cases in which a numeracy decade was located in the first third of the half century, and if a decade was located in the last third of the half century. We do not include all confounding variable together, because if we would, too few observations would be left. *Sources*: see text.

	(1)	(2)	(3)	(4)	(5)	(6)
Excluded:	UK	NL	DK	RU	Witch	Death
Second stage:	The dependent variable is numeracy					
Female age at marriage	6.678***	6.902***	9.284***	6.787***	6.787***	5.683***
	(1.371)	(1.446)	(2.712)	(1.467)	(1.467)	(1.455)
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Source type	Yes	Yes	Yes	Yes	Yes	Yes
Century FE	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-66.95***	-75.65***	-118.1***	-73.39***	-73.39***	-58.96**
	(23.38)	(25.76)	(45.43)	(25.74)	(25.74)	(24.70)
First stage:	The dependent variable is female age at marriage					
Lactose tolerance x pasture suitability	2.127***	2.187***	1.503***	2.111***	2.061***	2.111***
	(0.537)	(0.549)	(0.478)	(0.539)	(0.679)	(0.535)
Constant	15.03***	15.83***	16.03***	15.45***	15.26***	15.45***
	(1.218)	(1.117)	(1.124)	(1.101)	(1.141)	(1.094)
Observations	72	73	73	78	77	67
R-squared	0.658	0.630	0.560	0.648	0.625	0.700
F-stat	15.67	15.86	9.89	15.34	15.54	9.22

Table G2. 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. In the table above we excluded several groups in order to make sure the inclusion of these in the baseline results did not 5divert 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. In column (4) we exclude Denmark as it might be another extreme case with high levels of numeracy and female age at marriage. We have re-estimated the regressions by omitting death records in column (5), witch trials in column (6). We added a dummy variable for the cases in which a numeracy decade was located in the first third of the half century, and if a decade was located in the last third of the half century. *Sources*: see text.

	(1)	(2)
Female age at marriage	3.498***	2.599***
	(0.766)	(0.600)
Protestant	6.456**	7.077**
	(3.031)	(2.769)
Orthodox	-21.02***	-21.55***
	(5.735)	(5.354)
Time FE	N	Y
Region FE	Y	Y
Source FE	Y	Y
	-6.643	0.755
	(14.96)	(11.76)
Observations	91	91

R-squared	0.712	0.789

Table G3. 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. 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. *Sources*: see text.

Lactose tolerance x pasture	14.33***
suitability	(3.095)
Controls	Y
First and last third	Y
Region FE	Y
Source type	Y
Century FE	Y
Constant	31.46**
	(11.98)
Observations	78
R-squared	0.585

Table G4. Reduced form Regression

Notes: for the description of the specification, see Table 5, Col. 2.

	(1)	(2)
Second stage:	The dependent variable is numeracy	
Female age at marriage	6.425***	8.528***
	(1.357)	(1.467)
First and last third	Yes	Yes
Controls	Yes	Yes
Regional Fixed Effects	Yes	Yes
Source type	Yes	Yes
Century FE	Yes	Yes
Constant	-67.79***	-101.0***
	(19.04)	(51.31)
First stage:	The dependent variable is female age at marriage	

	(1)	(2)
Lactose tolerance	8.865***	
	(1.675)	
Pasture suitability		1.546**
		(0.719)
	(1.024)	(1.101)
Observations	78	85
R-squared	0.741	0.701
F-stat	27.97	4.63

Table G5. 2SLS regressions measuring the effect of female age at marriage on numeracy, using lactose tolerance and pasture suitability separately.

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 instrument lactose tolerance is time-invariant, we cannot control for country FE, and use region FE instead. We use lactose tolerance interacted and pasture suitability as the two instruments in separate regressions. 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). We include all the control variables of Table 7, Col. 10: war, civil war and serfdom. Finally, we added a dummy variable for the cases in which a numeracy decade was located in the first third of the half century, and if a decade was located in the last third of the half century. *Sources*: see text.

Appendix H: Can we obtain additional evidence from Murdock's (1976) data collection?

Murdock's (1967) evidence, recently processed by Nathan Nunn, has been used for African, American and Asian ethnic groups. Consequently, a look on his European data is a very sensible idea. However, for Europe, his research strategy needs to be taken into account.

Murdock did not collect data on Europe in a representative way, as he aimed (a) to provide data for "pre-industrial" societies and (b) he compiled many ethnological studies that happened to be available. These were often interested in special societies about which otherwise not so much was known (such as the Saami).

Scandinavian evidence from Murdock refers mostly to Sami populations (Lule Sami, North Saami, Pite Saame, South Saami, Inari Saami etc.), Karelians, Faroese, and early Icelanders. "Sweden" refers to the territory, not to the society or ethnic group of the Swedish. Murdock has no data on the Swedish, Norwegian, Finnish or Denish societies. "Icelanders" refers to "Early Icelanders" (focal year 1100 CE). Only for this society, respondents mentioned "Males" in the question about dairy farming (they might have thought about managing the herds that were kept for meat production (slaughtering) on the pastures of Iceland, rather than the milking work on the farm).

