Journal of Comparative Economics 51 (2023) 133–159

Contents lists available at ScienceDirect

Journal of Comparative Economics

journal homepage: www.elsevier.com/locate/jce

The roots of female emancipation: Initializing role of Cool Water ☆

Manuel Santos Silva [a,*], Amy C. Alexander [b,1], Stephan Klasen [c,2], Christian Welzel [d,e,3]

[a] University of Münster, Germany
[b] University of Gothenburg, Sweden
[c] University of Goettingen, Germany
[d] Leuphana University, Lueneburg, Germany
[e] National Research University – Higher School of Economics, Moscow, Russia

ARTICLE INFO

JEL classification:
J12
J16
N30
O15

Keywords:
Cool Water
Climate
Gender equality
Marriage
Household formation
Historic drivers

ABSTRACT

The Cool Water condition is a climatic configuration that combines periodically frosty winters with mildly warm summers under the ubiquitous accessibility of fresh water. Historically, it embodied opportunity endowments that weakened fertility pressures, resulting in household formation patterns that empowered women and reduced gender inequality. Reviewing the literature on the deep historic roots of gender inequality, this paper theorizes and provides evidence for a trajectory that (1) originates in the Cool Water climatic configuration, (2) leads to late female marriages in preindustrial times, and (3) eventually paves the way for various gender-egalitarian patterns of the present.

1. Introduction

Scholars have traced variation in contemporary gender roles across societies to several deep-rooted features of their agrarian or religious history. In the footprints of Boserup (1970), one group of researchers sees the transition to agriculture in and by itself as the origin of greater gender inequality (Hansen et al., 2015), others identify the plow-using type of agriculture as the source (Alesina et al., 2013; Giuliano, 2015), and still others argue that historical scarcity of arable land strengthens agriculture's contribution to gender inequality (Hazarika et al., 2019). With respect to religion, modern Islam has been viewed as a source of gender inegalitarian

☆ We would like to thank Timur Kuran (editor), two anonymous reviewers, Laura Barros, Lisa Höckel, Krisztina Kis-Katos, Susan Steiner, Holger Strulik, Mikołaj Szołtysek, participants at the Gender Governance Link conference in Goettingen, the Quality of Government Institute's Internal Conference in Budapest, and seminars at the Universities of Goettingen and Freiburg for valuable comments, and Florian Klassen and Alexander Stöcker for research assistance. We thank Maria Kravtsova for assistance with Russian Empire's historical data. We gratefully acknowledge financial support from Lower Saxony's Ministry for Science and Culture for the project *The Gender-Governance Link: Gender Equality and Public Goods Provision* under the initiative "Geschlecht–Macht–Wissen: Genderforschung in Niedersachsen". Alexander was also supported through the European Union Seventh Framework Research Project PERDEM (Performance of Democracy, grant agreement 339571). The views expressed here are ours and do not represent those of the funders.

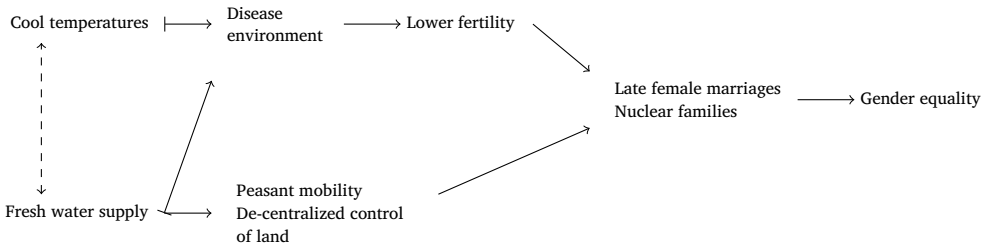
* Correspondence to: Scharnhorststr. 100, 48151 Münster, Germany.
E-mail addresses: manuel.santos.silva@uni-muenster.de (M. Santos Silva), amy.catherine.alexander@gmail.com (A.C. Alexander), cwelzel@gmail.com (C. Welzel).
[1] Sprängkullsgatan 19, 41123 Göteborg, Sweden.
[2] Passed away on October 2020.
[3] Universitätsallee 1, C4.008b, 21335 Lüneburg, Germany.

https://doi.org/10.1016/j.jce.2022.11.001
Received 21 September 2021; Received in revised form 27 October 2022; Accepted 2 November 2022
Available online 24 November 2022

norms (Inglehart and Norris, 2003).[4] In contrast, more egalitarian family patterns have been attributed to historical exposure to the Roman Catholic Church (Goody, 1983; Schulz et al., 2019; Henrich, 2020), while Protestantism has been linked to lower gender inequality in education (Becker and Woessmann, 2008).

Building on this scholarship, we theorize and provide evidence for an overlooked trajectory originating in a climatic configuration, called the "Cool Water" (CW) condition, that leads to late female marriages in preindustrial times, which in turn result in various gender egalitarian outcomes today. This article argues that the CW condition can be understood as a precursor to the agrarian and religious deep-roots discussed in the literature.

The CW condition is a specific climatic configuration that combines periodically frosty winters with mildly warm summers under the ubiquitous accessibility of fresh water. It is most prevalent in Northwestern Europe and some of its former colonial offshoots (Welzel, 2013, 2014). We argue that, over time, household formation patterns of an agrarian economy evolve, in part, as a response to the challenges and opportunities ingrained in its climate. The CW condition embodies opportunity endowments that significantly reduced fertility pressures on women and favored late female marriages already in the preindustrial era. In CW areas, the relatively favorable disease environment implied that lower fertility was needed to achieve the desired number of children that survive into adulthood. In addition, the ubiquitous availability of water, as a fundamental agricultural input, reduced mobility costs of farming households and defied centralized control of land, which favored nuclear over extensive families and neo-locality over patri-locality. The resulting family and household patterns placed women into a better position to struggle for more gender equality during the subsequent transitions toward the industrial and post-industrial stages of development. Hence, enduring territorial differences in the CW condition predict differences in preindustrial female marriage ages, which in turn predict differences in gender equality today. Below is a schematic overview of the hypothesized causal chain:



In support of this causal chain, we use the Cool Water index: a combination of absolute latitudes, continuous rain, and mild summers (Welzel, 2013, 2014). We present cross-country evidence showing that CW has a sizeable and robust impact on current female marriage ages as well as the age gap between the spouses. A one standard deviation increase in the CW index is associated with a 10-month reduction of the male–female difference in mean ages at first marriage. Furthermore, we can trace this effect back to the preindustrial epoch using European historical data on marital ages from Dennison and Ogilvie (2014). Both these effects are robust to including the other 'deep drivers' discussed in the literature. In line with these results, we show in reduced-form estimates that CW has a large and significant impact on today's gender gaps in labor force participation and life expectancy.

In the absence of experimental control, the two lurking threats to causal identification are reverse causality and omitted variable bias. Reverse causality is of no concern in our setting. The reason is the distinct temporal ordering of our variables along a far-reaching sequence of separate historic epochs, extending from (a) *original* environmental conditions manifest in CW to (b) *preindustrial* marriage patterns to (c) *post-industrial* gender-egalitarian outcomes.

With respect to omitted variable bias, there is always the risk that unobserved country characteristics might account for the relationship between CW and its hypothesized outcomes. In order to address this problem, we go at length in double-checking our results against many possible confounding factors. First, we conduct multiple sensitivity tests to the composition of the sample and find that our conclusions are not driven by certain world regions or subregions. Second, we control for a battery of additional variables proposed in the literature and find that the CW effect persists throughout. Third, following Oster (2019), we estimate how large the effect of unobservables needs to be to explain away the CW effect. For reasonable upper bounds on the full model's R^2, the degree of selection on unobservables would have to be almost as large as the degree of selection on the observable variables of our most restricted models. Thus, it seems unlikely that omitted variable bias alone could account for our results.

Another concern relates to the fact that countries did not exist in today's borders invariantly throughout the temporal scope of our analyses. To address this issue, we "ancestry-adjust" the countries' CW-scores using Putterman and Weil's (2010) post-1500 World Migration Matrix. Doing so changes the unit of analysis from today's countries to ancestral populations. Since this exercise fully reproduces all of our major results, it is safe to conclude that they are not an artifact of using countries in today's borders as the unit of analysis.

As additional evidence, we show that, at the turn of the twentieth century (1880–1930), the CW condition of the country of origin correlates negatively with the marriage probability among female immigrants (first and second generation) to the US. These

[4] Pre-modern Islamic law, however, was very egalitarian in key dimensions, such as property rights. "Once a [Muslim] woman came of age [...], [s]he had full legal capacity as far as her private property was concerned. The absence of discrimination against women as property-holders contrast with the situation in Europe well into the modern period" (Tucker, 2008, p.25).

individual-level results suggest that the CW-effect has persisted over time because it became encoded as a cultural norm. Moreover, since immigrants living in a US state face the same institutional environment, the results further alleviate the concern that omitted institutional factors might be driving the cross-country estimates.

Finally, we present subnational evidence from the late nineteenth century Russian Empire. Using the Russian Empire's 1897 census and exploiting the large geographical variation available, we document a strong negative association between regional CW and the share of young women (15–19) who are married. This association, while descriptive, is robust to controlling for several local measures of historical economic development.

Our main contribution is to propose a novel deep-rooted geo-climatic factor, the Cool Water condition, to explain historical and contemporary variation in sex-specific ages at first marriage. There are two prominent explanations for Northwestern Europe's pattern of neo-locality and late ages at first marriage for women. Voigtländer and Voth (2013) argue that the mid-fourteenth century Black Death, by making land relatively more abundant, shifted production away from labor-intensive agricultural cultivation towards land-intensive animal husbandry, an activity where women have a comparative advantage. As a response to higher female labor demand, women postponed marriage and childbirth, a process which, over time, gave rise to the European marriage pattern. We show that our results are not simply capturing a Northwest European idiosyncrasy, as the Black Death explanation would suggest. Instead, the CW-effect is fully robust to excluding Northwestern Europe or controlling for the importance of animal husbandry for preindustrial subsistence.

The second explanation argues that the emergence of the European marriage pattern was a direct result of family policies promoted by the Catholic Church during the medieval period. This hypothesis, mostly prominently developed in Goody (1983), has been recently revived by Schulz et al. (2019) and Henrich (2020). In summary, it claims that the Catholic Church (or Western Church) actively undermined kin-based family patterns by, among other policies, banning polygyny and cousin marriage, and promoting marriage based on spousal consent, rather than parental consent. We show that our main findings are fully robust to controlling for a country's length of exposure to the Western Church as measured by Schulz et al. (2019). In addition, we reproduce the main findings across regions of the historic Russian Empire, where the Orthodox (Eastern) Church has been the most dominant branch of Christianity.

Our article relates to several additional strands of literature. Our starting point is the growing consensus that gender equality in its various manifestations goes together with a host of beneficial outcomes, from economic productivity (Klasen, 2002; Klasen and Lamanna, 2009; Duflo, 2012) to distributional justice, physical security, generalized trust (Alexander and Welzel, 2015), honest government (Brollo and Troiano, 2016), effective democracy (Fish, 2002; Inglehart and Norris, 2003) and other quality of life aspects (e.g., World Bank, 2001, 2011; Branisa et al., 2013). Many studies also highlight the importance of female agency during the preindustrial Malthusian era for the timing and pace of the fertility transition and consequent projection into the modern era of sustained economic growth.[5]

Previous studies identify several deep historic roots of unequal gender roles that generate contemporary variation in gender equality in the labor market, political representation, and other outcomes.[6] Related to societies' agrarian past, unequal gender roles have been linked to the use of the plow (Alesina et al., 2013), longer histories of agriculture (Hansen et al., 2015), and historical scarcity of arable land Hazarika et al. (2019). We differ from these studies by proposing a novel deep-rooted factor, the CW condition, and by focusing on preindustrial marriage patterns as a key mechanism along the long-run causal chain.

In doing so, we contribute to the large literature, dating back to Hajnal (1965, 1982), arguing that family systems and marriage patterns decisively shape gender roles in the preindustrial era. Several authors since Hajnal link preindustrial ages at first marriage to persistent norms of gender equality (e.g., Hartman, 2004; De Moor and Van Zanden, 2010). Empirically, historical family systems are highly persistent over time, suggesting they became embedded in cultural norms (Rijpma and Carmichael, 2016). The persistence of different regional family structures across Europe is documented, among others, by Reher (1998), Duranton et al. (2009), and Szołtysek et al. (2017). Features of historic family systems have been empirically associated with several contemporary socio-economic outcomes: patri-locality is positively associated with the number of missing women (i.e., larger sex ratios at birth) (Ebenstein, 2014), nuclear household structures are positively associated with democracy (Dilli, 2016), while Spanish regions with historical stem family systems exhibit less intimate partner violence (IPV) today than regions with historical nuclear families (Tur-Prats, 2019). We contribute to this literature by emphasizing ages at first marriage as a widely available measure of different marriage and family systems, documenting the persistence of ages at first marriage over time, showing its association with multi-dimensional gender egalitarian outcomes, and showing, across different units of analysis and time periods, how ages at first marriage are systematically explained by geo-climatic conditions.

Our results for first- and second-generation immigrants in the US relate to a literature showing that norms on gender roles are transmitted from countries of origin to emmigrant descendants through intergenerational cultural transmission (e.g., Antecol, 2000; Fernández, 2007; Bredtmann et al., 2020). Similarly, Giuliano (2007) shows that European family structures are replicated by second-generation European immigrants in the US, which is consistent with intergenerational transmission of cultural norms on family structures.

The remainder of our article proceeds in the following steps. Section 2 reviews the literature and derives from this discussion our theoretical propositions. Section 3 introduces the data and variables used to demonstrate the empirical validity of our propositions. The fourth section describes the empirical strategy while Section 5 presents the findings. We conclude in Section 6.

[5] See, among others, Galor and Weil (1996), Lagerlöf (2003), Diebolt and Perrin (2013) and Prettner and Strulik (2017). For a survey of this literature, see Santos Silva and Klasen (2021).

[6] Jayachandran (2015) and Giuliano (2018) survey this literature.

2. Theoretical discussion

In Section 2.1, we start by reviewing existing hypotheses that link gender equality to specific features of a society's agrarian past. We then describe in detail how historical household formation patterns varied across societies, highlighting the distinction between those where women marry at relatively later ages vis-à-vis those where they marry immediately after puberty (Section 2.2). In late marriage societies, women are relatively more empowered (Section 2.3). Finally, in Section 2.4, we build our main hypothesis that the norm of late marriages is, in part, determined by temperature and water patterns, through their impact on fertility pressures on women.

2.1. Original sources of gender (in)equality

Three recent studies suggest alternative origins as the decisive historic drivers of gender inequality today. To begin with, Alesina et al. (2013) revive Boserup's (1970) thesis that the participation of women in preindustrial agriculture differed significantly between plow-using and plow-free cultivation systems. The plow constituted a gender-biased technology as it required more upper body strength than alternative tools, such as the digging stick or the hoe. As a result, in societies that adopted the plow, women reallocated their time away from farming towards domestic activities. This labor division along gender lines became gradually encultured into enduring norms. Alesina et al. show that the fraction of a country's population whose ancestors practiced preindustrial plow agriculture is negatively correlated with contemporary female participation in the labor force, politics, and corporate ownership. Moreover, among children of immigrants in the US and Europe, those with plow-using ancestors hold less egalitarian beliefs about the appropriate role of women.

In contrast, Hansen et al. (2015) argue that the transition from humans' original foraging lifestyle to sedentary agriculture *as such* is a driver of preindustrial disparities in gender roles, no matter what particular cultivation methods have been used. As these authors suggest, the earlier the transition to sedentary agriculture, the more intense the cultivation methods became thereafter. Intense methods of cultivation generate a demand for cheap mass labor, which in turn increases the fertility pressure on women. As a consequence, women reallocate their time from fieldwork to raising children and other indoor activities related to caretaking. Societal beliefs about gender roles then incrementally evolved in support of this labor division. Accordingly, societies with longer histories of agriculture had more time to enculture patriarchal values in their moral systems. Indeed, Hansen et al. show that longer histories of agriculture are negatively correlated with female participation in the labor force, politics, education, as well as positively correlated with the sex ratio at birth (boys/girls). The correlation remains significant even after controlling for ancestral plow use, which retains its significance. Not only is this relationship present in a cross-country sample but also for a sample of European regions, and for a sample of second-generation immigrant children in the US.

The third study by Hazarika et al. (2019) argues that historic resource scarcity shaped cultures of gender discrimination. This claim is consistent with patterns of sex-inequality and resource availability in some non-human primate species, gender gaps in prehistoric human skeleton sizes, and contemporary evidence on the relationship between material deprivation and gender bias in intra-household resource allocation. According to this thesis, prehistorical differences in resource scarcity gave rise to a persistent culture of gender discrimination. The authors measure historic resource scarcity by limitations of arable land and show that these limitations are negatively correlated with present-day measures of gender equality across countries and positively correlated with population sex-ratios across districts in India.

While these insights are valuable, we offer a fourth explanation that we believe adds an important element to understanding the deep drivers of gender inequality and also provides a plausible transmission channel.