Appendix I: Culture and lactose tolerance

We use lactose tolerance as an instrument. How exogenous is this factor? Did a gender equality culture perhaps exist in the regions of later lactose tolerance, or might lactose tolerance be the result of a coevolution of culture and genes which might have taken place several millennia before the early modern period? Hansen et al.'s (2015) discussion of hunter-gatherer society evidence (based on the last two centuries) and some small-sample archaeological evidence suggest that the possibility might exist. More solid archaeological evidence would be needed to make the case that this played a role, but we would like to mention the potential caveat. Fortunately, we arrive at very similar results using the pasture suitability evidence in our panel and regional analysis. This consistency of pasture suitability and lactose tolerance and their functioning as instrumental variables increase the likelihood that both are exogenous, but the caveat remains to be noted.

Are changes of lactose tolerance within the 350 years covered by our panel data sample likely? The most dramatic changes towards the lactose tolerance took place – according the present status of our knowledge -- between 6000 and 4000 years before present, while for the last 350 years, we can be relatively sure that the genetic variation of lactose tolerance did not change between regions and over time (Rosenstock et al. 2015).

For our regional approach in the latter part of the paper – for which we use pasture suitability as an instrument – one might also wonder whether a "coevolution of genes and culture" might have stimulated dairy farming agricultural production, but it seems not likely that exactly in the Alpine regions (incl. Northeastern Italy and mountainous Slovenia), the Czech lands, Northeast and Southeast Russia, and Northern Spain, gender-egalitarian culture-gene clusters developed millennia ago that had a stronger influence than the soil suitability for dairy farming in the very same regions. Whether or not this gene-cultural clusters developed, the differences also in the 19th century seem to have been clearly predetermined, which is relevant for our regional analysis. Nevertheless, we admit the possibility of this "coevolution of genes and culture" that might be studied more carefully in the future.

Appendix J: Church exposure

Schultz et al. (2019) recently studied the role of church exposure in various regions of the world on behavioral patterns. They argue that the policy of the Western Church (later: the Catholic Church) aimed at reducing cousin marriages and changing other marriage practices during the medieval period. These marriage behaviors resulted in intensive kinship in other regions, but also in the regions of the Western Church before this program was introduced. The specific marriage and family planning program changed this and resulted in more individualistic behavior, and also in more social behavior and trust towards foreigners, while the other world populations concentrated their activities on their own kinship groups. This might also have resulted in a higher female status in Northwest Europe and hence in a higher numeracy. Hence, we need to control for Church exposure in our regressions.

We have investigated the relationships very carefully and found that while Schultz et al. observe a strong effect on the psychological values today, surprisingly the correlation between church exposure and human capital formation does not hold if female age at marriage is included (see results in table J1). This might be caused by the fact that within Europe, Scandinavian countries have a very modest church exposure, because missionary activity took place very late in Scandinavia. On the other hand, human capital values are very high due to the female autonomy that we observe in Scandinavia (Figure J1). Similarly, the kinship intensity – according to the data in Schultz's et al.'s replication files -- was quite homogeneously low in Europe, except for two exceptions: Hungary and Finland (Figure J2). It seems unlikely that these two – economically very heterogenous, and not large – countries determined overall European development. Hence, it seems that female age at marriage (and the female autonomy it proxies) seems to be the driving force of early numeracy formation, not the Western Church's family program against cousin marriage and related family-specific behavior.

	(1)	(2)	(3)	(4)
	The dependent variable is numeracy			
Female age at marriage	3.089***		3.851***	
	(0.936)		(0.923)	
Cousin mariage	-2.694**			
	(1.077)			

Church exposure		1.429***	-2.041**	
		(0.179)	(0.861)	
Kinship intensity				-6.097***
				(0.977)
Controls	Yes	Yes	Yes	Yes
Source type	Yes	Yes	Yes	Yes
Century FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Constant	3.591	50.86***	-8.653	55.35***
	(16.90)	(5.703)	(16.12)	(5.646)
Observations	56	85	85	85
R-squared	0.840	0.887	0.902	0.887

Table J1. Controlling for Church Exposure.

Notes: The effects of Church Exposure and Cousin Marriage alone are visible (Col. 1 and 2), but not if female age at marriage is controlled for (Col 3 and 4). *Controls, source type etc*: see table 7. *Sources*: See text.