2.2. Historical household formation patterns

The feature in preindustrial household types receiving most attention refers to a bundle of elements that Hajnal (1965, 1982) branded as the "Western family pattern" and described as unique to Northwestern Europe (Todd, 1985, 1987; Hartman, 2004). Based on archival records from the seventeenth and eighteenth centuries, Hajnal (1982) stresses four exceptional features of household formation in Northwestern Europe, which supposedly date back at least to medieval times[7]: (a) late ages at first marriage for women, resulting in smaller age gaps between husband and wife; (b) a considerable proportion of women (and men) never married, which, together with late marriage, implied for women that they lived under lower pressure to maximally exploit their reproductive potential; (c) neo-local residence: couples move into their own household upon marriage—a feature that reduces obligations to the extended family and, thus, fosters individual autonomy; and (d) a widespread practice among both adolescent men and women to work as contracted servants in non-kin households until marriage.

In sharp contrast, household formation patterns in the then advanced areas of the Middle East, China, and India—which survived with remarkably small changes until the twentieth century—involve the exact opposite features: (a) much earlier marital ages for women, often resulting in large age gaps between husband and wife; (b) near universal marriage, indicating an intolerance of women

[7] For example, in his classical study of the Germanic 'barbarian' tribes, *De Germania*, from around 98 AD, Tacitus already describes a Germanic late marriage pattern, in contrast to the early marriage tradition of the Roman Empire: "The [Germanic] young men marry late and their vigor is thereby unimpaired. The girls, too, are not hurried into marriage. As old and full-grown as the men, they match their mates in age and strength, and the children reproduce the might of their parents" (quoted in Herlihy, 1985, p.73). See Hartman (2004, chapter 3) for a review of the historical evidence.

just living by themselves (including quick remarriage of widows); (c) patri-local residence: freshly married couples move into the household of the husband's parents under their headship; and (d) an almost exclusive use of family labor for household production.[8]

From the gender perspective, the most fundamental element of the Northwest European household system is late ages at first marriage for women (Hajnal, 1965, 1982; Smith, 1981; Hartman, 2004). In Northwestern Europe, women would typically marry in their twenties, while in the early marriage systems of the Middle East, India, and China, brides would marry in their early teens, immediately after puberty. In late marriage societies, women's ages at first marriage were elastic with respect to local economic conditions (Smith, 1981; Carmichael et al., 2016; Cinnirella et al., 2017) and, indeed, for certain societies, at certain points in time, ages at first marriage were very high even by contemporary standards. For example, the mean age at first marriage from Norway's 1875 census was 27.3 years for women and 28.5 for men (Ruggles, 2009, Appendix Table). Similarly, it was on average 26 for women and 28 for men in rural Germany in 1740–1860 (Klasen, 1998). As a consequence of late marital ages for women, the age gap between the spouses was substantially lower than in early marriage societies, where the husband would often be ten to fifteen years older than his wife. For example, in Bangladesh, the mean age at first marriage in 1974 was 16.4 years for women and 24 for men (UN, 2009).

Another important aspect of late marriage societies is the high proportion of never married individuals of both sexes. At times, the never married constituted more than ten percent of the adult population in preindustrial Northwestern Europe, whereas they were rarely more than five percent of the adult population in early marriage societies (Hajnal, 1965, 1982).

In late medieval Northwestern Europe, late marriages were intimately linked to the institution of service through which a large share of young boys and girls would leave the parental household in their early- to mid-teens and circulate as servants on a contractual basis, typically for several years until marriage. The practice of contracted, non-kin service means that households form on the basis of consent instead of lineage. The fact that marriage itself has been an act of agreement among adult non-relatives, instead of being pre-arranged among people belonging to the same family circle, further detached household formation from lineage and strengthened its contractual character. Importantly, service did not mean *domestic* service, in fact "[m]ost servants were not primarily engaged in domestic tasks, but were part of the workforce of their master's farm or craft enterprise" (Hajnal, 1982, p. 473). Contractual household service defied a vertical allocation of labor across social strata: "It was not only the poor and landless whose children went into service. Those who operated their own farms and even farmers with large holdings sent their children into service elsewhere, sometimes replacing them with hired servants in their own household" and "[t]here was no assumption that a servant, as a result of being in service, would necessarily be socially inferior to his or her master" (Hajnal, 1982, pp. 471 and 473). Thus defined, the institution of service fulfilled two roles in late marriage societies: (i) it provided a subsistence base for unmarried young men and women as well as a means to accumulate sufficient savings to then start a family, and (ii) it created a flexible workforce for ageing farm tenants in need of support.[9]

The same logic clarifies why contracted household service was not a widespread phenomenon in early marriage societies: whenever girls married at puberty, there was no chance for service to be part of their premarital life-cycle, and the availability of family labor, including children, in multi-generational households meant that local labor markets were peripheral.

2.3. Implications for gender equality

The Northwest European system of late marriage for women and life-cycle service for the young had profound implications for gender relations. While Hajnal already outlined the most important of these implications, Hartman (2004) revisits and extends his analysis.

First, marrying later reduces women's childrearing burden during the most productive years of their life-cycles. In addition, women marrying in their mid-twenties are more mature, experienced, and arguably more confident than women marrying at puberty. Combined with a smaller age gap to their spouses, women in late marriage societies accordingly enjoy more intra-household bargaining power than their early marriage counterparts. Moreover, in late marriage societies, the intervals between generations are inevitably larger than in early marriage societies. Therefore, while it was common for newly married couples to head their own household, it was typical of earlier marriage societies for the wife to join the husband's multi-generational household. As a result, in early marriage household systems, there are often more individuals positioned above the young wife: not only her parents in-law but also other members of the husband's kin, like his siblings, uncles, and cousins. In such a setting, a freshly married woman has little bargaining power. To sum up, the absolute age at first marriage, the age gap between the spouses, and the amount of household members who are hierarchically superior to the wife are all important determinants of adult women's status in the household.

Second, in late marriage societies, individual consent is a necessary condition for marriage, whereas in early marriage societies parental consent is decisive. The mere fact of marrying later meant that a considerable proportion of the parents of freshly married couples were no longer alive at the time of marriage. For instance, "as in sixteenth-century Lyon, one-third of teenagers becoming apprentices and *one-half of the young women marrying for the first time* were fatherless; [and] as in seventeenth-century Bordeaux, more than one-third of the apprentices had neither parent alive" (Davis, 1977, p. 87; emphasis added). In contrast, in earlier marriage systems, young women were more likely to have living parents at the time of marriage and living in-laws throughout a substantial portion of their marital years. Unsurprisingly, marriage norms in these systems were tilted towards parental arrangement rather

[8] Of course, substantial variation existed around these two stylized household systems (e.g., Szołtysek, 2014). But for reasons of clarity, it is helpful to take this simplification as a point of departure.

[9] For historical evidence on the contractual household system from mid-fourteen century England see Poos (1991). For older, yet more scattered, evidence from ninth century France (around Paris) see Herlihy (1985).

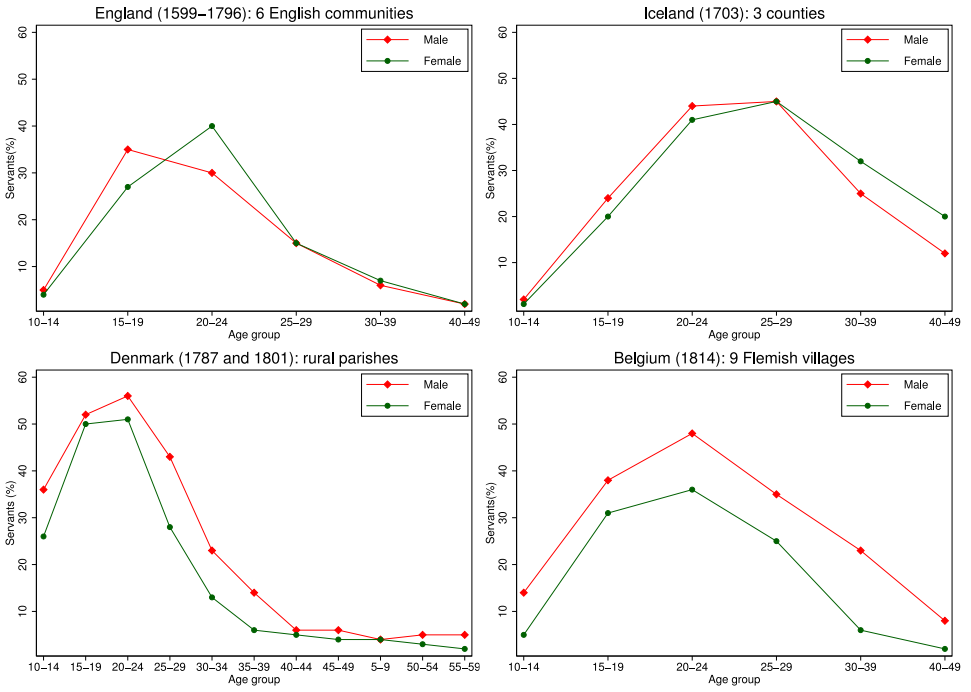


Fig. 1. Service prevalence by age and sex for selected Northwest European countries.
Source: Hajnal (1982): Tables 13, 14, and 17.

than individual choice (e.g., Edlund and Lagerlö, 2006). It is plausible to assume that individual choice lowers the intra-household bargaining costs for both spouses; either because they match with partners that have similar preferences, or because mutual consent brings about a higher degree of altruism to individual preferences.

Third, the institution of contractual farm service is a driver of gender equality norms and behaviors in its own right. As shown in Fig. 1, contractual service was a common phase in the life-cycle of *both* men and women across Northwestern Europe.[10] Indeed, it is striking how similar the patterns of service were among the life-cycles of both sexes: the prevalence rates are not too different in absolute terms (although, in general, they were higher for men across all age-groups) and their age-distributions are remarkably similar. Over the life-cycle, service started roughly at 15–19 years of age for both sexes—an age range at which most women in early marriage societies were already married. Service peaked at 20–24 years of age and then declined as people left service to get married. As a result, the life-cycle of young men and women in late marriage societies is much more similar than in early marriage societies, where young brides are separated early on from their male peers by the experiences of marriage and motherhood, often living in seclusion. The convergence of life experiences between the sexes in Northwestern Europe was thus a force challenging the boundaries between gender identities. By contrast, the divergence of life experiences early on in the Middle East, India or China, kept gender identities and domains of activity strictly separate (Hartman, 2004).

In Northwestern Europe, the years of service were an ideal opportunity for young people to accumulate savings, skills, and meet potential marriage partners (Hajnal, 1982; De Moor and Van Zanden, 2010). For the young servants, high mobility strengthened their maturity and weakened patriarchal norms, to the benefit of women (Reher, 1998; Hartman, 2004). As summarized by Hajnal (1982, pp.474–475), "While in service, women were not under the control of any male relative. They made independent decisions about where to live and work and for which employer. There was also financial independence even though women servants' wages were lower than men's."[11]

Remarkably, the implications of gender gaps at first marriage were well understood early on, as recorded in legal codes and councils from early medieval Europe (Herlihy, 1985). For example, a late eighth-century council (796–97) held at Fréjus, in today's Southeastern France, "explicitly states that the groom and bride should be 'not of dissimilar age but of the same age.' Many abuses take place, it warns, when the groom is adult and the bride immature, or *vice versa*. [...] The council strongly advocated that marriage unite mature partners of equal age" (Herlihy, 1985, pp. 75–76).

[10] A similar institution existed in Japan (Cornell, 1987). Among all non-Western agrarian civilizations, Japan is the one whose geo-climatic configuration is by far the closest to Northwestern Europe's (as defined by the Cool Water index, see Section 3.1).

[11] A recent study on eighteenth-century Amsterdam shows that wages earned by female servants were substantial: "most servant women could save between one-third and half of the amount of money an unskilled man could save in the same period of time" (Boter, 2017, p. 71).

To avoid misunderstandings, we are by no means claiming that late female marriage ages created anything close to perfect gender equality. For example, studies have found gender gaps in survival opportunities in late marriage societies (e.g., Klasen, 1998), even though the magnitude of the gaps tended to be smaller than in current-day South Asia or China (Klasen, 2003). Instead, we suggest that late marriage societies generated opportunities for female individualization that were not available for their early marriage counterparts.[12] While these opportunities did not eradicate women's discrimination, they built the basis for emancipatory struggles against it. The consequences of these struggles are evident in the global variation in gender discrimination across countries today. Indeed, women of late marriage societies were in a better position to turn external events to their own advantage. The experience of service, for instance, could be turned into labor force participation during war periods (Goldin, 1991; Acemoglu et al., 2004). Centuries of high fertility elasticity with respect to environmental factors facilitated the adoption of modern family planning technologies, such as the contraceptive pill (Goldin and Katz, 2002). Moreover, a legacy of consensual marriages, joint decision making, pooling of resources at marriage, and higher female bargaining power meant that women could reap the benefits of new consumer technologies that greatly reduced the burden of household chores (Greenwood et al., 2005; Cavalcanti and Tavares, 2008; Coen-Pirani et al., 2010). In early marriage societies, these opportunities passed by without major changes in the position of women.

2.4. Origins of household formation patterns

Diamond (1997) reinvigorates the notion that agriculture is forcefully shaped by an environment's natural endowments. Specifically, what kind of crops can be grown and what type of cattle can be bred depends directly on climatic conditions, especially a climate's thermo-hydrological configuration, that is, temperature and water patterns. Consequently, the ways in which farming households form in an agrarian economy should be a response to the challenges and opportunities embodied in this economy's climatic conditions.

Welzel (2013, 2014) introduces a particular thermo-hydrological constellation: the "Cool Water" (CW) condition. The CW condition combines periodically (albeit not permanently) frosty winters with mildly warm summers under the ubiquitous and permanent accessibility of fresh water sources. Fig. 2 presents the world distribution of the CW condition, as measured by the CW index (presented in detail in the next section). This condition is most prevalent in Northwestern Europe and fades gradually away as one moves to Eastern and Southern Europe. Outside Europe, the CW condition only prevails in the former settler colonies of Northwestern Europe (i.e., parts of North America and Australia and New Zealand) and—at a less pronounced level—in Japan, the Korean peninsula, and adjacent territories in East Asia, as well as the isolated Southern tips of South America and Africa. According to Welzel, the significance of the CW condition originates in the fact that it bestows on people some very basic existential autonomies that are absent under other conditions. These autonomies incentivize a late marriage type of agrarian household formation. Two types of existential autonomy are particularly noteworthy: *reproductive autonomy* and *water autonomy*.[13]

Reproductive autonomy is the degree to which people are exempted from the pressure to maximize fertility. A crucial determinant of fertility pressure in agrarian societies is the infant mortality rate incurred naturally by an environment's pathogen load; the infectiousness of its water sources; the scarcity of fresh water, dairy products, and other foods; and heat stress. The thermo-hydrological features of the CW condition reduce these risk factors. As a result, women in CW-areas could afford lower fertility levels to achieve the desired number of children that survive into adulthood.[14] Moreover, the type of combined cereal-livestock farming made possible by the CW condition requires lower labor inputs than, for instance, irrigation-managed agriculture and rice or tropical crop cultivation. Lower demands for labor inputs, including child labor, further reduce fertility pressures. Hence, women enjoyed more reproductive autonomy with respect to the timing of first marriage, the interval between pregnancies, and a higher recognition of their worth beyond the birth and care of children.

Reproductive autonomy partly originates in climatically induced disease security. The prevalence of non-host pathogens—from Malaria to yellow fever, dengue fever and other tropical diseases—increases with average temperatures, a fact known as the latitudinal gradient (Guernier et al., 2004; Cashdan, 2014). Before 1750, infant mortality was mainly determined by infectious diseases (e.g., Deaton, 2013, pp. 81–87).[15] In CW-areas, both cold temperatures and abundance of fresh water lower the pathogen

[12] Centuries-old household formation systems shape women's self-perceived agency and expectations about, say, partner selection and marital consent. For example, in the Indian Human Development Survey of 2005, 55 percent of married women who felt they played a role in the selection of their husbands first met them close to or on the wedding day (Banerji et al., 2008); a situation hard to conceive of wherever mutual consent, instead of parental consent, is the prevailing norm.

[13] See Welzel (2013, 2014) for broader implications of the CW condition for long-run development. Here, the discussion focuses exclusively on household formation patterns.

[14] We speculate that this is the mechanism implicit in Alfred Marshall's observation, back in 1890, that: "The age of marriage varies with the climate. In warm climates where childbearing begins early, it ends early, in colder climates it begins later and ends later; but in every case the longer marriages are postponed beyond the age that is natural to the country, the smaller is the birth-rate; the age of the wife being of course much more important in this respect than that of the husband" (Marshall, 2013, p. 150).

[15] Of course, Northwest European countries were early adopters of modern public health innovations, such as vaccinations, from the second half of the eighteenth century onwards. Ager et al. (2018), for example, show how the introduction of the smallpox vaccine in Sweden in 1801 reduced infant mortality and, consequently, gross fertility. In this article, we are referring to long-run preindustrial differentials in infant mortality *before* such innovations became available. For example, in 1750–1800, before the introduction of smallpox vaccination, Sweden's infant mortality rate was fluctuating around a stable mean of 211 deaths out of 1000-live births (Ager et al., 2018, Figure 3, Panel A). In 1960, the corresponding figure was 210 in Egypt and 220 in Nepal; in 1965, it was 225 in Afghanistan and 263 in Yemen (estimates from World Bank, 2016). If we assume that these latter countries experienced some improvement in infant mortality rates in the two-hundred-year period 1750–1960, they must have exhibited historical infant mortality rates that were substantially higher than those of Sweden. The lack of available cross-country data on preindustrial infant mortality rates, particularly for non-European countries, prevents us from testing this mechanism in an econometric framework.

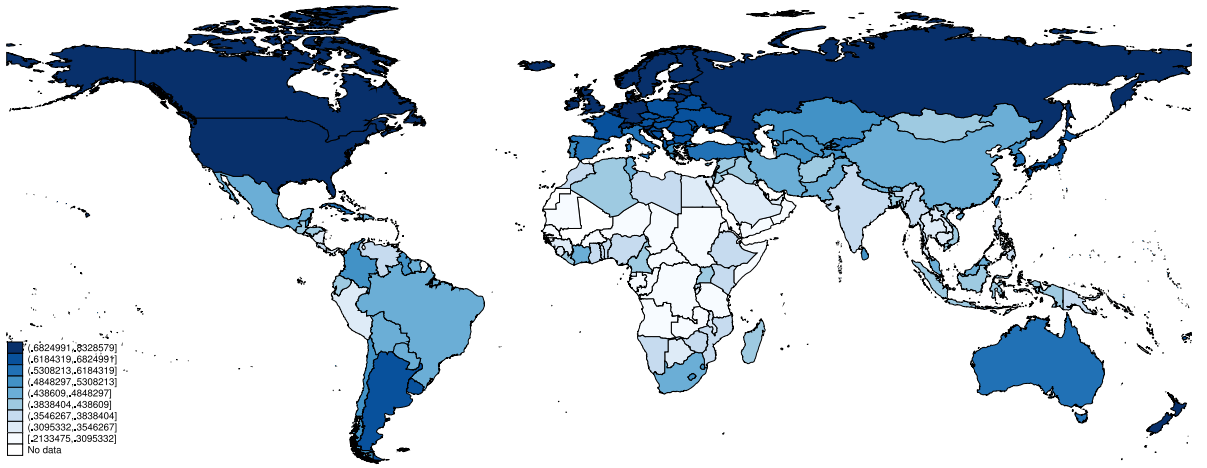


Fig. 2. World distribution of the Cool Water condition. Notes: Cool Water condition is measured by the Cool Water index (Welzel, 2013, 2014). It combines the latitude of a country's centroid, the mean temperature in the hottest month, and mean rainfall in the driest month. Higher values of the index represent areas with cold winters, mild summers, and continuous rainfall. Technical details on the index's construction are presented in Section 3.1.
Source: Temperature and precipitation data are from Parker (2000) and Gallup et al. (2010).

load, also because fresh water is safer to drink under cold temperatures (Cashdan, 2014). The resulting lower fertility requirement in CW-areas meant that women's time was less constrained by childrearing: young women could therefore postpone marriage in response to labor market incentives (Smith, 1981; Cinnirella et al., 2017).[16]

Water autonomy is the ability to access fresh water freely and permanently either for consumption, fishing, agriculture, transportation, or water power. Water autonomy is higher in colder, rainy areas. The availability of water was a major constraint for farming households in traditional agrarian societies. Areas suitable for large-scale irrigation agriculture developed more autocratic institutions and labor-repressive regimes compared to rain-fed agricultural societies (Bentzen et al., 2017). The coordination of large irrigation infrastructure projects and the allocation of scarce water resources among farmers were best achieved by a centralized authority that could enforce water-sharing rules and collect taxes to finance infrastructure investment. The monopoly of scarce water resources then became an instrument through which a small landowning elite extracted rents from a mass of peasant households. This system of agricultural production was typical of Middle Eastern, Andean and Mexican societies, as well as India and China (Wittfogel, 1957). In contrast, rain-fed agricultural production was highly decentralized. The level of rents extracted by large feudal landowners in Northwestern Europe never reached the amount extracted by their irrigation-areas counterparts (e.g., Jones, 1981; Powelson, 1994; Mitterauer, 2010). Moreover, the ubiquitous availability of fresh water makes most available land arable, thus turning valuable land into a ubiquitous asset that defies centralized control. Entry barriers and fixed costs of farming are lower when there is no need to maintain expensive irrigation (Haber, 2012). Finally, the lush pastures typical of CW-areas lend themselves to a combined form of cereal and livestock farming that widens nutritional options and easily feeds a small family while keeping the demand for child labor at the low end.[17] Together with a lower infant mortality, the weak demand for child labor further reduces the fertility pressures on women.

Thus, areas with cool temperatures and rainfall throughout the seasons enhance *water autonomy* and *reproductive autonomy*, thus proliferating the natural basis of egalitarian individualism and its expression in the late-marriage system. Over time, these marriage patterns became embedded in both formal and informal institutions, thereby persisting until today. We note that, in theory, the CW condition could still have a direct impact on present day marriage ages. We believe however that, if present, this direct effect is at most marginal. The mechanisms of *reproductive* and *water* autonomy lost relevance in most modern countries either because family farming is now a residual activity in many countries, or because medical innovations and public health measures have weakened the link between geography and infant mortality. Therefore, we posit that any persistence of the CW effect on contemporary marriage patterns runs through the long-term effect of CW on historical marriage patterns.

From the theoretical discussion above follow four propositions, which the remainder of this article will investigate:

(1) A society's CW condition correlates positively with its historical age at first marriage for women. Because of data limitations, we demonstrate this proposition mostly for European countries and a selection of non-European countries, representing major civilizations in pre-colonial history, but also for subnational regions of the historic Russian Empire, and US first- and second-generation immigrants at the turn of the twentieth century.

[16] Some authors have proposed a direct link between high pathogen prevalence and collectivist cultures, as societies develop strong distrust of strangers and in-group bias as an evolutionary strategy against infectious diseases (Fincher et al., 2008; Murray and Schaller, 2010; Höckel, 2018). This theory could also explain why, in areas with high pathogen prevalence, preferences for extended households based on kinship were relatively stronger.

[17] In thirteenth- and fourteenth-century England, for example, the labor inputs, per unit of land, required for pastoral agriculture were one-fifth of those required for arable cultivation (Campbell, 2000, p. 10).

(2) Because centuries-old household formation patterns become encultured in a society's enduring norms, a society's CW condition correlates positively with contemporary ages at first marriage for women. We demonstrate this proposition for all countries in the world with available data.
(3) Historically as well as contemporary late ages of marriage for women are positive determinants of various gender egalitarian outcomes today, mediating the initial impact of the CW condition.
(4) The CW condition's emancipatory impulse is not absorbed by other deep historic drivers championed in the literature, but is a significant driver even when controlling for these other effects.

3. Data and descriptives

We briefly describe the variables used in the empirical section and present selected descriptive statistics. A full list of all variables used, their sources and summary statistics can be found in Table A.1.

As main contemporary measures of gender equality, we use the average female to male ratio in 1990–2010 of: (i) labor force participation rates for the age group 25–59 from the ILO Laborsta EAPEP 6th Revision (2011), (ii) life expectancy at birth from the World Bank's World Development Indicators, and (iii) mean years of education of the 25+ years old from Barro and Lee (2013). In addition, we consider the share of firms with some degree of female ownership, the share of parliament seats held by women, and the UNDP's Gender Development Index. These variables capture different dimensions of gender equality and are all positively pairwise correlated in our sample.

Our contemporary data for ages at first marriage are from the United Nations' *World Marriage Data* (UN, 2009) and cover the period 1960–2008. We use the singulate mean age at first marriage (SMAM) for men and women.[18] For each country, we select data from its first available year, which ranges from 1960 to 2006 (see Table 1) with the period 1960–1980 accounting for 80 percent of the observations.[19]

We take historical data on European female ages at first marriage from Dennison and Ogilvie (2014). They collect marital ages between 1500 and 1900 from 365 studies on historical demography and harmonize the data by regressing the marital ages on several characteristics of the sources from which they were extracted.[20] We use the country-specific coefficients from that multivariate regression, where England is the omitted country (Dennison and Ogilvie, 2014, Table 2). The fact that female ages at first marriage were rising, on average, throughout the whole 1500–1900 period is not a major issue since Dennison and Ogilvie's regression also controls for the historical time period covered by each of the demographic studies, thereby purging the country-specific estimates from overall trends.[21] Table 1 shows that, on average, the 28 European countries for which data are available have female ages at first marriage around 2 years below that of England. Yet, there is substantial variation: Belarusian brides were almost 7 years younger (at first marriage) than English brides, whereas Danish brides were approximately 2 years older. The female age at first marriage for England, the reference country, is 25.26 years.

For non-European historical data, we take Gapminder's female SMAM for the period 1800–1900. Of the 27 countries with available data, 11 are non-European.[22] We interpret these data with caution, since they are collected from a variety of sources and, in some cases, supplemented by qualitative adjustments or backward extrapolations by Gapminder.[23]

Considering the deep historic determinants of gender equality, we use the three variables most prominently discussed in the literature. To begin with, *years of agriculture* is the number of thousands of years from 1500 C.E. since the Neolithic revolution, as provided by Putterman and Trainor (2006). *Plow usage* and *agricultural suitability* are taken from Alesina et al. (2013), with plow usage indicating the proportion of a country's population with ancestors that used the plow in preindustrial agriculture. Agricultural suitability measures the suitability of ancestral land for the cultivation of barley, wheat, sorghum, rye, foxtail millet, or pearl millet. Summary statistics for these variables are shown in Table 1.

3.1. The cool water index

To measure the Cool Water condition, we rely on geo-climatic data from Parker (2000) as well as Gallup et al. (2010). These data have no specific time frame. Thus, apart from minor short-term fluctuations—such as the little ice age and interglacial warming—the CW condition captures territorial differences in thermo-hydrological conditions that have been relatively constant over the past 11,000 years when the last ice age ended.

[18] The SMAM estimates the mean number of years before marriage for those marrying before age 50. The SMAM is a widely used statistic for cross-country comparisons, because it is indirectly estimated from the sex- and age-specific distribution of the never-married population. As such, the SMAM does not rely on self-reported ages at first marriage data, which are unavailable for many countries.

[19] We make only two adjustments. First, for Malta, we do not use data from its first available year, 1967, because it includes nationals living outside the country; instead we use data for the next available year, 1985. Second, for Jamaica, we do not use the first available year, 1970, because it is an extreme outlier from trend, as shown in Figure A1; instead we use data for the next available year, 1982. Our manual adjustments do not substantially influence the results.

[20] Carmichael et al. (2016, p. 200) discuss limitations of these data.

[21] Denmark's 2.36 value, for example, "shows that its female age at first marriage was 2.36 years higher than England's, controlling for time period, unit of observation, settlement size, publication type, and sources and methods used" (Dennison and Ogilvie, 2014, p.663).

[22] Armenia, Azerbaijan, Bangladesh, China, Egypt, Georgia, India, Japan, Kazakhstan, Pakistan, Sri Lanka, and United States.

[23] For more details, see the original documentation of these data at https://www.gapminder.org/wp-content/uploads/2008/10/gapdoc009.pdf. For the 10 European countries for which both Gapminder and Dennison and Ogilvie (2014) report data, the correlation coefficient is 0.95. Non-European data are of much lower quality. They are only used for an additional analysis; results are posted in the Online Appendix.

Table 1
Descriptive statistics for selected variables.

	Mean	(Std. Dev.)	Min.	Max.	N
Average female–male ratio in 1990–2010:					
Labor force participation	0.68	(0.21)	0.15	1.01	191
Life expectancy	1.07	(0.04)	0.99	1.21	202
Years of education	0.82	(0.22)	0.21	1.41	146
Ages at first marriage:					
Contemporary					
Female	21.96	(2.88)	15.56	32.19	214
Year of obs.	1975.97	(9.26)	1960	2006	214
Male	26.13	(2.28)	21.13	34.49	209
Female/male	0.84	(0.07)	0.64	0.98	209
Male–female	4.12	(1.76)	0.5	9.93	209
Historical (Europe only, ref. = England)					
Female	−2.07	(3.1)	−6.81	2.36	28
Deep determinants:					
Cool Water	0.48	(0.15)	0.21	0.83	183
Years of agriculture	4.31	(2.42)	0	10	165
Plow	0.48	(0.48)	0	1	227
Agricultural suitability	0.54	(0.33)	0	0.98	214

Notes: The unit of observation is a country. Differences in the number of observations across variables reflect data availability. *Sources:* Labor force participation refers to ages 25–59 and is taken from the ILO Laborsta EAPEP 6th Revision (2011). Life expectancy at birth is from the World Bank's World Development Indicators. Mean years of education of the 25+ years old are from Barro and Lee (2013). Contemporary ages at first marriages are from the United Nations' *World Marriage Data* (UN, 2009). For each country, earliest year available between 1960 and 1990 is selected. Historical ages at first marriage for Europe are from Dennison and Ogilvie (2014). Cool Water is the Cool Water index described in Section 3.1. It combines the latitude of a country's centroid, the mean temperature in the hottest month, and mean rainfall in the driest month. Higher values of the index represent areas with cold winters, mild summers, and continuous rainfall. Years of agriculture is the number of years (in thousands) since the Neolithic revolution (from 1500) from Putterman and Trainor (2006). Plow is the proportion of population with ancestors that used the plow in preindustrial agriculture from Alesina et al. (2013). Agricultural suitability, also from Alesina et al. (2013), is the share of ancestral land suitable for the cultivation of barley, wheat, sorghum, rye, foxtail millet, or pearl millet.

The CW condition constitutes a specific thermo-hydrological configuration, namely, the combination of cold winters with mild summers under the ubiquitous availability of fresh water. This condition is prevalent in territories of high latitude in coastal proximity.

To capture the CW condition's thermal aspect, we take latitude as the starting point, using each country-centroid's latitude in degrees.[24] Higher latitudes get us away from the tropics and into areas with lower seasonal temperatures on average.

However, not all high latitude areas comprise the CW condition's moderate seasonality, which combines winter cold with mostly mild summer heat. Thus, we need to further qualify latitude for this additional thermal condition. To do so, we take the usual peak temperature (in degrees Celsius) in a country-territory's hottest month of the year, which is July or August in the Northern hemisphere and February or March in the South. Interestingly, while average annual temperature correlates strongly and negatively with latitude,[25] summer heat is uncorrelated with latitude.[26] For instance, summer heat peaks are as high in Mongolia as they are in Somalia. This pattern reflects the fact that mild summers mostly prevail in high latitudes but not all high latitudes belong into this category because they are divided into continental climates (with high summer peaks) and maritime climates (with low summer peaks). Thus, latitude is an ideal representative of cold winters, yet not of mild summers. To measure indeed the mildness of summers, we calculate the inverse of peak summer heat.

For the very same reason, high latitudes include most of the countries that possess the hydrological features of the CW condition, and yet this is only a subset of the countries in high latitudes. Therefore, it is necessary to capture the hydrological features of the CW condition by an additional, independent measure: continuous rain.

To capture continuous rainfall, average annual precipitation per month would be a misleading yardstick. The reason is that average annual precipitation correlates strongly with tropical climates and negatively with latitude.[27] and is, thus, untypical for CW-regions. Most tropics have a monsoon season in which the extreme amount of rain is excessive and water, while abundant,

[24] The highest latitudes are 90 degrees at the poles. No country-centroid comes even close to that. In the Northern hemisphere, the highest latitude is obtained by Norway at about 67 degrees. In the Southern hemisphere it is New Zealand at about 42 degrees. We equate Norway's 67 degrees with 1 and standardize all other country-centroids' latitudinal degrees to this maximum.

[25] The country-centroids' latitudes correlate with the countries' annual mean temperatures at $r = -0.89$ (N = 183; p-value = 0.001, 2-tailed).

[26] $r = 0.07$ (N = 183; p-value = 0.336, 2-tailed).

[27] Average annual rainfall per month and latitude correlate at $r = -0.53$ (N = 177; p-value = 0.00, 2-tailed).

is wasted and harmful (e.g., floods). Misleadingly, these extremes inflate the measure of average annual precipitation per month. Thus, the necessary qualification needed to capture the CW condition's typical precipitation pattern is a focus, instead, on whether the rainfall in a region's driest month is high. To capture this feature, we use the typical rainfall level (in cubic millimeters) in a country's driest month.[28] To correct a distribution skewed to the top in this measure, we calculate its square root. Doing so moves extreme outliers at the high end of this measure (i.e., Pacific islands) closer to the center of the distribution. We call this measure henceforth *continuous rain*

The next question is how to combine (1) latitudinal height, (2) summer mildness, and (3) continuous rain. These measures should be combined in a way that best represents in a single indicator the thermo-hydrological configuration typical of the CW condition. Instead of imposing a theoretical solution on this problem, we subject the three measures to an exploratory factor analysis. As it turns out, the three measures reflect two independent dimensions. Continuous rain and mild summers represent one dimension, with factor loadings of 0.80 (mild summers) and 0.77 (continuous rain). High latitudes, by contrast, represent a separate dimension, on which only this measure shows a major loading (i.e., a factor loading of 0.95).[29]

In terms of substance, the first dimension of this factor solution represents *maritime climates*, which capture the *water*-component of the CW condition. The second dimension, by contrast, represents what is unique to high latitudes irrespective of maritime climates, which is *cold winters*. In other words, the second dimension captures the *coolness*-component of the CW condition.