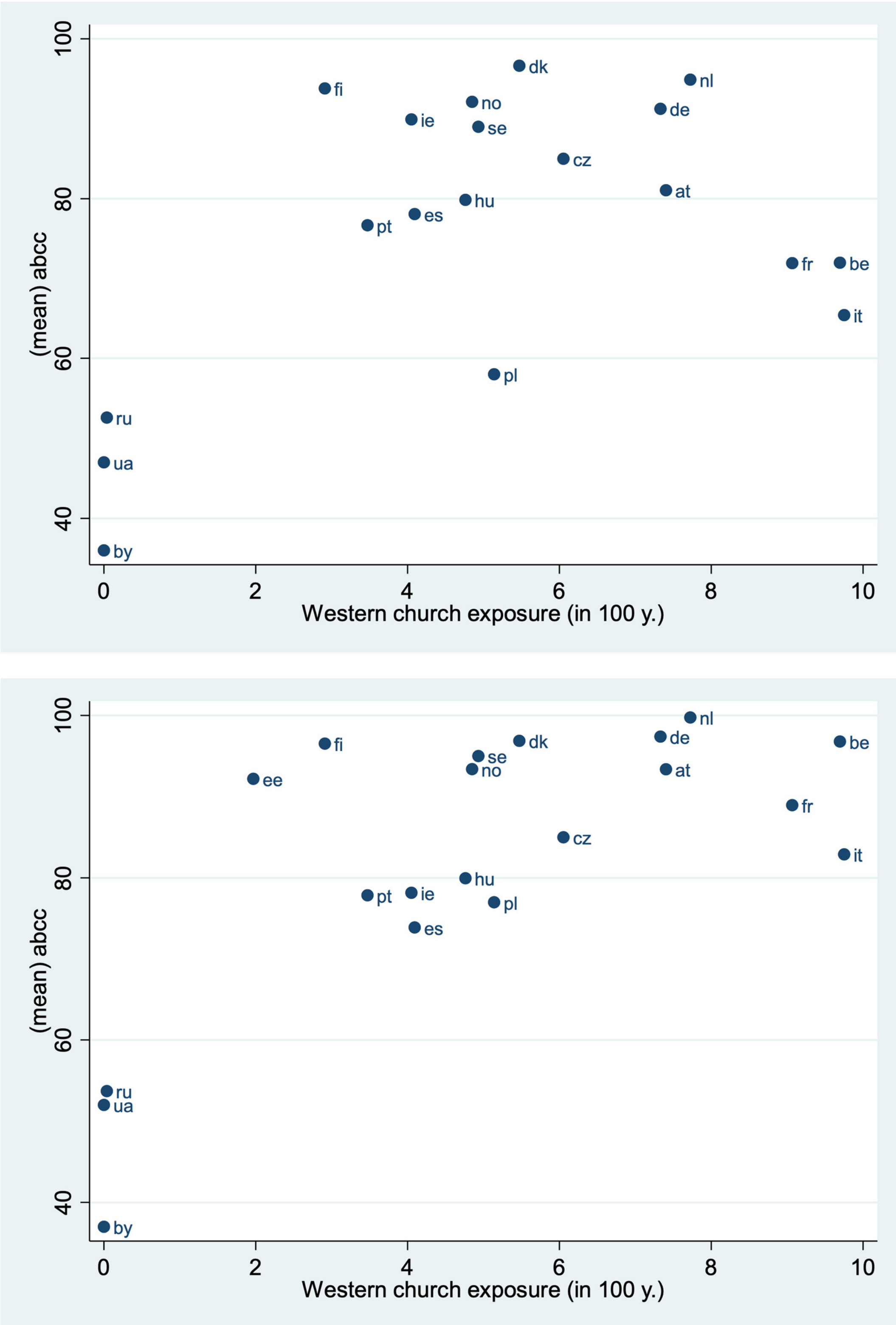


Figure J1. Western church exposure and numeracy for 1700-49 (upper panel) and 1750-99 (lower panel).

Sources: See text.