These results show that CW is a condition that combines two independent components into a single configuration. Accordingly, the measurement of CW should represent this pattern and proceed as an additive combination of its two independent components. Following this premise, we calculate for each country its factor score on the first dimension (i.e., the water factor) and on the second dimension (i.e., the coolness factor).[30] The latter represents the coldness of winters independent of maritime climates, for which reason country scores on the first dimension (i.e., the water factor) and the second one (i.e., the coolness factor) are uncorrelated. Hence, we can calculate CW as an additive combination of two independent components by taking the arithmetic mean of the water and the coolness factor.[31]

Should the effects of CW's coolness component and its water component indeed add on each other, this additive combination will capture both effects in a single measure. In other words, if it is really the combination of coolness and water that makes the difference, this additive measure will isolate that effect.

The CW index correlates with two widely-used geographical proxies of development: it correlates negatively with the share of land in tropical and subtropical latitudes, and it correlates positively with the share of land area in the cold-to-temperate zones that lack a dry season, corresponding to climate zones *Cf* and *Df* of the Koeppen–Geiger classification of climate zones (Peel et al., 2007). Empirically, most countries' territory falls entirely within or outside tropical climates, or entirely within or outside *Cf* or *Df* zones. As a result, these two variables are essentially binary variables. In contrast, the CW index varies smoothly and continuously within and outside tropical or cold-to-temperate climate zones, and, thus, measures more fine-grained geo-climatic variation across countries. For visual support of this argument, see Figure A2, which plots the CW index against the share of a country's territory in Koeppen–Geiger *Cf* or *Df* climate zones, and Figure A3, which plots the CW index against the share of ancestral land in tropical or subtropical latitudes. In the Online Appendix, we also show that the CW index outperforms the two alternative variables in our regression models.

We also acknowledge that a key concern with our CW-index relates to differences in country area size. Indeed, scores on the CW-index might not be comparable across countries with different area sizes when bigger size implies higher within-country variability in the CW condition. In the Online Appendix, we deal with this issue.[32] Territorial country size is entirely unrelated to within-country variability in the CW condition. Moreover, only 14 percent of the total variation in the CW condition is within countries, while 86 percent is between countries. Hence, country-mean differences in the CW condition are significant and meaningful as they capture by far most of the existing territorial variation in CW conditions.

For Europe, we find that ages at first marriage for women have persisted over centuries. Fig. 3 displays the positive correlation between historical and contemporary marital ages for women. Countries with higher marital ages for women in 1500–1900 have older brides at first marriage in the postwar period. Moreover, as hypothesized, there is a negative correlation between the husband–wife age gap and the CW condition (Fig. 4). The relationship looks fairly linear; furthermore, it is not just a "European artifact" but holds for different world regions. To test more rigorously if these descriptive correlations are in fact meaningful, we move to a multivariate regression framework which is outlined in the next section.

[28] Continuous rain in this definition is uncorrelated with average annual rainfall per month.

[29] The factor analysis has been conducted across 183 countries for which all three measures are available. The analysis has been conducted under the Kaiser-criterion, advising the extraction of as many factors as there are with Eigenvalues above 1. The factor loadings we report are obtained after a varimax-rotation. The factor solution explains 75 percent of the variance.

[30] We compute regression-based factor scores, i.e., weighted averages of factor loadings, the raw variables of interest, and the inverse of their covariance matrix (Thompson, 1951).

[31] We do this after having normalized the two factor-z-scores for each country to the unit interval.

[32] A previous version of this index (Welzel, 2014) also included the coastline share of a country's borders as a proxy for temperate maritime climates. We discuss this change and compare the two versions in the Online Appendix.

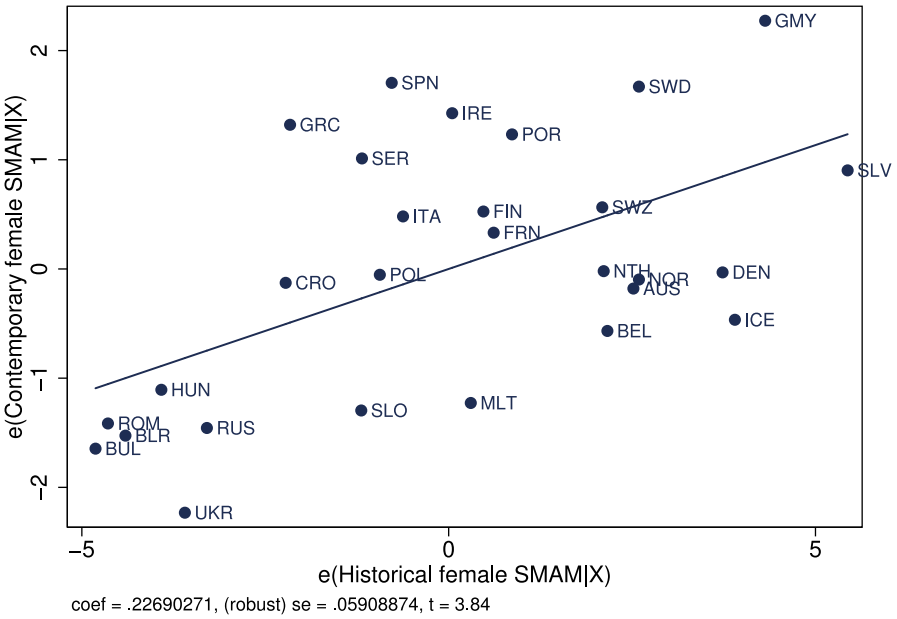


Fig. 3. Historical and contemporary female SMAM for 28 European countries. Notes: Linear correlation controlling for the year of observation of contemporary female SMAM, 1966–2002. Historical data from Dennison and Ogilvie (2014, Table 2); contemporary data from UN (UN, 2009). SMAM is singulate mean age at first marriage.

4. Empirical strategy

We estimate regressions of the form:

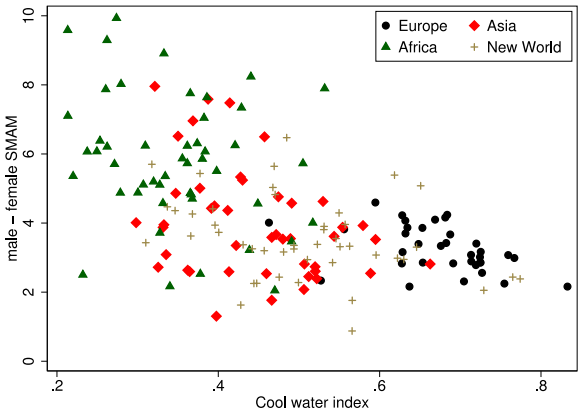
$$y_i = \alpha + \boldsymbol{H}_i'\boldsymbol{\beta} + \boldsymbol{X}_i^{H\prime}\boldsymbol{\gamma} + \boldsymbol{X}_i^{C\prime}\boldsymbol{\delta} + \theta_c + \epsilon_i \tag{1}$$

where y_i is the outcome variable of interest for country i; $\boldsymbol{H}_i$ is a vector of potential deep determinants—*CW-index*, *years of agriculture*, *plow*, and *agricultural suitability*; $\boldsymbol{X}_i^H$ is a vector of historical controls of country i's ancestral population; $\boldsymbol{X}_i^C$ is a vector of contemporary control variables; θ_c is a continent fixed effect. To ensure comparability to the previous literature $\boldsymbol{X}_i^H$ is the same set of historical controls used by Alesina et al. (2013): (1) the presence of large domesticated animals, (2) the number of levels in political hierarchies, and (3) the level of economic complexity proxied by the type of settlement patterns (e.g., nomadic vs. complex settlements). The only exception is that we do not include the proportion of ancestral land that is tropical or subtropical in our baseline specification. As a robustness check, we later estimate additional regressions where tropical climate is included to show that the CW-index is not merely capturing different degrees of exposure to tropical climates.[33] As contemporary controls ($\boldsymbol{X}_i^C$), we also follow previous literature in using the natural log of per capita income and its square, typically referring to the same time period of the outcome variable, y.[34] The inclusion of income levels on the right-hand side of the regression equation raises the issue of endogeneity, to the extent that most of the other historical regressors are thought to partially determine current income levels. However, since we are interested in the persistent effect of deep determinants in contemporary outcomes, it seems natural to condition on contemporary income levels. The empirical exercise then asks the following question: how relevant are deep rooted variables for explaining the share of variation in y that is left unexplained once current income levels have been taken into account?

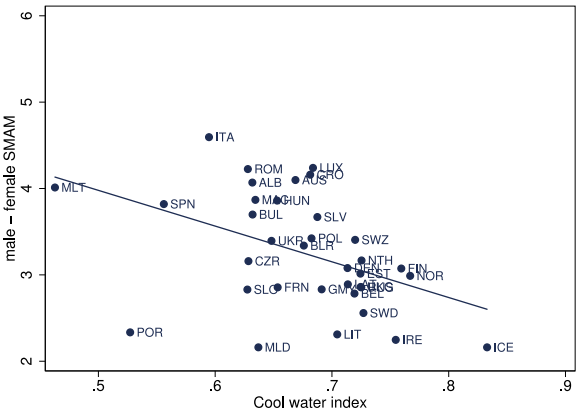
The main potential flaw of our empirical strategy is that the effect of the CW-index might be spurious due to omitted variable bias. Given the cross-sectional nature of our data, we cannot remove time-invariant unobservables by employing country fixed effects. We go at great lengths to convince the reader that our relationships of interest are not driven by third factors. First, we explicitly control for several candidates of such omitted factors in additional regressions. Second, we estimate how large selection on unobservables needs to be relative to selection on observable characteristics in order to fully explain away the CW-effect (Altonji et al., 2005; Oster, 2019). Third, we test the sensitivity of our estimates to subsample selection. In particular, we differentiate between Old versus New World samples to prevent mass migration movements post-1500 from biasing our results (following Olsson and Paik,

[33] We use a large set of additional control variables in our robustness checks. For convenience, these are introduced in the text whenever necessary. For a full list see Table A1.

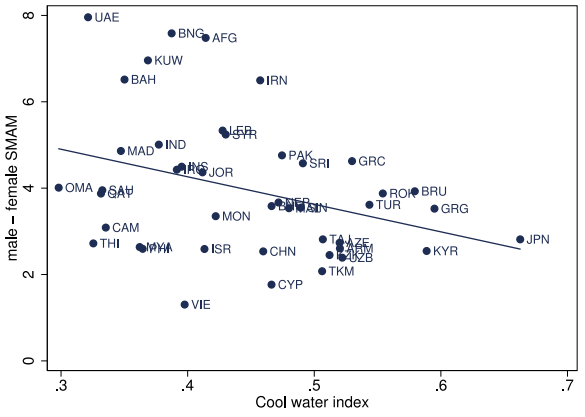
[34] Modeling gender gaps as a quadratic function of per capita income is standard due to the feminization-U hypothesis. The hypothesis argues that women's participation in the labor force first declines with economic development at low income levels, but then increases with per capita income after an income threshold is reached (Boserup, 1970; Goldin, 1995; Mammen and Paxson, 2000). Once modern panel data techniques are applied, however, the evidence for the feminization-U is weak (Gaddis and Klasen, 2014).



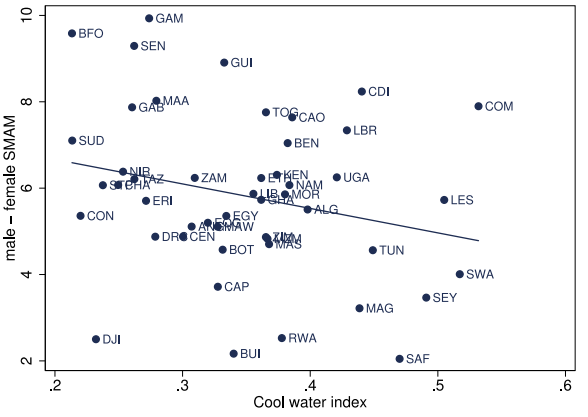
(a) World



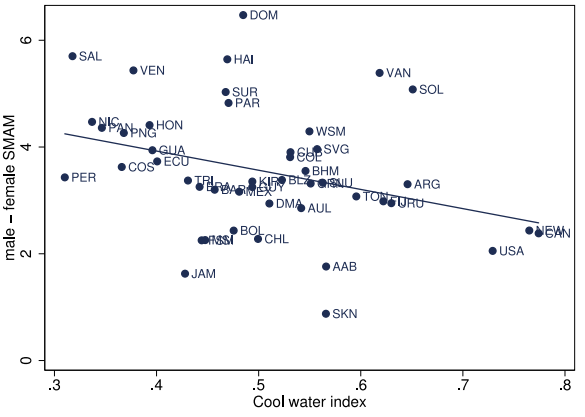
(b) Europe



(c) Asia



(d) Africa



(e) New World

Fig. 4. Cool Water condition and male–female difference in SMAM. *Notes:* Cool Water is the Cool Water index described in Section 3.1. It combines the latitude of a country's centroid, the mean temperature in the hottest month, and mean rainfall in the driest month. Higher values of the index represent areas with cold winters, mild summers, and continuous rainfall. Contemporary ages at first marriages are from the United Nations' *World Marriage Data* (UN, 2009). SMAM is singulate mean age at first marriage. Period is 1960–2006. For each country first year available is shown.

2016, p. 205), and exclude, respectively, Europe and Sub-Saharan Africa from the sample because the former is home to the most extreme historical version of a late marriage pattern and the latter is unique in its prevalence of polygynous marriages. Later on, we will extend our contemporary cross-country setting with evidence from preindustrial Europe, first- and second-generation US immigrants, and regions of the former Russian Empire.

5. Results

5.1. Ages at first marriage and gender equality

We start by establishing that contemporary ages at first marriage between the spouses correlate with gender equality today (Table 2). We use several alternative measures for contemporary ages at first marriage: the female and male SMAM, the ratio of female-to-male SMAM, and the difference between male and female SMAM.[35] The indicators of gender equality are the average female-to-male ratios in labor force participation rates, life expectancy, and years of education for the period 1990–2010.[36] To prevent reverse causality, we use the earliest available year of SMAM data for each country and exclude countries whenever this year is later than 1990. To account for possible worldwide trends in SMAM over time, we include the year of the SMAM observation as a control variable.[37]

As shown in Table 2, countries with older brides and younger grooms have higher female-to-male labor force participation ratios (column 1), higher female-to-male life expectancy ratios (column 4), and higher female-to-male years of education ratios (column 8). These relationships are confirmed by the positive and highly significant effect of the female-to-male SMAM ratio for gender ratios in labor force participation and life expectancy (columns 2 and 5), as well as the negative and highly significant effect of the male to female difference in SMAM for the same outcomes (columns 3 and 6). For gender education ratios, the coefficients have the expected sign but are either statistically insignificant (column 9) or marginally significant (column 8). However, this result is driven by Middle Eastern and North African countries. The MENA region experienced large increases in female education (e.g., in the Gulf States, by allowing women to pursue higher education) without corresponding improvements in labor market participation or ages at first marriage.[38]

Overall, the estimated coefficients of Table 2 have economic relevance. They suggest, for example, that a one-year reduction in the average age difference between the groom and the bride is associated, on average, with a 3.5 percentage point increase in the ratio of female-to-male labor force participation and 0.6 percentage point increase in the ratio of female-to-male life expectancy years.

Consistent with Hansen et al. (2015), longer histories of agriculture are negatively and significantly correlated with gender equality in labor force participation, health, and education. Ancestral plow use is also negatively correlated with gender equality in labor force participation, which replicates the findings of Alesina et al. (2013). It is also the case that historical agricultural suitability is positively associated with higher female participation in the labor force, as hypothesized by Hazarika et al. (2019). However, both the plow and agricultural suitability are uncorrelated to gender equality in life expectancy and education.

Furthermore, higher SMAM for women (and lower SMAM for men) positively and significantly correlates with additional measures of gender equality such as the share of firms with some degree of female ownership, the share of parliament seats held by women, and the UNDP's Gender Development Index, as shown in Table A4.

We explore the sensitivity of these results to subsample analyses (Table A5). Excluding Sub-Saharan Africa, the Americas and Oceania (i.e., the New World), or Europe does not affect the overall finding that higher ages at first marriage for women and lower ages at first marriage for men are positive and significant correlates of gender equality in labor force participation and life expectancy.

5.2. Cool water breeds late-marriage societies

Having shown that ages at first marriage are indeed important factors for gender equality, broadly defined, we now test if the Cool Water condition is a relevant determinant of gender gaps in ages at first marriage. We estimate the baseline specification, with the alternative SMAM variables on the left-hand side and include the CW-index as a new explanatory variable.

Table 3 shows the results. As hypothesized, the CW-index has a positive and significant effect on female SMAM and the ratio of female-to-male SMAM and a negative effect on the difference between male and female SMAM. These effects persist once the other

[35] Note that these three measures have different interpretations. When including both the female and male SMAM as separate regressors, their coefficients estimate the effect on y of one additional SMAM, on average, for gender X, holding the SMAM for gender Y constant. When using the ratio between female and male SMAM, one implicitly weights the age differences between the spouses in the inverse proportion of their age levels. When using the simple difference between male and female SMAM, one weights age differences equally, irrespective of the average age level of the spouses at first marriage.

[36] We average the dependent variables over a 20-year period because a single year might be unrepresentative of the actual cross-sectional differences between countries. The results are robust to using the dependent variables for 2000, instead of averaging between 1990–2010 (Table A2).

[37] As a robustness check, we run the analyses taking the earliest data point for the period 1985–1994. By reducing the time window considerably, we can avoid potential problems of inter-temporal comparison. While our sample size is reduced, the results do not change qualitatively (Table A3).

[38] If we exclude the MENA countries for the regression in Table 2, column 9, the effect of male-to-female SMAM difference increases in absolute magnitude by a factor of 1.7 and becomes highly significant: $\hat{\beta}_{w/o\ MENA} = -0.024$, robust s.e. $= 0.009$, p-value $= 0.008$. Including the MENA region, but measuring the ratio of female-to-male education in 1990 (instead of the 1990–2010 average) also produces a larger coefficient (in absolute terms) than that of column 9: $\hat{\beta}_{1990} = -0.018$, robust s.e. $= 0.008$, p-value $= 0.028$.

Table 2
Determinants of gender gaps: ages at first marriage.

	Average female–male ratio in 1990–2010								
	Labor force participation			Life expectancy			Years of education		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ages at first marriage:									
Female	0.037***			0.006***			0.017*		
	(0.010)			(0.002)			(0.009)		
Male	−0.032***			−0.006***			−0.005		
	(0.012)			(0.002)			(0.010)		
Female/male		1.014***			0.164***			0.425*	
		(0.269)			(0.043)			(0.245)	
Male–female			−0.035***			−0.006***			−0.014
			(0.010)			(0.002)			(0.009)
Year of obs.	−0.002	−0.002	−0.002	0.001	0.000	0.001	−0.003	−0.003	−0.002
	(0.003)	(0.003)	(0.003)	(0.000)	(0.000)	(0.000)	(0.003)	(0.003)	(0.003)
Deep determinants:									
Years of agriculture	−0.041***	−0.040***	−0.041***	−0.004**	−0.004**	−0.004**	−0.024**	−0.025**	−0.026**
	(0.010)	(0.010)	(0.010)	(0.002)	(0.002)	(0.002)	(0.010)	(0.010)	(0.011)
Plow	−0.096*	−0.095*	−0.099*	−0.004	−0.002	−0.003	0.023	0.017	0.015
	(0.052)	(0.052)	(0.052)	(0.009)	(0.009)	(0.009)	(0.098)	(0.096)	(0.096)
Agricultural suitability	0.115**	0.118**	0.118**	−0.001	−0.001	−0.001	−0.046	−0.044	−0.045
	(0.052)	(0.053)	(0.052)	(0.010)	(0.010)	(0.010)	(0.071)	(0.069)	(0.069)
Historical controls:									
Large animals	0.018	0.015	0.023	−0.020	−0.022	−0.020	−0.190*	−0.187*	−0.186*
	(0.081)	(0.083)	(0.081)	(0.020)	(0.020)	(0.020)	(0.104)	(0.106)	(0.106)
Political hierarchies	−0.005	−0.006	−0.005	0.001	0.001	0.001	0.006	0.007	0.008
	(0.020)	(0.020)	(0.020)	(0.003)	(0.003)	(0.003)	(0.016)	(0.016)	(0.016)
Economic complexity	0.012	0.012	0.013	−0.002	−0.002	−0.002	−0.011	−0.010	−0.010
	(0.013)	(0.013)	(0.013)	(0.002)	(0.002)	(0.002)	(0.014)	(0.014)	(0.014)
Tropical climate	−0.145***	−0.140***	−0.139***	−0.024***	−0.025***	−0.025***	−0.048	−0.036	−0.036
	(0.050)	(0.047)	(0.048)	(0.009)	(0.008)	(0.008)	(0.048)	(0.048)	(0.048)
Contemporary controls:									
Income per capita (log)	−0.547***	−0.557***	−0.527***	0.137***	0.130***	0.135***	0.588***	0.618***	0.629***
	(0.159)	(0.158)	(0.160)	(0.033)	(0.030)	(0.031)	(0.192)	(0.183)	(0.183)
(Income per capita (log))^2	0.028***	0.029***	0.027***	−0.008***	−0.008***	−0.008***	−0.030***	−0.031***	−0.031***
	(0.009)	(0.009)	(0.009)	(0.002)	(0.002)	(0.002)	(0.011)	(0.011)	(0.011)
Continent dummies:	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	131	131	131	132	132	132	119	119	119
R^2	0.581	0.585	0.580	0.571	0.574	0.571	0.630	0.627	0.625
adj. R^2	0.518	0.527	0.521	0.507	0.515	0.511	0.568	0.568	0.566

Notes: OLS estimates are reported with robust standard errors in parentheses. "Ages at first marriage" are singulate mean years at first marriage (SMAM) from UN (2009) for the period 1960–1990. For each country, earliest year available is selected and controlled for with variable "Year of obs.". "Years of agriculture" is the number of years (in thousands) since the Neolithic revolution (from 1500) from Putterman and Trainor (2006). "Plow" is the proportion of population with ancestors that used the plow in preindustrial agriculture from Alesina et al. (2013). "Agricultural suitability", also from Alesina et al. (2013), is the share of ancestral land suitable for the cultivation of barley, wheat, sorghum, rye, foxtail millet, or pearl millet. *Historical controls* are from Alesina et al. (2013). The natural log of per capita income and its square are measured in the same time period as the dependent variable. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

deep determinants are included. In fact, the CW-impact on the gender differences becomes *larger* in absolute terms. The estimates in column 8, for example, imply that a one standard deviation increase in the CW-index is associated with a 10-month reduction of the average gap between male and female ages at first marriage.[39] This is a sizeable effect: 10 months corresponds to roughly 20 percent of the world's average gender gap in ages at first marriage, which is 4.12 years in the period considered (see Table 1). In contrast, none of the other deep determinants—whether years of agriculture, historical plow use, or historical agricultural suitability—are significant at the 5 percent level, once the CW-index is included. Interestingly, the ages at first marriage for women and men increase with per capita income (averaged over the period 1960–1980), following an inverted U-shaped function that peaks around 13,000–13,500 PPP-$. But there is no evidence that women's SMAM approaches men's as countries get richer, since the income coefficients for the female–male age ratio or the male–female age difference are not statistically significant.[40] These results suggest that while marital ages do respond to economic development they do so similarly for *both* sexes, thus leaving ratios and differences unaffected. Over time, as income levels rise, there is no evidence of convergence between female and male ages at first marriage, which supports the view that persistent, deep-rooted patterns dominate this relationship. The results hold for the usual three subsamples: without Sub-Saharan Africa, without the New World, and without Europe, as shown in Table A6.

[39] $\hat{\beta}_{cw} * \sigma_{cw} = -5.902 * 0.145 \approx -0.856$ (years (≈ -10.27 months).

[40] The coefficients for the year of the SMAM observation tell a similar story: while ages at first marriage have increased over time for both women and men (columns 2 and 4), the time trends for the ratio or differences between the sexes are not statistically significant.

Table 3
Determinants of ages at first marriage.

	Singulate mean age at first marriage							
	Female		Male		Female/male		Male–female	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deep determinants:								
Cool Water	6.727***	4.316*	1.949	−1.555	0.195***	0.221***	−4.672***	−5.902***
	(2.546)	(2.494)	(2.423)	(2.402)	(0.047)	(0.075)	(1.302)	(2.078)
Years of agriculture		−0.312**		−0.159		−0.007*		0.154*
		(0.147)		(0.112)		(0.004)		(0.092)
Plow		−1.105		−0.253		−0.034		0.890
		(1.025)		(0.860)		(0.026)		(0.686)
Agricultural		0.310		0.728		−0.019		0.523
suitability		(0.906)		(0.824)		(0.024)		(0.646)
Historical controls:								
Large animals	1.074	0.731	0.694	0.336	0.020	0.019	−0.361	−0.395
	(1.121)	(1.363)	(1.035)	(1.096)	(0.023)	(0.035)	(0.575)	(0.877)
Political hierarchies	0.658***	0.753**	0.329	0.254	0.013**	0.020**	−0.303*	−0.482**
	(0.245)	(0.295)	(0.203)	(0.259)	(0.006)	(0.008)	(0.163)	(0.218)
Economic complexity	−0.242	−0.061	−0.030	0.064	−0.007	−0.003	0.190	0.099
	(0.156)	(0.121)	(0.106)	(0.110)	(0.005)	(0.004)	(0.128)	(0.112)
Contemporary controls:								
Income p.c. (log)	5.726**	6.092***	6.471***	6.632***	−0.002	0.010	0.918	0.759
	(2.369)	(2.113)	(1.951)	(1.807)	(0.062)	(0.061)	(1.625)	(1.631)
(Income p.c. (log))^2	−0.307**	−0.321***	−0.350***	−0.349***	0.000	−0.001	−0.052	−0.039
	(0.137)	(0.119)	(0.115)	(0.102)	(0.003)	(0.003)	(0.092)	(0.092)
Year of SMAM obs.	0.080***	0.059***	0.061***	0.050**	0.001**	0.001	−0.025	−0.015
	(0.022)	(0.022)	(0.020)	(0.021)	(0.001)	(0.001)	(0.016)	(0.016)
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	138	125	134	121	134	121	134	121
R^2	0.459	0.433	0.312	0.253	0.495	0.517	0.468	0.498
adj. R^2	0.407	0.355	0.244	0.146	0.445	0.448	0.415	0.426

Notes: OLS estimates are reported with robust standard errors in parentheses. "Singulate mean age at first marriage" data are from UN (2009) for the period 1960–2006. For each country, earliest year available is selected and controlled for with variable "Year of SMAM obs." "Cool Water" is the Cool Water index described in Section 3.1. It combines the latitude of a country's centroid, the mean temperature in the hottest month, and mean rainfall in the driest month. Higher values of the index represent areas with cold winters, mild summers, and continuous rainfall. "Years of agriculture" is the number of years (in thousands) since the Neolithic revolution (from 1500) from Putterman and Trainor (2006). "Plow" is the proportion of population with ancestors that used the plow in preindustrial agriculture from Alesina et al. (2013). "Agricultural suitability", also from Alesina et al. (2013), is the share of ancestral land suitable for the cultivation of barley, wheat, sorghum, rye, foxtail millet, or pearl millet. *Historical controls* are from Alesina et al. (2013). The natural log of per capita income and its square are averaged over the period 1960–1980. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

In Table A7, we unpack the CW index by replacing it in the regressions with its constitutive components. All three individual variables of the CW-index (absolute latitude, mild summers, and continuous rain) have a negative effect on gender differences in marital ages, as hypothesized (although the coefficient for continuous rain is not statistically significant). The two index components (*Coolness* and *Water*) also have the hypothesized negative effects when regressed separately, although the *Coolness* estimate is small and statistically insignificant. However, when both components are included simultaneously in the regression, they both have larger and highly significant coefficients. We interpret these results as suggesting that it is indeed the combination of the two orthogonal components that matters the most, and not their independent effects. Consistent with this interpretation, adding an interaction term between the two components does not change their coefficient magnitudes, and the interaction term itself is statistically insignificant. In a nutshell, the overall effect of the CW condition is not driven by one of its particular components but is truly the result of their combination, as depicted in our summary CW-index.

We have thus far a negative correlation between the CW condition and gender gaps in ages at first marriage in the postwar period. Our interpretation is however that this present-day correlation emerges from the effect of geo-climatic conditions on *preindustrial* marriage patterns. Over time, these marriage patterns became embedded in both formal and informal institutions, thereby persisting until today. Although we cannot empirically exclude that the CW condition has a direct contemporary effect of ages at first marriage, we can ask whether the CW effect is mainly tied to a geographical area *as such* or to its ancestral inhabitants (Putterman and Weil, 2010). Hansen et al. (2015) show that the negative effect of longer histories of agriculture on gender equality becomes stronger after weighing their variable on the timing of the Neolithic revolution with post-1500 migration flows from Putterman and Weil (2010). In the same spirit, we create an alternative version of the CW index weighted by Putterman and Weil's World Migration Matrix data.[41] Adjusting for ancestry strengthens the CW effect (see Table A8 for point estimates), suggesting that the historical CW condition of a population matters more than the CW condition of its present-day place of residency. This is consistent with our hypothesis

[41] The correlation coefficient of the CW indexes adjusting or not for post-1500 migration is 0.96. For comparison, the correlation between the adjusted and unadjusted years of agriculture variable used by Hansen et al. (2015) is 0.85.

that the results reveal the long-run persistence of a preindustrial relationship, rather than a contemporary one, between a society's geo-climatic configuration and its household formation patterns. However, adjusting the CW index in this way is problematic if the migration flows are endogenous to marriage patterns, as in the case where areas with favorable CW conditions were to attract immigrants with strong preferences for late marriage patterns of household formation. Indeed, Northwestern Europeans have largely settled in the regions of the New World with the highest score of the CW index. The unadjusted CW measure is free from this specific source of endogeneity bias. Thus, even though the estimated effects are stronger with the ancestry-adjustment, we decide, as a matter of caution, to present the remaining results without this adjustment.

5.2.1. Selection on observables

It is important to note that if, by construction, the CW index would uniquely fit the geo-climatic features of Northwestern Europe and its New World colonies, then it would necessarily be also correlated with all the unobservable factors that might explain these regions' (potentially) unique preindustrial late marriage pattern. If this were the case, our results would be completely spurious. It is therefore essential to refute this possibility. First, we use a data-driven approach (factor analysis) for the construction of the CW index in order to minimize the concern that the world distribution of this variable results from *ad hoc* weighting decisions. Second, we show that our results are not dependent on Northwestern Europe or Western offshoots. Both including dummy variables for these groups of countries or excluding them altogether from the estimation sample does not affect the main result: a highly significant and negative effect of the CW index, only slightly weaker in magnitude (Table A9).[42] In other words, the association of the CW index with ages at first marriage is not a spurious idiosyncrasy of Northwestern Europe and its offshoots, but a broader relationship that holds for the rest of the globe.[43]

The inclusion of further control variables does not affect the relationship between the CW-index and the average age gap between groom and bride (Table 4). We start by including tropical climate in column 2 of Table 4, since it could be that the relevant variation captured by the CW-index is that between tropical and non-tropical countries. The results show otherwise; the coefficient for CW remains negative and highly significant, whereas the tropical climate coefficient is statistically indistinguishable from zero. This demonstrates that the CW-index is more than just an inverse measure of tropical temperatures. What distinguishes it from such an inverse measure is that it gives a premium not just on high latitude but more specifically on high latitude with minimized seasonal extremity and continuous rain.[44]

Another possible source of error is that the CW-index captures the fact that European colonizers settled by and large in all the major CW-areas outside Europe, with the exception of Japan and the Korean peninsula. Thus, it could simply be that settlers from late marriage European societies "exported" the late marriage pattern to their overseas offshoots. Even though the subsample analyses of Tables A6 and A9 do not support this argument, we provide further evidence by including the weighted genetic distance between each country and the United Kingdom from Spolaore and Wacziarg (2009).[45] If the CW-index was indeed a mere proxy for areas of European settlement around the world, we would expect the relationship between CW and ages at first marriage to vanish once the genetic distance from Western Europe is held constant. However, as seen in column 3, the coefficient of CW remains negative, statistically significant, and, if anything, the effect becomes *stronger*. On column 4, we go beyond controlling for contemporary income difference across countries and also control for preindustrial differences in the level of development. Following the literature, we use population density in 1500 as a proxy for development in the Malthusian era (e.g., Spolaore and Wacziarg, 2013, footnote 3). The CW-effect does not change.

Alternatively, it could be the case that per capita income levels do not reasonably proxy other developmental dimensions that might be driving the correlation between the CW index and ages at first marriage. In particular, education levels and formal institutions are plausible candidates for such omitted factors. More educated individuals marry later, and better formal institutions could be stronger at enforcing minimum-marital-ages legislation, or recognizing individual consent as the basis for a lawful marriage. In columns 5–6 we include, respectively, the mean years of education in 1950 for the total population and by gender; in columns 7–8, we include the *polity2* score in 1980 as a measure of democracy and the World Bank's rule of law variable in 2000 (Kaufmann et al., 2011) as a measure of institutional quality. The CW coefficient remains negative and significant, and of comparable magnitude throughout.

Religion poses a particular challenge. While certain authors regard religion as a crucial determinant of gender norms (e.g., Fish, 2002; Inglehart and Norris, 2003; Lagerlöf, 2003; Carmichael, 2011), others have argued that religion is endogenous to pre-existing factors. As such, religion would be a "bad control" to include since it would shut down important transmission mechanisms from the deep determinants to the outcome of interest. Boserup (1970) recognized this problem in her original plow vs. shifting agriculture argument. In particular, she claims that the use of the veil or the burqa was a direct consequence of female domestic seclusion due to plow agriculture; only afterwards did it become incorporated in the religious practices of those societies. Being

[42] Northwestern Europe includes: Belgium, Denmark, France, Germany, Iceland, Ireland, Netherlands, Norway, Sweden, and the United Kingdom. Excluding France from this list has no impact on the results. Western offshoots are Australia, Canada, New Zealand, and the United States.

[43] A somewhat similar concern is that the CW index is potentially noisier for large countries. Reassuringly, we find that large countries do not drive the CW effect. The CW estimate remains stable when the largest 10 countries in landmass area are excluded from the sample (Table A9, column 6).

[44] Similarly, the CW-index 'outperforms' the Koeppen–Geiger *Cf* and *Df* variables. As shown in Table A10, the coefficient of the CW-index remains highly significant and roughly of the same size, whereas the coefficient estimates for the Koeppen–Geiger variables, in turn, despite having the expected sign (columns 2 and 4), are imprecise and statistically insignificant.