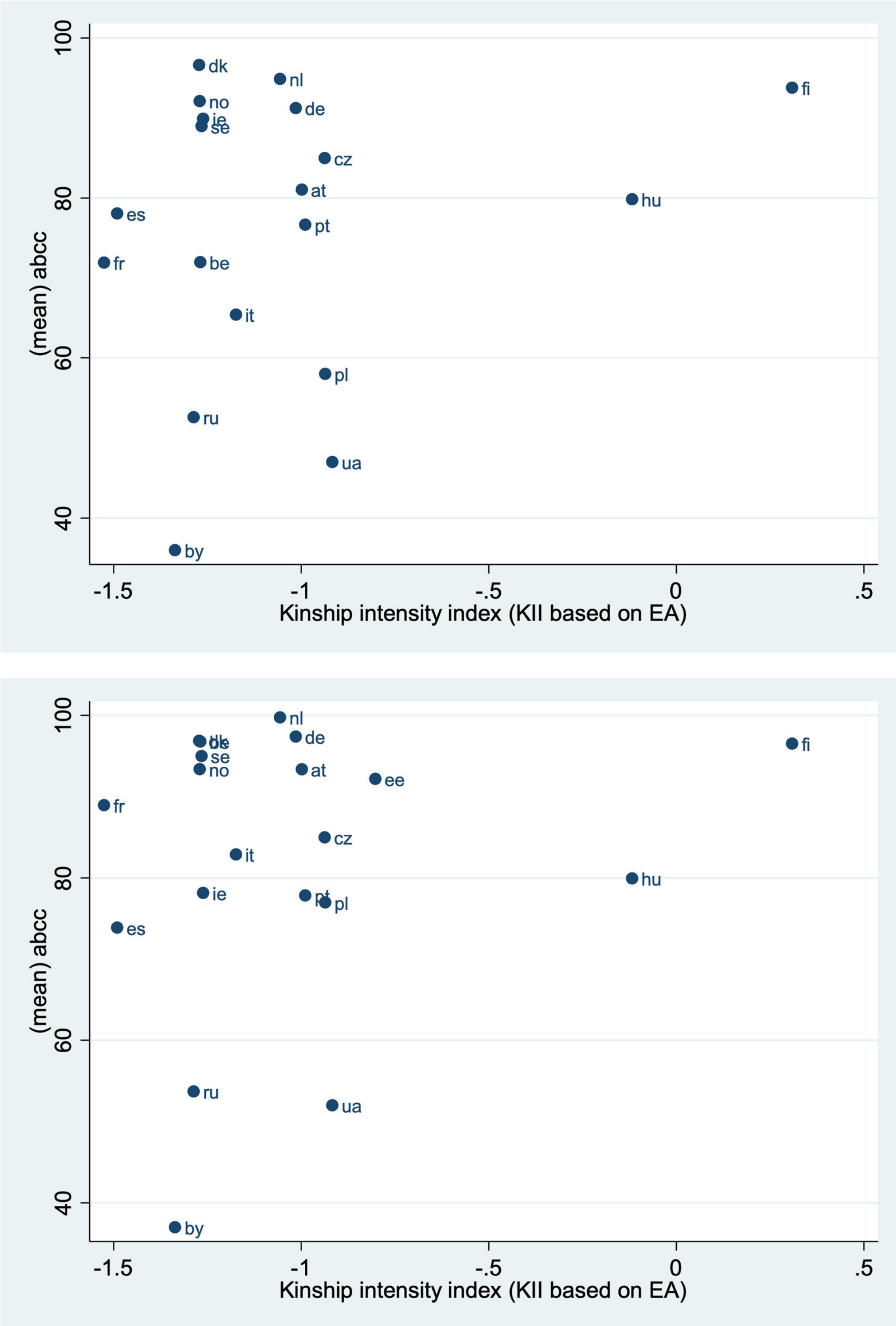


Figure J2. Kinship intensity index and numeracy for 1700-49 (upper panel) and 1750-99 (lower panel).

Sources: See text.

Appendix K: Why was Northwestern Europe not developing earlier?

Clearly, one important question that we need to answer is why north-western Europe was not already ahead of Italy and the Mediterranean in general in 1500. After all, the dairy-farming-driven gender equality mechanism might have been working already quite early, perhaps already during the medieval period, and hence the whole human capital revolution could have happened in earlier centuries already. What could be possible hypotheses to explain this?

Hypothesis 1: Recent research shows that the Western Mediterranean, including Italy, was always more advanced economically – from Roman antiquity to the 15th century. This could be quantitatively assessed by measuring its high elite human capital (Keywood and Baten 2021, see their Figure 8). The superiority of the Western Mediterranean might even stretch back in time to the Etruscan economic growth period of the late iron age around the fifth century BCE. This high level of economic development, and high state capabilities, was soon taken over by the Roman economy, laying the foundations of its Imperial expansion. The period of the 8th to 10th century saw a minor and temporary dip for the Western Mediterranean (partly affected by the expansion of Islamic Empires), but this region still remained the top performing region within Europe, as the dip was stronger in other regions (Keywood and Baten 2021). Hence the strong position of Italy in the 14th and 15th centuries can be well explained by path-dependency, relative to the poorer regions of Europe that were always less developed since Roman antiquity. The astonishing development that we need to explain was rather the catching up of the Northwest and the subsequent overtaking of the Mediterranean during the following centuries.

Moreover, the dominant position of Italy in the 14th and 15th centuries is actually not a contradiction to the hypothesis of high gender equality resulting in high numeracy. We have not much evidence on female age at marriage during the medieval period, and the available data are heavily disputed in the literature (Edwards and Ogilvie 2018, for example, criticise of Voigtlander and Voth 2006). But what we know is that interestingly, Northern Italy had a relative high female age at marriage during the high and late medieval period, as 14 studies cited by Dennison and

Ogilvie indicate. Jointly with Belgium, the female age at marriage was higher, while it was lower in other regions, such as in Northern France, Southern France, and Southern Italy (Dennison and Ogilvie 2014).[22] It is not unlikely that relatively high female age at marriage in Northern Italy (and Belgium) – both dairy farming regions, especially during the lower-population-density period after the Black death -- reflects female autonomy during the medieval period, which contributed to the Renaissance human capital and general growth of these regions. Later population growth might have reduced dairy farming and female autonomy, reflected in lower ages at marriage.

Hypothesis 2: Another plausible mechanism could be that low welfare initially leads to high female ages at marriage, which might later lead to high human capital and again higher welfare. We would like to stress up-front that the two hypotheses are not excluding each other. Both could be at work (hence we control for this effect in our regression). Carmichael et al. (2016) report some examples such as the migrants to South Africa who had higher income levels than the population in their country of origin, and they married earlier than they could have at home in the Netherlands. Moreover, the case of the UK might suggest an increase of female age at marriage during the 16th century, when real wages fell, suggesting the adoption mechanism hypothesized by Dupâquier (1972). Le Bris and Tallec (2021) discuss a substantial amount of evidence for and against this hypothesis. For example, they emphasize that while Dutch migrants in the Cape Colony might have married early, the US migrants married late. Their own study on a town in Southwestern France shows that in this town there was a marriage age response to more adverse circumstances, but then conclude that this village did not develop successfully in the long run, hence if there was an EMP pattern here, it was not the gender-equality-generates-high-numeracy (and subsequently, development) that we have in mind.