[45] The weighted genetic distance is the expected value of the genetic distance between two randomly picked individuals for each pair of countries. See more details in Spolaore and Wacziarg (2009).

a consequence of the plow, religion would mistakenly absorb much of the plow's gender-inegalitarian effect in any regression of gendered outcomes. Alesina et al. (2013) find indeed that including religion reduces the effect of the plow by 20 percent.

In general, this strand of argument claims that emerging religions absorbed, incorporated and codified many pre-existing local practices and beliefs, rather than having introduced them. Hartman (2004) argues that, in medieval Europe, the Catholic doctrine of individual consent being a sufficient condition for the validity of marriage was widely followed in the Northwestern societies but rarely so in (deeply Catholic) Italy, Spain, and Southern France. Similarly, Hansen et al. (2015, p. 378) and Hazarika et al. (2015, pp. 19–20) discuss how pre-existing gender norms influenced early Islamic doctrine.

Despite the controversy on whether religion is a "bad control", we report results controlling for two sets of religion variables. First, we control for the length of a country's exposure to the Roman Catholic (Western) Church and to the Orthodox (Easter) Church. These variables are from Schulz et al. (2019), who present evidence supporting Goody's (1983) and Henrich's (2020) hypothesis that the origins of the European Marriage Pattern lay in medieval family policies enacted by the Western Church. While the CW-effect remains fully robust, the variables capturing exposure to medieval churches are statistically insignificant (column 9). Second, we include the population shares of Catholics, Protestants, and Muslims in 1980 (with other religions as the reference group). Protestantism has been associated with higher literacy rates and human capital (Becker and Woessmann, 2009), reduced gender gaps in educational attainment (Becker and Woessmann, 2008), and decreased fertility (Becker et al., 2010). Modern Islam, on the other hand, has been associated with more conservative gender roles and gender inequality in many dimensions (e.g., Fish, 2002; Inglehart and Norris, 2003). Consistent with previous literature, column 10 shows that, relative to other religious groups, a large Protestant share is associated with smaller age gaps between the spouses, whereas a larger Muslim share correlates with a wider age gap. Once the religious shares are included, the CW effect remains negative and statistically significant but, as expected, about 32 percent smaller in magnitude.

Finally, we add a country's per capita oil production in 2000 to capture Ross's (2008) argument that oil endowments crowd-out women from the labor force. The low employment prospects and, consequently, low returns to education for women could incentivize early female ages at first marriage. Indeed, per capita oil production is positively correlated with larger age gaps between the groom and the bride, but its coefficient is insignificant. In any case, the CW effect remains unchanged by the inclusion of this additional control (column 11). Even when we include all the additional controls simultaneously in column 12, the negative correlation between the CW-index and the male-to-female difference in age at first marriage remains significant at the 5 percent level.

In addition to the regressions of Table 4, we perform further robustness checks and present the results in Table A11 of the Online Appendix. As additional historical controls, we include a measure for the preindustrial intensity of agriculture and the proportion of ancestral subsistence provided by animal husbandry (taken from Alesina et al., 2013)[46] to account for the possibility that animal husbandry delays ages at marriage for women (Voigtländer and Voth, 2013). We also add the proportion of ancestral subsistence provided by hunting since hunter–gatherer societies display higher levels of gender equality (Dyble et al., 2015). To test the idea that male dominance over women derives from the emergence of private property (Engels, 1902), we also include the share of ancestors from ethnicities where land inheritance rules were absent. Furthermore, as additional contemporary controls, we include two warfare variables: both the number of years of civil and inter-state conflict for each country from 1816 until 2007 and the terrain ruggedness index from Nunn and Puga (2012). The latter is included because flatter regions are easier to invade but also easier to irrigate and plow than rugged terrain, and may also experience less rainfall than mountainous regions. War could either be detrimental for women if it reinforces gender violence and patriarchy in society, but it could also have positive effects if women are called to replace men in the labor force, thus postponing marriage (Whyte, 1978). A more direct effect of war is to reduce the supply of young men in the marriage market leading to later marital ages or higher proportions of never married women. To complement the genetic distance variable and the World Migration Matrix in measuring post-1500 global migratory flows, we add the share of a country's population (in 2000) that is of Western European descent. We also include a communist dummy since communist regimes had explicit policies to promote gender equality and, in some cases, fought traditional marriage practices such as arranged marriages, or child marriages. Finally, we add the share of GDP accruing to agriculture, manufacturing, or services in 2000 since labor demand in female-dominated sectors will likely impact female marital ages (e.g., Ross, 2008).

Controlling for these additional variables, both in a stepwise manner or simultaneously, does not affect our main result: the CW coefficient is always statistically significant at least at the 5 percent level and its size ranges from −5.206 to −6.296 (Table A11).[47] Overall, after controlling for many additional variables, we confirm our baseline finding that countries with a stronger CW condition have systematically lower age gaps at first marriage between spouses.

5.2.2. Selection on unobservables

Finally, we estimate how large the ratio (δ) of selection on unobservables relative to selection on observables needs to be in order to explain away the CW effect (Altonji et al., 2005; Oster, 2019). Table A15 presents the estimates for δ, using the method described in Oster (2019).[48] Altonji et al. (2005) propose $\delta = 1$ as an *ad hoc* cutoff for robustness. For reasonable upper bounds on the R^2 and models that exclude religious variables—as recommended by most of the literature—, the estimates of δ are close to the cutoff, ranging from 0.81 to 1.01. In sum, for omitted variable bias to fully explain away the CW effect, selection on unobservables would have to be 81 to 101 percent as large as selection on observables in the most restricted models. Thus, it seems unlikely that omitted variable bias alone could account for our results.

[46] Unless otherwise noted, all additional variables included are from Alesina et al. (2013). For the original source and construction method see their Online Appendix.

[47] Re-estimating the specifications of Tables 4, A6, and A11 using ancestry-adjusted CW index and years of agriculture produces even stronger CW effects, which are always highly statistically significant. Results are shown in Tables A12, A13, and A14.

[48] See section A.2 of the Online Appendix for more details.

Table 4
Determinants of ages at first marriage: additional controls.

	Singulate mean ages at first marriage: male–female											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Cool Water	−5.902***	−6.054**	−6.041***	−5.921***	−5.291**	−5.335**	−6.450***	−5.781***	−6.570***	−4.016**	−5.464**	−6.120***
	(2.078)	(2.315)	(2.087)	(2.051)	(2.395)	(2.404)	(2.372)	(2.056)	(2.235)	(1.787)	(2.173)	(2.304)
Years of agriculture	0.154*	0.155*	0.068	0.133	0.095	0.090	0.141	0.116	0.144	−0.071	0.149*	−0.133
	(0.092)	(0.093)	(0.109)	(0.094)	(0.099)	(0.105)	(0.100)	(0.097)	(0.094)	(0.098)	(0.090)	(0.093)
Plow	0.890	0.892	0.687	1.032	0.867	0.881	0.999	1.064	1.053	0.431	0.877	1.468*
	(0.686)	(0.689)	(0.734)	(0.715)	(0.766)	(0.786)	(0.743)	(0.671)	(0.713)	(0.669)	(0.679)	(0.737)
Agricultural suitability	0.523	0.523	0.556	0.362	0.606	0.621	0.551	0.699	0.636	0.685	0.574	0.699
	(0.646)	(0.649)	(0.643)	(0.667)	(0.716)	(0.723)	(0.656)	(0.628)	(0.675)	(0.572)	(0.650)	(0.635)
Tropical climate		−0.091										−0.845
		(0.443)										(0.546)
F_{ST} from U.K.			−5.727									4.488
(weighted)			(3.534)									(4.706)
Population density				0.018								0.021*
in 1500				(0.014)								(0.012)
Total years of schooling					−0.327							−0.028
in 1950 (log)					(0.250)							(0.242)
Male years of schooling						−0.147						
in 1950 (log)						(0.352)						
Female years of						−0.154						
schooling in 1950 (log)						(0.248)						
Polity2 in 1980							−0.011					0.047*
							(0.024)					(0.026)
Rule of law in 2000								−0.360				−0.551*
								(0.239)				(0.330)
Exposure to Medieval Churches:												
Western Church									−0.016			0.157
									(0.073)			(0.109)
Eastern Church									−0.071			0.047
									(0.143)			(0.208)
Catholic share in 1980										−0.009*		−0.007
										(0.005)		(0.008)
Protestant share										−0.017**		−0.012
in 1980										(0.007)		(0.010)
Muslim share in 1980										0.022***		0.027***
										(0.006)		(0.008)
Oil production											1.057	2.173*
(per capita)											(1.038)	(1.146)
Historical & contemporary controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	121	121	120	119	109	109	113	121	113	116	121	97
R^2	0.498	0.498	0.514	0.516	0.508	0.509	0.492	0.511	0.494	0.630	0.503	0.681

Notes: OLS estimates are reported with robust standard errors in parentheses. "Singulate mean age at first marriage" data are from UN (2009) for the period 1960–2006. For each country, earliest year available is selected and controlled for with variable "Year of SMAM obs." "Cool Water" is the Cool Water index described in Section 3.1. It combines the latitude of a country's centroid, the mean temperature in the hottest month, and mean rainfall in the driest month. Higher values of the index represent areas with cold winters, mild summers, and continuous rainfall. "Years of agriculture" is the number of years (in thousands) since the Neolithic revolution (from 1500) from Putterman and Trainor (2006). "Plow" is the proportion of population with ancestors that used the plow in preindustrial agriculture from Alesina et al. (2013). "Agricultural suitability", also from Alesina et al. (2013), is the share of ancestral land suitable for the cultivation of barley, wheat, sorghum, rye, foxtail millet, or pearl millet. All regressions include the same set of historical and contemporary controls as in Table 3. *Additional controls:* "Tropical climate" is from Alesina et al. (2013), "F_{ST} weighted genetic distance to the U.K." is from Spolaore and Wacziarg (2009), "Population density in 1500" is from Klein Goldewijk et al. (2010), years of education in 1950 are from Barro and Lee (2013), "Polity in 1980" is the polity2 score from the Center for Systemic Peace, "Rule of law in 2000" is the rule of law indicator from the World Bank's World Governance Indicators (Kaufmann et al., 2011), "Exposure to Medieval Churches" are measured in centuries of exposure and taken from Schulz et al. (2019), "Religious shares in 1980" are the shares of the population of different religions from La Porta et al. (1999), "Oil production (per capita)" is the number of barrels produced per person per day in 2000 from Alesina et al. (2013). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5.3. Cool Water and historic late marriages

5.3.1. Preindustrial Europe

We now turn to Europe to show that the relationship between the Cool Water condition and ages at first marriage is deep rooted in history. Between 1500 and 1900, across 27 European countries, the CW-index is positively associated with female marital ages (Table 5). Despite the much smaller sample, the CW coefficient continues to be statistically significant in all specifications, except

Table 5
Europe: historical female ages at first marriage.

	Historical female age at first marriage, 1500–1900							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deep determinants:								
Cool Water	14.424**	20.872***	13.733	18.809	12.814**	19.033**	12.873**	18.894***
	(5.175)	(6.894)	(10.357)	(11.132)	(5.708)	(7.291)	(5.443)	(6.492)
Years of agriculture			0.105	−0.005				
			(0.703)	(0.643)				
Plow					−1.668	−2.028		
					(1.168)	(1.206)		
Agricultural suitability							−2.988*	−4.361**
							(1.728)	(1.613)
Preindustrial development:								
Population density in 1500		0.091***		0.093***		0.093***		0.098***
		(0.032)		(0.032)		(0.033)		(0.031)
N	27	26	26	25	27	26	27	26
R^2	0.152	0.299	0.105	0.267	0.161	0.313	0.187	0.374
adj. R^2	0.118	0.238	0.027	0.163	0.091	0.219	0.119	0.289

Notes: OLS estimates are reported with robust standard errors in parentheses. "Historical female age at first marriage" data are from Dennison and Ogilvie (2014), see more details in Section 3. Countries included: Austria, Belarus, Belgium, Bulgaria, Croatia, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Malta, Netherlands, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, and Ukraine. "Cool Water" is the Cool Water index described in Section 3.1. It combines the latitude of a country's centroid, the mean temperature in the hottest month, and mean rainfall in the driest month. Higher values of the index represent areas with cold winters, mild summers, and continuous rainfall. "Years of agriculture" is the number of years (in thousands) since the Neolithic revolution (from 1500) from Putterman and Trainor (2006). "Plow" is the proportion of population with ancestors that used the plow in preindustrial agriculture from Alesina et al. (2013). "Agricultural suitability", also from Alesina et al. (2013), is the share of ancestral land suitable for the cultivation of barley, wheat, sorghum, rye, foxtail millet, or pearl millet. "Population density in 1500" is from Klein Goldewijk et al. (2010). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

when years of agriculture are included in columns 3 and 4. In those specifications, while the size of the CW effect is not much affected, its standard errors increase due to multicollinearity between the CW-index and the timing of the Neolithic revolution ($r \approx -0.70$) for this sample of European countries.

Importantly, once again the results are not a statistical artifact driven by more developed areas being located in CW regions. The CW effect is robust to the inclusion of population density in 1500 as a proxy for preindustrial development. While it is true that societies with higher population densities had, on average, older brides at first marriage, controlling for this relationship actually increases the estimates of the CW variable. The reason is that most societies with extremely high CW scores (e.g., Iceland, Sweden, Norway, and Denmark) had relatively lower population densities and were *not* among the wealthiest and most developed nations of Europe in this period (Dennison and Ogilvie, 2014).

Moreover, the positive and significant relationship between the CW-index and historical female ages at first marriage holds for a nineteenth century sample of both European and non-European countries, using data from Gapminder (Table A16). In this setting, the CW-effect retains significance, even taking the timing of the Neolithic Revolution into account.[49]

5.3.2. First and second-generation immigrants in the US, 1880–1930

We now provide evidence, at the individual level, that the Cool Water condition of the country of origin correlates with an immigrant's age at marriage. We use data from five US censuses: 1880, 1900, 1910, 1920, and 1930.[50] We chose these censuses because they are the only to include mother's and father's country of birth, years since immigration, and marital status. We construct two samples of women aged 15–39: (1) immigrants to the US (first generation), and (2) US-born children of immigrants, defined by either paternal or maternal foreign birthplace (second generation). We then estimate two linear probability models. For the sample of first-generation immigrants, the linear probability model is

$$M_{istb} = \alpha + \beta_1 CW_b + \beta_2 CW_b \times \text{Years in US}_{istb} + \beta_3 \text{Years in US}_{istb} + \boldsymbol{X}_{\boldsymbol{b}}^{\boldsymbol{CH}\prime}\boldsymbol{\delta} + \boldsymbol{D}_{\boldsymbol{istb}}'\boldsymbol{\gamma} + \eta_s + \omega_t + \epsilon_{istb} \quad (2)$$

where M is a ever-married dummy for woman i, living in US state s in census-year t, born in foreign country b. CW is the Cool Water index of the immigrant's country of birth b, and 'Years in US' are the number of years since the woman migrated to the US. $\boldsymbol{X}^{\boldsymbol{CH}}$ includes the same set of historical and contemporary controls of the cross-country regressions, this time defined at the immigrant's country of birth b. In addition, we control for individual characteristics $\boldsymbol{D}$ (age, age squared, literacy fixed effects), US state fixed effects (η_s), and census-year fixed effects (ω_t).

[49] Once again, the results are robust to ancestry-adjusting the CW index and years of agriculture (Table A17).

[50] All census samples are from IPUMS-US. We use the following samples: 10 percent sample (1880), 5 percent sample (1900, 1930), and 1 percent sample (1910, 1920).

Table 6
Marriage probability of immigrants in the US.

	Female immigrants: 1880–1930					
	(1)	(2)	(3)	(4)	(5)	(6)
Age group:	15–39	15–19	20–24	25–29	30–34	35–39
Cool Water	−1.928**	−1.096***	−3.135***	−2.298**	−1.232**	−0.825
	(0.771)	(0.211)	(1.073)	(0.948)	(0.597)	(0.548)
Cool Water × Years since immigration	0.034***	0.026***	0.040***	0.032**	0.016***	0.004
	(0.010)	(0.006)	(0.015)	(0.013)	(0.004)	(0.004)
Years since immigration	−0.022***	−0.019***	−0.028***	−0.021**	−0.010***	−0.002
	(0.006)	(0.004)	(0.010)	(0.008)	(0.003)	(0.003)
Other deep determinants	Yes	Yes	Yes	Yes	Yes	Yes
Ancestry-country controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Census-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	295,864	29,598	53,128	65,705	70,616	76,817
R^2	0.370	0.129	0.151	0.098	0.059	0.047
adj. R^2	0.370	0.127	0.150	0.097	0.058	0.046
Sample share of ever-married	0.723	0.117	0.502	0.769	0.885	0.921

Notes: OLS estimates are reported with standard errors clustered at the ancestry-country level in parentheses. The unit of observation is a female immigrant in the US. The data are from US censuses samples from IPUMS-US for 1880, 1900, 1910, 1920, and 1930. The dependent variable is a dummy taking value one if the woman has ever married and zero otherwise. "Years since immigration" is the number of years since the woman migrated to the US. All deep determinants are measured at the immigrant's country of birth. "Cool Water" is the Cool Water index described in Section 3.1. It combines the latitude of a country's centroid, the mean temperature in the hottest month, and mean rainfall in the driest month. Higher values of the index represent areas with cold winters, mild summers, and continuous rainfall. *Other deep determinants* are: the number of years (in thousands) since the Neolithic revolution (from 1500) from Putterman and Trainor (2006), the proportion of population with ancestors that used the plow in preindustrial agriculture from Alesina et al. (2013), and the share of ancestral land suitable for the cultivation of barley, wheat, sorghum, rye, foxtail millet, or pearl millet from Alesina et al. (2013). *Ancestry-country controls* are all measured at the immigrant's country of birth: ancestral domestication of large animals, ancestral settlement patterns, ancestral political complexity, fraction of ancestral land that was tropical or subtropical, the natural log of per capita income and its square, and continent dummies. *Individual controls* are: age, age squared, and literacy fixed effects. US state and census-year fixed effects included. Constants are not reported. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

For the sample of second-generation immigrants, the linear probability model is

$$M_{istfm} = \alpha + \beta_1 CW_f + \boldsymbol{X}_f^{CH\prime}\boldsymbol{\delta} + \boldsymbol{D}'_{istfm}\boldsymbol{\gamma} + \eta_s + \theta_m + \omega_t + \epsilon_{istfm} \tag{3}$$

where M is a ever-married dummy for a US-born woman i, living in US state s in census-year t, with a father born in foreign country f, and a mother born in foreign country or US state m. CW is the Cool Water index of the father's country of birth f. $\boldsymbol{X}^{\boldsymbol{CH}}$ includes the same set of historical and contemporary controls of the cross-country regressions, this time defined at the father's country of birth f. In addition, we control for individual characteristics $\boldsymbol{D}$ (age, age squared, literacy fixed effects), US state fixed effects (η_s), mother's birthplace fixed effects (country of birth or US state of birth, θ_m), and census-year fixed effects (ω_t).

We present all linear probability model estimates of Eqs. (2) and (3) for different age groups: the pooled sample of 15–39 and five-year age groups from 15–19 to 35–39. Standard errors are always clustered at the father's country of birth.

At arrival, female immigrants originating from areas with higher CW condition were less likely to be married at all ages (Table 6). The negative effect of the ancestral CW weakens slightly with years lived in the US, as shown by the positive interaction effect with years since immigration.[51] This could either suggest assimilation effects or capture different marriage selectivity of migrant cohorts. The implied effect is relevant in magnitude. At age 25–29, 77 percent of female immigrants in the (unweighted) sample were ever married (column 4). At arrival in the US, one standard deviation increase in the CW-index of the father's country of birth is associated with a 33 percentage point reduction in the likelihood of being ever married. For those living in the US for 15 years, the effect weakens to (minus) 26 percentage points.

The negative correlation between CW and marriage probability persists for the next generation. Among female children of immigrants, the CW of their fathers' country of birth correlates negatively with marriage at all ages (Table 7). Among women born to a foreign father, 68 percent were ever married at ages 25–29 (unweighted sample average). A one standard deviation increase in the CW-index of the father's country of birth is associated with a 28 percentage point reduction in marriage probability. Among women born to a foreign mother, the CW-index of the mother's country of birth is also negatively correlated with the daughter's marriage probability, although the effects are generally smaller (in absolute terms) in comparison to father's ancestral CW (Table A19).

[51] See Table A18 for estimates without the years since immigration interaction term.

Table 7
Marriage probability of children of immigrants in the US, defined by the father's country of origin.

	Female children of immigrants: 1880–1930					
	(1)	(2)	(3)	(4)	(5)	(6)
Age group:	15–39	15–19	20–24	25–29	30–34	35–39
Deep determinants:						
Cool Water	−1.304***	−0.282**	−1.695***	−1.933***	−1.346***	−1.696***
	(0.408)	(0.116)	(0.619)	(0.554)	(0.382)	(0.288)
Years of agriculture	−0.009	−0.005*	−0.010	−0.012	−0.004	−0.010*
	(0.007)	(0.002)	(0.011)	(0.009)	(0.006)	(0.005)
Plow	0.616*	0.179	0.251	1.245*	1.231**	1.627***
	(0.334)	(0.111)	(0.533)	(0.713)	(0.563)	(0.186)
Agricultural suitability	−0.481**	−0.106	−0.471	−0.782***	−0.651***	−0.871***
	(0.199)	(0.067)	(0.305)	(0.277)	(0.195)	(0.111)
Ancestry-country controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Census-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Mother BPL fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	662,914	185,709	162,830	125,311	100,343	88,721
R^2	0.402	0.073	0.092	0.060	0.046	0.041
adj. R^2	0.402	0.073	0.091	0.059	0.044	0.040
Sample share of ever-married	0.472	0.057	0.385	0.680	0.798	0.839

Notes: OLS estimates are reported with standard errors clustered at the ancestry-country level in parentheses. The unit of observation is a US-born female child of a immigrant, as defined by the father's country of origin. The data are from US censuses samples from IPUMS-US for 1880, 1900, 1910, 1920, and 1930. The dependent variable is a dummy taking value one if the woman has ever married and zero otherwise. *Deep determinants* are all measured at the father's country of origin. "Cool Water" is the Cool Water index described in Section 3.1. It combines the latitude of a country's centroid, the mean temperature in the hottest month, and mean rainfall in the driest month. Higher values of the index represent areas with cold winters, mild summers, and continuous rainfall. "Years of agriculture" is the number of years (in thousands) since the Neolithic revolution (from 1500) from Putterman and Trainor (2006). "Plow" is the proportion of population with ancestors that used the plow in preindustrial agriculture from Alesina et al. (2013). "Agricultural suitability", also from Alesina et al. (2013), is the share of ancestral land suitable for the cultivation of barley, wheat, sorghum, rye, foxtail millet, or pearl millet. *Ancestry-country controls* are measured at the father's country of birth: ancestral domestication of large animals, ancestral settlement patterns, ancestral political complexity, fraction of ancestral land that was tropical or subtropical, the natural log of per capita income and its square, and continent dummies. *Individual controls* are: age, age squared, and literacy fixed effects. Mother's birthplace (*Mother BPL*), US state, and census-year fixed effects included. Constants are not reported. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5.3.3. Subnational evidence from the Russian Empire, 1897

In the third and last historical exercise, we exploit subnational variation in the Cool Water condition across one of the largest landmass polities in history: the Russian Empire. Using data from the 1897 census, the first (and only) census of the Russian Empire, we compute the marriage share for young women, ages 15–19, for each governorate (*gubernia*), which was the Empire's main administrative subunit. Using temperature and precipitation data for each governorate, we then construct a regional CW index, which is described in detail in Online Appendix A.1 and is conceptually similar to the country-level CW index described in Section 3.1.

Governorates of the Russian Empire with larger values of the CW index exhibit substantially lower marriage rates for young women (Fig. 5). To show that this relationship is not confounded by local economic development, we present simple regressions in Table 8, controlling for three measures of preindustrial local economic development: urbanization, population density, and literacy rate (all measured from the 1897 census). Including these controls, either separately or simultaneously, does not qualitatively change the coefficient of the CW-index, which remains negative and highly statistically significantly. The estimates of Table 8 suggest that a one standard deviation increase in the CW index is associated with an average reduction of the local marriage share of young girls (15–19) of 8 to 11 percentage points. The magnitude of this estimated effect is substantial: across governorates, the average marriage share for young girls in 19.3 percent.

5.4. Cool Water and contemporary gender equality

We have shown that the CW-index is associated with smaller male-to-female differences in ages at first marriage which, in turn, are positively correlated with contemporary female-to-male ratios in labor force participation and life expectancy.

Now, we estimate the reduced-form impact of the CW-index on those present-day measures of gender equality. The reduced-form coefficient of the CW-index will be a composite of the effect of CW operating through reduced sex differences in marital ages *plus* all the other potential transmission channels that are not controlled for in our regression setup.

The results, displayed in Table 9, suggest that the reduced-form effect of the CW-index on the female–male labor force participation ratio (Panel A) is positive and robust in terms of statistical significance to the inclusion of other deep determinants.

Table 8
Russian Empire (1897): share of married women, ages 15–19.

	Share of married women, ages 15–19: 1897 census				
	(1)	(2)	(3)	(4)	(5)
Cool Water	−1.196***	−1.231***	−1.124***	−1.000***	−0.874***
	(0.135)	(0.139)	(0.133)	(0.122)	(0.126)
Urbanization		−0.254***			0.303***
		(0.093)			(0.105)
Population density			−0.002***		−0.002***
			(0.000)		(0.000)
Literacy				−0.404***	−0.445***
				(0.085)	(0.089)
Constant	0.924***	0.980***	0.942***	0.891***	0.837***
	(0.088)	(0.095)	(0.087)	(0.080)	(0.083)
N	88	88	88	88	88
R^2	0.513	0.537	0.612	0.623	0.695
adj. R^2	0.507	0.526	0.603	0.614	0.681

Notes: OLS estimates are reported with robust standard errors in parentheses. The unit of observation is a governorate (*gubernia*). The dependent variable is the share of women aged 15–19 that are married. "Cool Water" is the Cool Water index; it is the simple average of two sub-indices: *Coolness* and *Water*. *Coolness* is higher when mean temperatures in the coldest and hottest months are lower (cold winters and mild summers). *Water* is higher when mean rainfall in the driest month is larger relative to mean rainfall in the wettest month. Temperature and precipitation data at the town-level are taken from www.meteoblue.com and refer to 1988–2018. See Online Appendix A.1 for more details. "Urbanization" is the share of population living in cities. "Literacy" is the share of the population of ages 9 and above that can read and write. "Population density" is the number of residents per squared kilometer. All variables other than the CW index are from the 1897 census of the Russian Empire. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

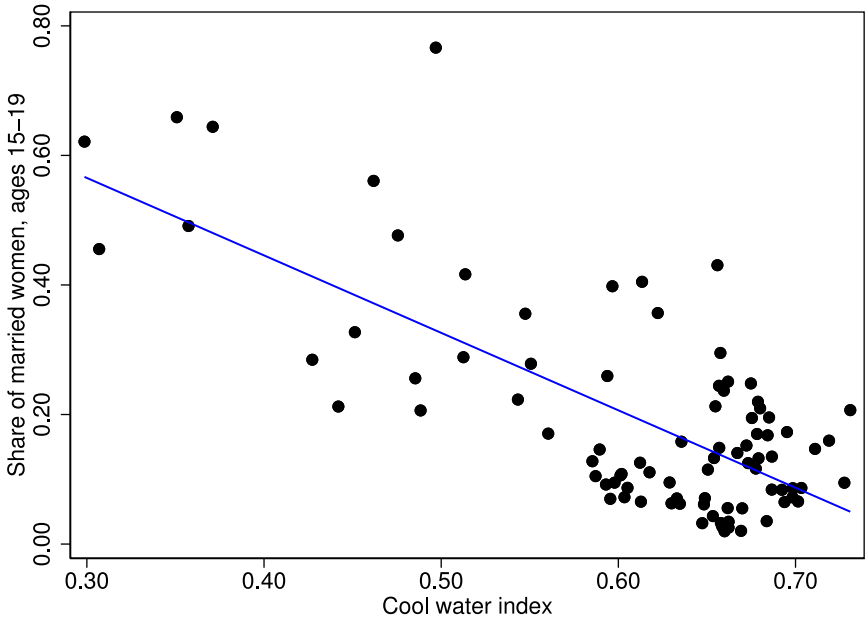


Fig. 5. Russian Empire: Marriage rate of 15–19 years-old women and Cool Water condition. *Notes:* The share of women aged 15–19 that are married is taken from the 1897 census of the Russian Empire. The unit of observation is a governorate (*gubernia*). The CW index is the simple average of two sub-indices: *Coolness* and *Water*. *Coolness* is higher when mean temperatures in the coldest and hottest months are lower (cold winters and mild summers). *Water* is higher when mean rainfall in the driest month is larger relative to mean rainfall in the wettest month. Temperature and precipitation data at the town-level are taken from www.meteoblue.com and refer to 1988–2018. See Online Appendix A.1 for more details.

One standard deviation increase in the CW index is associated with a 0.34 standard deviations increase in the female to male labor force participation ratio.

Table 9
Determinants of gender gaps: reduced form estimates.

	Average female–male ratio in 1990–2010							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Labor force participation							
Deep determinants:								
Cool Water	0.626***	0.447***	0.715***	0.480**	0.612***	0.310*	0.578***	0.492***
	(0.178)	(0.165)	(0.195)	(0.185)	(0.175)	(0.171)	(0.199)	(0.182)
Years of agriculture		−0.055***			−0.048***	−0.054***		−0.048***
		(0.011)			(0.011)	(0.011)		(0.011)
Plow			−0.190***		−0.188***		−0.181***	−0.170***
			(0.062)		(0.056)		(0.055)	(0.054)
Agricultural				0.149**		0.136**	0.136**	0.104*
suitability				(0.067)		(0.062)	(0.059)	(0.055)
Historical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contemporary controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	156	146	156	156	146	146	156	146
R^2	0.282	0.452	0.332	0.309	0.493	0.471	0.354	0.504
	Panel B: Life expectancy							
Deep determinants:								
Cool Water	0.073***	0.060**	0.078***	0.061**	0.073***	0.052*	0.066**	0.068**
	(0.025)	(0.025)	(0.026)	(0.027)	(0.025)	(0.028)	(0.029)	(0.028)
Years of agriculture		−0.005***			−0.005**	−0.005***		−0.005**
		(0.002)			(0.002)	(0.002)		(0.002)
Plow			−0.011		−0.015		−0.011	−0.015
			(0.011)		(0.011)		(0.010)	(0.011)
Agricultural				0.012		0.007	0.011	0.004
Suitability				(0.010)		(0.012)	(0.010)	(0.011)
Historical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contemporary controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	158	147	158	158	147	147	158	147
R^2	0.408	0.466	0.413	0.413	0.474	0.468	0.418	0.475
	Panel C: Years of education							
Deep determinants:								
Cool Water	0.245	0.114	0.255	0.254	0.087	0.119	0.265	0.085
	(0.205)	(0.200)	(0.190)	(0.198)	(0.189)	(0.195)	(0.179)	(0.175)
Years of agriculture		−0.029***			−0.030***	−0.029***		−0.030***
		(0.008)			(0.009)	(0.008)		(0.009)
Plow			−0.017		0.029		−0.017	0.029
			(0.073)		(0.085)		(0.074)	(0.087)
Agricultural				−0.009		−0.004	−0.010	0.001
Suitability				(0.057)		(0.057)	(0.056)	(0.059)
Historical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contemporary controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	130	127	130	130	127	127	130	127
R^2	0.581	0.618	0.582	0.581	0.619	0.618	0.582	0.619

Notes: OLS estimates are reported with robust standard errors in parentheses. "Cool Water" is the Cool Water index described in Section 3.1. It combines the latitude of a country's centroid, the mean temperature in the hottest month, and mean rainfall in the driest month. Higher values of the index represent areas with cold winters, mild summers, and continuous rainfall. "Years of agriculture" is the number of years (in thousands) since the Neolithic revolution (from 1500) from Putterman and Trainor (2006). "Plow" is the proportion of population with ancestors that used the plow in preindustrial agriculture from Alesina et al. (2013). "Agricultural suitability", also from Alesina et al. (2013), is the share of ancestral land suitable for the cultivation of barley, wheat, sorghum, rye, foxtail millet, or pearl millet. *Historical controls* are: ancestral domestication of large animals, ancestral settlement patterns, and ancestral political complexity from Alesina et al. (2013). *Contemporary controls* are the natural log of per capita income and its square, measured in the same time period as the dependent variable. Continent dummies are included. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The CW-index is also a significant positive correlate of contemporary female–male ratio in life expectancy (Table 9, Panel B). One standard deviation increase in the CW index is associated with a 0.26 standard deviations increase in the female to male life expectancy ratio.

Finally, consistent with the lack of correlation between ages at first marriage and the gender ratio in years of education, the reduced-form coefficient of the CW-index is small and statistically insignificant (Table 9, Panel C). The only robust negative deep determinant of gender equality in education is years of agriculture.

6. Conclusion

Reviewing the burgeoning literature on the remote historic drivers of gender inequality, we presented evidence for an overlooked trajectory that (1) originates in the Cool Water condition, from where the path leads to (2) late female marriages in preindustrial times, which eventually pave the way towards (3) various gender egalitarian outcomes today.

In theorizing this evidence, we argue that the CW condition embodies opportunity endowments that significantly reduced fertility pressures on women, which favored late female marriages in the preindustrial era. The resulting family and household patterns placed women into a better position to struggle for more gender equality during subsequent economic transitions toward the industrial and post-industrial stages of development. Hence, enduring territorial differences in the CW condition predict preindustrial female marriage ages, which in turn predict gender equality today.

In support of these hypotheses, we provide cross-country evidence that, in preindustrial Europe, women married later in countries with a higher Cool Water index. We are able to trace this effect, at the turn of the twentieth century (1880–1930), among first- and second-generation immigrants in the US. Female immigrants whose country of origin had a higher CW condition were less likely to marry over the 15–39 age distribution. Likewise, in the late nineteenth century, regions of the Russian Empire with a higher CW condition exhibited substantially lower marriage rates for young girls (15–19). Finally, we show that the CW effect has persisted into the contemporary period. Using a cross-section of countries, from 1960 onwards, we estimate a sizeable effect of a country's CW score on female ages at first marriage. Altogether, the empirical evidence spans a long period of time, involving several independent datasets and levels of analysis (country, ancestry, region, individual).