Moreover, there is quite a bit of evidence that lower welfare does not usually lead to higher marriage age. Voigtlaender and Voth studied the marriage ages after the Black Death and found that the substantial improvement in land per population ratio was correlated with high ages at marriage in England (critical: Edwards and Oglivie 2018). We observe low pressure demographic regimes, for example in Alpine regions including the Slovenian part of the Alps and some of the Italian Alpine regions, where generally high ages at marriage correspond with a relatively good standard of living, and, as soon as the Alpine regions are left towards the southern parts of Slovenia and northern Italy, the marriage ages decline jointly with the human capital level (Manfredini et al. 2013). In the developing world of the last two centuries, we always observe higher ages at marriage combined with better human capital, such as in Africa during the 20th century, where we observe higher human capital and higher ages at marriage in the southern part of the continent and the lowest human capital and the lowest ages at marriage in the Sahel zone around Niger and northern Nigeria (Garenne 2004).

In sum, we agree with Carmichael et al. (2016) that female age at marriage can decline in a new situation of higher wages, such as among the Dutch in South Africa or (negatively) in Southwest France in adverse crisis situations. Hence our results are not opposing this view. We also take it into consideration by running one specification with female age at marriage rates that were pre-adjusted for GDP/c. However, it turns out empirically that for most cases female age at marriage tends to be correlated with higher gender equality, and hence can be used as a measure for the latter. The high development level of the Western Mediterranean is not a contradiction to our hypothesis, as (a) it was always more developed since Roman antiquity and (b) in Northern Italy, female age at marriage was relatively high during the medieval period, as between-region comparisons showed.

Appendix L: What are "Proxy controls"?

The concept of "bad controls" was explained Angrist and Pischke (2009). If control variables are variables that are themselves potentially outcome variables or dependent variables, the coefficient on the main variable of interest might be biased. As an example, Angrist and Pischke (2009) explain the case of the decision about a college degree, and which impact it has on the wage of an individual. The assumption is that the decision about the college degree is randomly assigned to the persons who decide about it. The question is, should the researcher include occupation as an explanatory variable, as the right hand side variable? The answer is no, because occupation is also determined by the decision about the college degree completion. In an empirical example, the decline of the college degree coefficient that Angrist and Pischke observe, when occupation is included, cannot be sensibly interpreted.

Interestingly, Angrist and Pischke distinguish “bad controls” and “proxy controls”, the latter are variables situated between good and bad controls. For example, in the study mentioned above, ideally an IQ of the eighth grade should be included to control for the innate ability of the pupils. However, it might be that such an IQ is not available, as it is often the case in empirical research. Instead, only a variable on an IQ measure at the time of the first job interview is available. However, this IQ is both influenced by the individual's innate ability, that the researcher wants to control for, and their college decision, which would lead to endogeneity. Again, the coefficient of college degree completion is not precisely the true coefficient of interest. However, including this proxy control might be "be an improvement over an estimation with no control at all" (Angrist and Pischke 2009). If the college degree completion coefficient is positive, both with the IQ proxy control included and not included, the researcher might expect to find a true coefficient of interest between the two.

In our paper, the trade fleet is such a proxy control between good and bad, because we can control for an underlying variable that we cannot measure, namely the accumulation of trading knowledge over the centuries before the measurement was taken, going back in the Netherlands case to the Frisian trading empire and the interactions with the Hansa, which would be a good control. At the same time, it is a bad control because numeracy might lead to a larger trading fleet. By comparing the sign of the coefficient between including this proxy control variable and not including it, we can more safely assure that the age at marriage effect is significant. Similarly ambivalent are some of the other control variables, hence we term them “proxy controls”.

Appendix M: Dairy farming and the role of women

In this Appendix we discuss some additional evidence that women worked in dairy farming in Northwestern Europe. For example, women worked as milkmaids, cheesemakers or herders, which required less physical strength than plough agriculture (Smith 1981). This picture is confirmed by Broadberry et al. (2015) who demonstrates that dairy output was high in late medieval England, and women clearly had their share in this output. In several Dutch regions of the late medieval and early modern period, cultivation of grains was unproductive and as a result the agricultural output mix contained notably livestock products (van Leeuwen and van Zanden 2012). The role played by women in the expansion of dairy farming has also been established: For early modern England Allen (1991) shows that dairy farming was associated with a large role for women in the labor force. And using medieval data on dairy production and age at first marriage in England, Voth and Voigtlander (2013, pp. 2255) have shown that in areas where agriculture was specialized on cattle, women on average married later, especially after the Black Death which reduced demographic pressure in the Netherlands and the UK. In these most critical centuries, before the early modern growth began to become path dependent and self-sustainable, the reduction of the demographic pressure played an important role.

How important was dairy farming in Northwestern Europe, compared to the South and East? We considered cattle per capita in the early 19th century to obtain an idea which countries specialized in dairy farming (for milk, meat and other products, source: Clio-infra.eu). Sweden had the maximum number with 0.7 cattle per (human) population, whereas Italy had only 0.19 and France 0.16. Other Scandinavian countries such as Norway had very high values as well (0.39), similarly to Finland, while Denmark became the major supplier of butter to the UK. Interestingly, the UK and Germany were also on the higher side with roughly twice the per capita value of Italy. Russia had a very low value of 0.20, i.e., very similar to the Italian case.

We focus in our study on dairy farming in Europe, while Becker (2019) studies the current situation of sexual constraints in the developing world, finding a negative effect of pastoralism on female agency (focusing on sexuality). Especially evidence from Sahel zone Africa and the Arabian peninsula are driving her results that the absence of male pastoralists resulted in more constraints on female sexuality, the terrible extreme being infibulation. While the constraints on sexuality are probably correlated with less female autonomy, this does not automatically imply that our IV using specialization of dairy farming is affected, for the following reasons: First of all, pastoralism was not very widespread in Europe. Where pastoralism was practiced (in the Alps, for example), women also contributed to the work in pastoralism (Tollnek and Baten 2017, see their Appendix). Becker (2021) also studies the difference between pastoral and dairy-farming regions, the former consisting of large herds moved over wide distances and focusing mainly on meat production, the latter using more intensive methods to produce milk, butter, cheese, and also meat, typically closer to the village of the family households. Becker runs a separate regression for dairy-oriented regions in today's Africa. She finds that in these regions, the gender gap was significantly lower (see her Appendix C.2, Table A6).

Secondly, Baten et al. (2021) observe for Africa a drastic historical change in gender inequality in cattle-oriented versus other regions: In the early and mid-twentieth-century regions that specialized on cattle-oriented agriculture did not have substantial gender gaps in schooling. not even in the 1940s and 1950s, when schooling levels had already reached a quite substantial level by the standards of the time. In the 1960s and 1970s in contrast, Baten et al. observe a significant gender gap of schooling between cattle-oriented regions and others. The reasons for this drastic historical change are not studied yet. We could imagine that the modern development of these societies (perhaps reinforced by the independence of African countries) could have resulted in a certain polarization between more central, market-integrated, non-pastoralism-oriented regions with more gender equality, and more sparsely populated, sometimes very conservative (in several regions strongly Islamic), pastoral regions. Moreover, the terrible devastation during the 1970 famine period in the Sahel Zone might have resulted in more extreme behavior. For example, religious behavior might have become more extreme in most devastated pastoral regions, and religious extremism in these regions often results in larger gender gaps.

Appendix N: Estimating the number of underlying cases of female age at marriage data

	1500-99	**1600-99**	**1700-99**	**1800-99**
at			11216	*14020*
be		*14020*	*72904*	36452
bg				*7010*
by			*1402*	*1402*
ch				*2804*
cz		5608	47668	86924
de	*7010*	*56080*	*144406*	222918
dk		*1402*	*7010*	*16824*
ee			1402	
es	*1402*	*1402*	*7010*	5608
fi			*2804*	*1402*
fr	*4206*	1402	18226	51874
hr				*1402*

hu			57482	63090
ie			1402	*53276*
it	*11216*	*43462*	*100944*	5608
lt				*1402*
nl		*2804*	15422	16824
no			*1402*	*7010*
pl			9814	*2804*
pt		*1402*	11216	11216
ro				*4206*
ru			*7010*	50472
se		*22432*	*2804*	12618
sk				*2804*
ua			*1402*	*1402*
uk	4206	92532	154220	*67296*

Table N1: Female age at marriage data: the estimated number of underlying cases.

Notes: Dennison and Ogilvie reported the number of studies on which their estimates of female age at marriage relied, and the 175 original titles. We could generate a sample of these studies, based on 96 of these titles that we could access. Sometimes the titles do not report number of cases, if they are citing other titles that report the number of cases (often in the language of the country studied, such as Estonian or Albanian), hence we had a sample of 60 studies that reported underlying cases between 111 and 60,000 on which the female age at marriage was calculated, with a median of 1,402. We multiplied this with the number of studies that Dennison and Ogilvie reported. These are the estimates in italics above. We refined the estimates for those cases in which we had at least one underlying number of cases by country and century, and multiplied the median and number of studies country/century-specific. These are not in italics.

In general, we can note that every study reported numbers of cases was based on at least 111, and typically hundreds of cases (and sometimes several thousand individuals). Hence there is not a small-number-of-cases-problem. See also Data Appendix F. ISO-2-country abbreviations: see Figure 4.

References (online) appendices

A'Hearn, B., Peracchi, F., & Vecchi, G. (2009). Height and the normal distribution: evidence from Italian military data. *Demography*, *46*(1), 1-25.

Acemoglu, D., Johnson, S., & Robinson, J.A. (2005). The rise of Europe: Atlantic trade, institutional change and growth. *American Economic Review*, *95*(3), 546-79.

Acemoglu D., & Robinson, J.A. (2012). *Why nations fail: the origins of power, prosperity and poverty*. Crown Publishing Group.

Allen, R.C. (1991). The two English agricultural revolutions: 1450–1850. In B. Campbell & M. Overton (Eds.), *Land, labour and livestock: historical studies in European agricultural productivity* (pp. 236-254). Manchester University Press.

Allen, R.C. (2003). Progress and poverty in early modern Europe. *Economic History Review*, *56*(3), 403-443.

Angrist, J.D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, *15*(4), 69-85.

Angrist, J.D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: an empiricists guide*. Princeton University Press.

Baten, J. (2016). *A history of the global economy*. Cambridge University Press.

Baten, J. (2021). Technical paper: sources to update the Clio-Infra Data Set for Europe. *University of Tuebingen Working Paper* 103. https://uni-tuebingen.de/de/18732 Last accessed 2022, Feb 23rd.

Baten, J., Benati, G., and Ferber, S. (2022). Rethinking age heaping again for understanding its possibilities and limitations. *Economic History Review* 00(2022), pp. 1–12. https://doi.org/10.1111/ehr.13139

Baten, J., & Blum, M. (2014). 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., de Haas, M., Kempter, E. & Meier zu Selhausen, F. (2021). Educational gender inequality in sub-Saharan Africa: a long-term perspective. *Population and Development Review*, *47*(3), 813-849.