In sum, we argue that the challenges and opportunities historically embodied in an agrarian society's thermo-hydrological condition shaped the evolution of peasant household formation, and, as a result, generated part of the wide, long-run variation in gender norms and gender inequality observed across countries today.

Our theory is compatible with, and actually integrates, several separate theories on the historic origins of gender (in)equality. First, the argument that scarcity in arable land favored historic gender inequality is incorporated, because the Cool Water condition explains the absence of such scarcity. Second, the argument that irrigation dependence favored historic gender inequality is incorporated, because the CW condition explains the absence of such dependence. Third, the argument that disease prevalence favored historic gender inequality is incorporated, because the CW condition explains the absence of such prevalence. Fourth, our theory suggests that the long-run evolution of social institutions that regulate family and gender norms (such as religion) is endogenous to local climatic conditions and can thus be partly explained by the prevalence of the CW condition in some areas but not in others. Fifth, the argument that European descent favored historic gender equality is incorporated, because European descent is linked to historic gender equality only in CW areas but not outside them. In conclusion, we suggest that our theory of female emancipation provides a credible umbrella in unifying previous theories of gender equality.

Appendix A. Online appendix

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jce.2022.11.001.

References

Acemoglu, D., Autor, D.H., Lyle, D., 2004. Women,war, and wages: The effect of female labor supply on the wage structure at midcentury. J. Polit. Econ. 112, 497–551.

Ager, P., Worm Hansen, C., Sandholt Jensen, P., 2018. Fertility and early-life mortality: Evidence from smallpox vaccination in Sweden. J. Eur. Econom. Assoc. 16, 487–521.

Alesina, A., Giuliano, P., Nunn, N., 2013. On the origins of gender roles: Women and the plough. Q. J. Econ. 128, 469–530.

Alexander, A.C., Welzel, C., 2015. Eroding patriarchy: The co-evolution of women's rights and emancipative values. Int. Rev. Sociol. 25, 144–165.

Altonji, J.G., Elder, T.E., Taber, C.R., 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. J. Polit. Econ. 113, 151–184.

Antecol, H., 2000. An examination of cross-country differences in the gender gap in labor force participation rates. Labour Econ. 7, 409–426.

Banerji, M., Martin, S., Desai, S., 2008. Is Education Associated with a Transition Towards Autonomy in Partner Choice? a Case Study of India. India Human Development Survey Working Paper No. 8.

Barro, R.J., Lee, J.W., 2013. A new data set of educational attainment in the world, 1950–2010. J. Dev. Econ. 104, 184–198.

Becker, S.O., Cinnirella, F., Woessmann, L., 2010. The trade-off between fertility and education: Evidence from before the demographic transition. J. Econ. Growth 15, 177–204.

Becker, S.O., Woessmann, L., 2008. Luther and the girls: Religious denomination and the female education gap in nineteenth-century Prussia. Scand. J. Econ. 110, 777–805.

Becker, S.O., Woessmann, L., 2009. Was Weber wrong? A human capital theory of protestant economic history. Q. J. Econ. 124, 531–596.

Bentzen, J.S., Kaarsen, N., Wingender, A.M., 2017. Irrigation and autocracy. J. Eur. Econom. Assoc. 15, 1–53.

Boserup, E., 1970. Woman's Role in Economic Development. George Allen and Unwin Ltd, London.

Boter, C., 2017. Marriages are made in kitchens: The European marriage pattern and life-cycle servanthood in eighteenth-century Amsterdam. Fem. Econ. 23, 68–92.

Branisa, B., Klasen, S., Ziegler, M., 2013. Gender inequality in social institutions and gendered development outcomes. World Dev. 45, 252–268.

Bredtmann, J., Höckel, L.S., Otten, S., 2020. The intergenerational transmission of gender role attitudes: Evidence from immigrant mothers-in-law. J. Econ. Behav. Organ. 179, 101–115.

Brollo, F., Troiano, U., 2016. What happens when a woman wins an election? Evidence from close races in Brazil. J. Dev. Econ. 122, 28–45.

Campbell, B.M.S., 2000. English Seigniorial Agriculture, 1250–1450. Cambridge University Press, Cambridge.

Carmichael, S.G., 2011. Marriage and power: Age at first marriage and spousal age gap in lesser developed countries. Hist. Fam. 16, 416–436.

Carmichael, S.G., de Pleijt, A., van Zanden, J.L., De Moor, T., 2016. The European marriage pattern and its measurement. J. Econ. Hist. 76, 196–204.

Cashdan, E., 2014. Biogeography of human infectious diseases: A global historical analysis. PLoS One 9, 1–11.
Cavalcanti, T.V.d.V., Tavares, J., 2008. Assessing the engines of liberation: Home appliances and female labor force participation. Rev. Econ. Stat. 90, 81–88.
Cinnirella, F., Klemp, M., Weisdorf, J., 2017. Malthus in the bedroom: Birth spacing as birth control in pre-transition England. Demography 54, 413–436.
Coen-Pirani, D., León, A., Lugauer, S., 2010. The effect of household appliances on female labor force participation: Evidence from microdata. Labour Econ. 17, 503–513.
Cornell, L.L., 1987. Hajnal and the household in Asia: A comparativist history of the family in preindustrial Japan, 1600–1870. J. Fam. Hist. 12, 143–162.
Davis, N.Z., 1977. Ghosts,kin, and progeny: Some features of family life in early modern France. Daedalus 106, 87–114.
De Moor, T., 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. Econ. Hist. Rev. 63, 1–33.
Deaton, A., 2013. The Great Escape: Health, Wealth, and the Origins of Inequality. Princeton University Press, Princeton, NJ.
Dennison, T., Ogilvie, S., 2014. Does the European marriage pattern explain economic growth? J. Econ. Hist. 74, 651–693.
Diamond, J., 1997. Guns, Germs, and Steel: The Fates of Human Societies. WW Norton & Company, New York.
Diebolt, C., Perrin, F., 2013. From stagnation to sustained growth: The role of female empowerment. Amer. Econ. Rev. 103, 545–549.
Dilli, S., 2016. Family systems and the historical roots of global gaps in democracy. Econ. Hist. Dev. Reg. 31, 82–135.
Duflo, E., 2012. Women, empowerment, and economic development. J. Econ. Lit. 50, 1051–1079.
Duranton, G., Rodríguez-Pose, A., Sandall, R., 2009. Family types and the persistence of regional disparities in Europe. Econ. Geogr. 85, 23–47.
Dyble, M., Salali, G.D., Chaudhary, N., Page, A., Smith, D., Thompson, J., Vinicius, L., Mace, R., Migliano, A.B., 2015. Sex equality can explain the unique social structure of hunter-gatherer bands. Science 348, 796–798.
Ebenstein, A., 2014. Patrilocality and Missing Women. Mimeo. Hebrew University.
Edlund, L., Lagerlö, N.P., 2006. Individual versus parental consent in marriage: Implications for intra-household resource allocation and growth. Amer. Econ. Rev. 96, 304–307.
Engels, F., 1902. The Origin of the Family, Private Property, and the State. Charles H. Kerr & Company Cooperative, Chicago.
Fernández, R., 2007. Women, work, and culture. J. Eur. Econom. Assoc. 5, 305–332.
Fincher, C.L., Thornhill, R., Murray, D.R., Schaller, M., 2008. Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. Proc. R. Soc. Lond. Ser. B: Biol. Sci. 275, 1279–1285.
Fish, M.S., 2002. Islam and authoritarianism. World Polit. 55, 4–37.
Gaddis, I., Klasen, S., 2014. Economic development, structural change, and women's labor force participation. J. Popul. Econ. 27, 639–681.
Gallup, J.L., Mellinger, A.D., Sachs, J.D., 2010. Geography Datasets. Center for International Development, Harvard University.
Galor, O., Weil, D.N., 1996. The gender gap, fertility, and growth. Amer. Econ. Rev. 86, 374–387.
Giuliano, P., 2007. Living arrangements in Western Europe: Does cultural origin matter? J. Eur. Econom. Assoc. 5, 927–952.
Giuliano, P., 2015. The role of women in society: From preindustrial to modern times. CESifo Econ. Stud. 61, 33–52.
Giuliano, P., 2018. Gender: A historical perspective. In: Averett, S.L., Argys, L.M., Hoffman, S.D. (Eds.), The Oxford Handbook of Women and the Economy. Oxford University Press, pp. 645–672.
Goldin, C., 1991. The role of world war II in the rise of women's employment. Amer. Econ. Rev. 81, 741–756.
Goldin, C., 1995. The U-shaped female labor force function in economic development and economic history. In: Schultz, T.P. (Ed.), Investment in Women's Human Capital and Economic Development. University of Chicago Press, Chicago, pp. 61–90.
Goldin, C., Katz, L.F., 2002. The power of the pill: Oral contraceptives and women's career and marriage decisions. J. Polit. Econ. 110, 730–770.
Goody, J., 1983. The Development of the Family and Marriage in Europe. Cambridge University Press.
Greenwood, J., Seshadri, A., Yorukoglu, M., 2005. Engines of liberation. Rev. Econom. Stud. 72, 109–133.
Guernier, V., Hochberg, M.E., Guégan, J.F., 2004. Ecology drives the worldwide distribution of human diseases. PLOS Biol. 2, e141.
Haber, S., 2012. Climate, Technology, and the Evolution of Economic and Political Institutions. PERC Research Paper No. 12/2, The Property and Environment Research Center.
Hajnal, J., 1965. European marriage patterns in perspective. In: Glass, D.V., Eversley, D.E.C. (Eds.), Population in History: Essays in Historical Demography. Edward Arnold Ltd, London, pp. 101–143, chapter 6.
Hajnal, J., 1982. Two kinds of preindustrial household formation system. Popul. Dev. Rev. 8, 449–494.
Hansen, C.W., Jensen, P.S., Skovsgaard, C.V., 2015. Modern gender roles and agricultural history: The neolithic inheritance. J. Econ. Growth 20, 365–404.
Hartman, M.S., 2004. The Household and the Making of History: A Subversive View of the Western Past. Cambridge University Press, Cambridge.
Hazarika, G., Jha, C.K., Sarangi, S., 2015. The Role of Historical Resource Scarcity in Modern Gender Inequality. Working Paper 2015-06, Louisiana State University.
Hazarika, G., Jha, C.K., Sarangi, S., 2019. Ancestral ecological endowments and missing women. J. Popul. Econ. 32, 1101–1123.
Henrich, J., 2020. The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous. Penguin UK.
Herlihy, D., 1985. Medieval Households. Cambridge University Press, Cambridge.
Höckel, L.S., 2018. Collectivism in the labor market: Evidence from second generation immigrants in the United States. J. Comp. Econ. 46, 1347–1369.
Inglehart, R., Norris, P., 2003. The true clash of civilizations. Foreign Policy 135, 62–70.
Jayachandran, S., 2015. The roots of gender inequality in developing countries. Annu. Rev. Econ. 7, 63–88.
Jones, E.L., 1981. The European Miracle: Environments, Economies and Geopolitics in the History of Europe and Asia. Cambridge University Press, Cambridge.
Kaufmann, D., Kraay, A., Mastruzzi, M., 2011. The worldwide governance indicators: Methodology and analytical issues. Hague J. Rule Law 3, 220–246.
Klasen, S., 1998. Marriage, bargaining, and intrahousehold resource allocation: Excess female mortality among adults during early German development, 1740–1860. J. Econ. Hist. 58, 432–467.
Klasen, S., 2002. Low schooling for girls, slower growth for all? Cross-country evidence on the effect of gender inequality in education on economic development. World Bank Econ. Rev. 16, 345–373.
Klasen, S., 2003. Weibliche übersterblichkeit in entwicklungsländern: Eine ökonomische analyse. In: Neuere AnsäTze Der Theoretischen Und Empirischen EntwicklungsläNderforschung. Duncker und Humblot, Berlin, pp. 257–282.
Klasen, S., Lamanna, F., 2009. The impact of gender inequality in education and employment on economic growth: New evidence for a panel of countries. Fem. Econ. 15, 91–132.
Klein Goldewijk, K., Beusen, A., Janssen, P., 2010. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. Holocene 20, 565–573.
La Porta, R., Lopez-de Silanes, F., Shleifer, A., Vishny, R., 1999. The quality of government. J. Law Econ. Organ. 15, 222–279.
Lagerlöf, N.P., 2003. Gender equality and long-run growth. J. Econ. Growth 8, 403–426.
Mammen, K., Paxson, C., 2000. Women's work and economic development. J. Econ. Perspect. 14, 141–164.
Marshall, A., 2013. Principles of Economics, eightth ed. Palgrave Macmillian.
Mitterauer, M., 2010. Why Europe? The Medieval Origins of Its Special Path. University of Chicago Press, Chicago.
Murray, D.R., Schaller, M., 2010. Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture. J. Cross-Cult. Psychol. 41, 99–108.
Nunn, N., Puga, D., 2012. Ruggedness: The blessing of bad geography in Africa. Rev. Econ. Stat. 94, 20–36.

Olsson, O., Paik, C., 2016. Long-run cultural divergence: Evidence from the Neolithic revolution. J. Dev. Econ. 122, 197–213.
Oster, E., 2019. Unobservable selection and coefficient stability: Theory and evidence. J. Bus. Econom. Statist. 37, 187–204.
Parker, P.M., 2000. Physioeconomics: The Basis for Long-Run Economic Growth. MIT Press, Cambridge, Massachusetts.
Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. Discuss. 4, 439–473.
Poos, L.R., 1991. A Rural Society After the Black Death: Essex: 1350–1525. Cambridge University Press, Cambridge.
Powelson, J.P., 1994. Centuries of Economic Endeavor: Parallel Paths in Japan and Europe and their Contrast with the Third World. University of Michigan Press, Ann Arbor, Michigan.
Prettner, K., Strulik, H., 2017. Gender equity and the escape from poverty. Oxf. Econ. Pap. 69, 55–74.
Putterman, L., Trainor, C.A., 2006. Agricultural transition year country data set. Online Resour..
Putterman, L., Weil, D.N., 2010. Post-1500 population flows and the long-run determinants of economic growth and inequality. Q. J. Econ. 125, 1627–1682.
Reher, D.S., 1998. Family ties in Western Europe: Persistent contrasts. Popul. Dev. Rev. 24, 203–234.
Rijpma, A., Carmichael, S.G., 2016. Testing Todd and matching Murdock: Global data on historical family characteristics. Econ. Hist. Dev. Reg. 31, 10–46.
Ross, M.L., 2008. Oil, Islam, and women. Amer. Polit. Sci. Rev. 102, 107–123.
Ruggles, S., 2009. Reconsidering the Northwest European family system: Living arrangements of the aged in comparative historical perspective. Popul. Dev. Rev. 35, 249–273.
Santos Silva, M., Klasen, S., 2021. Gender inequality as a barrier to economic growth: A review of the theoretical literature. Rev. Econ. Househ. 19, 581–614.
Schulz, J.F., Bahrami-Rad, D., Beauchamp, J.P., Henrich, J., 2019. The church, intensive kinship, and global psychological variation. Science 366, eaau5141.
Smith, R.M., 1981. Fertility, economy, and household formation in England over three centuries. Popul. Dev. Rev. 7, 595–622.
Spolaore, E., Wacziarg, R., 2009. The diffusion of development. Q. J. Econ. 124, 469–529.
Spolaore, E., Wacziarg, R., 2013. How deep are the roots of economic development? J. Econ. Lit. 51, 325–369.
Szołtysek, M., 2014. Toward a conceptual framework for the variation in historical family and household systems across Eurasia. Przeszl. Demogr. Polski 36, 55–86.
Szołtysek, M., Klüsener, S., Poniat, R., Gruber, S., 2017. The patriarchy index: A new measure of gender and generational inequalities in the past. Cross-Cult. Res. 51, 228–262.
Thompson, G.H., 1951. The Factorial Analysis of Human Ability. University of London Press, London.
Todd, E., 1985. The Explanation of Ideology: Family Structures and Social Systems. Blackwell, Oxford.
Todd, E., 1987. The Causes of Progress: Culture, Authority and Change. Blackwell, Oxford.
Tucker, J., 2008. Women, Family, and Gender in Islamic Law (Themes in Islamic Law). Cambridge University Press, Cambridge.
Tur-Prats, A., 2019. Family types and intimate partner violence: A historical perspective. Rev. Econ. Stat. 101, 878–891.
United Nations, Department of Economic and Social Affairs, Population Division, 2009. World marriage data 2008. (POP/DB/Marr/Rev2008).
Voigtländer, N., Voth, H.J., 2013. How the west invented fertility restriction. Amer. Econ. Rev. 103, 2227–2264.
Welzel, C., 2013. Freedom Rising. Cambridge University Press, Cambridge.
Welzel, C., 2014. Evolution, empowerment, and emancipation: How societies climb the freedom ladder. World Dev. 64, 33–51.
Whyte, M.K., 1978. The Status of Women in Pre-Industrial Societies. Princeton University Press, Princeton, NJ.
Wittfogel, K., 1957. Oriental Despotism: A Comparative Study of Total Power. Yale University Press, New Haven, Connecticut.
World Bank, 2001. Engendering Development Through Gender Equality in Rights, Resources, and Voice. The World Bank and Oxford University Press.
World Bank, 2011. World Development Report 2012: Gender Equality and Development. The World Bank, Washington, DC.
World Bank, 2016. World development indicators. [accessed on 24-09-2016].