Baten, J., & Hippe, R. (2018). Geography, land inequality and regional numeracy in Europe in historical perspective. *Journal of Economic Growth*, *23*(1), 79-109.

Baten, J., & 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., Szołtysek, M. & 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), 29-63.

Becker, A. (2019). On the economic origins of restricting women's promiscuity. *CESifo Working Papers* No. 7770.

Botao Rego, A., Amorim, M. N., Ferreira, A., Santos, C., Silva, M. N., & Guerreiro, M. J. S. (2016). The diversity in age at marriage in Portugal from the seventeenth century to the early twentieth century. *Revista de Demografía Histórica-Journal of Iberoamerican Population Studies*, *34*(2), 129-159.

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

Broadberry, S.N., Campbell, B., Klein, A., Overton, M. & van Leeuwen, B. (2015). *British economic growth, 1270-1870*. Cambridge University Press.

Carmichael, S., de Pleijt, A.M., van Zanden, J.L., & de Moor, T. (2016). The European marriage pattern and its measurement. *Journal of Economic History*, *76*(1), 196-204.

Caruthers, B.G. & Espeland, W.N. (1991). Accounting for rationality: double-entry bookkeeping and the rhetoric of economic rationality. *The American Journal of Sociology*, *97*(1), 31-69.

[dataset] Clio-infra.eu, last accessed Nov 1st, 2021

Coale, A.J. (1976). *Economic factors in population growth*. Springer.

Coale, A.J., & Watkins, S.C. (1986). *The decline of fertility in Europe*. Princeton University Press.

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

Crayen, D., & Baten, J. (2010). Global trends in numeracy 820-1949 and its implications for long-term growth. *Explorations in Economic History*, *47*(1), 82-99.

de Moor, T., & van Zanden, J. L. (2006). *Vrouwen en de geboorte van het kapitalisme in West-Europa,* International Institute of Social History Publisher.

de Pleijt, A.M., & 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.

Dennison, T., & Ogilvie, S. (2013). Does the European marriage pattern explain economic growth? *CESifo Group Munich Working Paper* no. 4244.

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

Dupâquier, J. (1972). La non-révolution agricole du XVIIIe siècle. In *Annales. Histoire, Sciences Sociales* (pp. 80-84). Cambridge University Press.

Edwards, J., & Ogilvie, S. (2018). Did the black death cause economic development by 'inventing' fertility restriction? *CESifo Working Paper Series* No. 7016.

[dataset] Familysearch: Familysearch.org database: a collection of historical genealogical information. Last accessed 2022, Feb. 22nd.

Flatz, G. (1995). The genetic polymophism of intestinal lactase activity in adult humans. In C.R. Scriver (Eds.), *The Metabolic and Molecular Bases of Inherited Disease*. Mc Graw-Hill.

Garenne, M. (2004). Age at marriage and modernization in sub-Saharan Africa. *South African Journal of Demography 9(2), 59-79.*

Goldthwaite, R.A. (1972). Schools and teachers of commercial arithmetic in Renaissance Florence. *The Journal of Economic History*, 1, 418-33.

Graff, H.J. (1987). *The legacies of literacy: continuities and contradictions in Western culture and society*. Indiana University Press.

Gruber, S., & Szoltysek, M. (2016). The patriarchy index: a comparative study of power relations across historic Europe. *The History of the Family*, *21*(2), 133-174.

Hansen, C.W., Jensen, P.S., & Skovsgaard, C.V. (2015). Modern gender roles and agricultural history: the Neolithic inheritance. *Journal of Economic Growth*, *20*, 365-404.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *25*(15), 1965-1978.

Hippe, R., & Perrin, F. (2017). Gender equality in human capital and fertility in the European regions in the Past, *Investigaciones de Historia Económica – Economic History Research,* 13, 166-179.

[dataset] https://opr.princeton.edu/archive/pefp/ last accessed 2021, Feb 12.

[dataset] https://opr.princeton.edu/archive/pefp/russia.aspx, last accessed 2021, Feb 12.

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*, 579–591.

Juif, D., Baten, J., & Pérez-Artés, M.C. (2020). Numeracy of religious minorities in Spain and Portugal during the inquisition era. *Revista de Historia Economica-Journal of Iberian and Latin American Economic History*, *38*(1), 147-184.

Keywood, T., and Baten, J. (2021). Elite violence and elite numeracy in Europe from 500 to 1900 CE: roots of the divergence, *Cliometrica*, *15*, 319–389.

Koepke, N., & Baten, J. (2005). The biological standard of living in Europe during the last two millenia. *European Review of Economic History*, *9*(1), 61-95.

Koepke, N., & Baten, J. (2008). Agricultural specialization and height in ancient and medieval Europe. *Explorations in Economic History*, *45*(2), 127-146.

Krishnan P. (1977). Age at marriage in a nineteenth century Indian parish. In *Annales de démographie historique* 1977, 271-284.

le Bris, D., and Tallec, R. (2021). The European marriage pattern and its positive consequences. Montesquieu-Volvestre, 1660-1789. *MPRA Working Paper* No. 105324.

Manfredini, M., Breschi, M., Fornasin, A., & Seghieri, C. (2013). Height, socioeconomic status and marriage in Italy around 1900. *Economics and Human Biology*, *11*(4), 465-473.

Maravall, L., & Baten, J. (2019). Valkyries: Was gender equality high in the Scandinavian periphery since Viking times? Evidence from enamel hypoplasia and height ratios. *Economics and Human Biology 34*, 181-193.

Martínez-Carrión, J.-M., & Moreno-Lázaro, J. (2016). Parámetros antropométricos de los reclutas españoles antes de la transición nutricional. Análisis de las desigualdades territoriales (1858-1913). *Nutrition hospitalaria: organo oficial de la Sociedad Espanola de Nutricion Parenteral y Enteral*.

Murdock, G.P. (1976). *Ethnographic atlas*. University of Pittsburgh Press.

North, D.C., & Weingast B.R. (1989). Constitutions and commitment: evolution of institutions governing public choice in seventeenth-century England. *Journal of Economic History*, *49*(4), 803-832.

Reis, J. (2005). Economic growth, human capital formation and consumption in western Europe before 1800. In R.C. Allen, T. Bengtsson & M. Dribe (Eds). *Living standards in the past: new perspectives on well-being in Asia and Europe* (pp. 195-225). Oxford University

Press.

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 dood., 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* (pp. 121-52). Habelt Publisher.

Schultz, D., Bahrami-Rad, J.F., Beauchamp, J. P., & Henrich, J. (2019). The church, intensive kinship, and global psychological variation. *Science*, 366, Article eaau5141.

Schumpeter, J. (1950). *Capitalism, socialism and democracy*. Harper Torchbooks.

Schwekendiek, D., & Baten, J. (2019). Height development of men and women from China, South Korea, and Taiwan during the rapid economic transformation period of the 1960s–1980s. *Economics and Human Biology*, *34*, 169-180.

Smith, R.M. (1981). Fertility, economy, and household formation in England over three centuries. *Population and Development Revie*w, *7*(4), 595–622.

Steckel, R., Larsen, C., Roberts, C., Baten, J. (Eds.), *The backbone of Europe: health, diet, work and violence over two millenia*. Cambridge University Press.

Tollnek, F., & 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.

Van Zanden, J.L., & 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., & Bosker, M. (2012). The rise and decline of European parliaments, 1188-1789. *Economic History Review*, *65*(3), 835-861.

Voigtländer, N., & Voth, H. J. (2006). Why England? Demographic factors, structural change and physical capital accumulation during the Industrial Revolution. *Journal of Economic Growth*, *11*(4), 319-361.

Voigtländer, N., & Voth, H. J. (2013). How the west 'invented 'fertility restriction. *American Economic Review*, *103*(6), 2227-2264.

Wagner, P. (2005). *Bauern, Junker und Beamte. Lokale Herrschaft und Partizipation im Ostelbien des 19. Jahrhunderts.* Wallstein Publisher.

Weber, M. (1930/1976). *The Protestant ethic and the spirit of capitalism*. Scribner/Simon & Schuster.

[dataset] www.polityIV.org, last accessed Nov. 4th 2021

[1] Zero refers to extremely low numeracy, 100 to a situation in which basic numeracy is achieved.

[2] Improving the position of women has positive effects children's educational attainment, as Currie and Moretti 2003; Schultz 1988, Strauss and Thomas 1995 show.

[3] 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.

[4] In Appendix M we discuss some additional evidence that women worked in dairy farming in Northwestern Europe.

[5] While their compilation work is a gigantic step ahead, similar work has not yet been done for male age at marriage, hence we cannot use the age gap as alternative indicator.

[6] Perhaps surprisingly, England only had moderate numeracy rates 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 may have reduced the demand for skilled workers (de Pleijt and Weisdorf 2017, de Pleijt et al. 2020).

[7] The relation between FAM and the share of female-headed households might be partly mechanical: if both females and males marry later, there might be a higher probability that the wife is widowed after the death of the husband.

[8] Both indicators reflect nutritional quality and the health situation of boys and girls after weaning, roughly age 2-6.

[9] Also male age at marriage is unavailable, hence the age gap at marriage between males and females cannot be computed. However, Le Bris and Tallec (2021) report a close correlation of FAM and this age gap in South-Western France.

[10] We also experimented with interpolation or with aggregating all variables at the century-country level. The results were very similar.

[11] Another alternative was vine and olive agriculture.

[12] 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. Unfortunately, Murdock (1976) does not report about majority populations (Appendix H).

[13] In contrast to Europe, sparsely settled regions of the developing world developed pastoral specializations. Becker (2019) found that these economies impose actually restrictions on female sexuality nowadays (and by implication, reduce female agency), especially in the region between Mali, Somalia and Saudi-Arabia. But Becker finds the opposite for dairy farming specialized regions, and hence presents evidence consistent with our European data (see her Appendix B.3 and our Appendix I)

[14] See also appendix K.

[15] Lactose tolerance is unfortunately not available at the regional level for the 19th century regional regressions.

[16] 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.

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

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

[19] Literacy for males was 34.2%, the one for females only 7.1, hence the gap was 27.1%.

[20] The literacy of males and females 50.1 and 41.8%, respectively.

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

[22] In general, female ages at marriage were low in the high and late medieval period, which corresponds to a society that was in general not characterized by very high general human capital levels, in spite of the relative differences mentioned above.