Max-Planck-Institut für demografische Forschung

Max Planck Institute for Demographic Research

Konrad-Zuse-Strasse 1 ▪ D-18057 Rostock ▪ Germany ▪ Tel +49 (0) 3 81 20 81 - 0 ▪ Fax +49 (0) 3 81 20 81 - 202 ▪ www.demogr.mpg.de

MPIDR Working Paper WP 2017-012 I May 2017

Family organisation and human capital inequalities in historic Europe: testing the association anew

Mikołaj Szołtysek I szoltysek@eth.mpg.de
Radosław Poniat
Sebastian Klüsener I kluesener@demogr.mpg.de
Siegfried Gruber

Mikołaj Szołtysek[*], Radosław Poniat^, Sebastian Klüsener^^, and Siegfried Gruber[++]

Family organisation and human capital inequalities in historic Europe: testing the association anew[1]

Abstract

There has been growing interest in the question of whether variation in family systems is a factor in the disparities in growth, development, and human capital formation. Studies by proponents of the field of New Institutional Economics have suggested that differences in family organisation could have considerable influence on regional developmental inequalities in today's world, while a number of economic historians have argued that certain systems of marriage and household structure in the European past might have been more conducive than others to economic growth. Despite recent criticism of these ideas by Dennison and Ogilvie, who argued that the family has no exogenous effects on growth, the debate over this potential relationship continues. However, we believe that this discussion has been suffering from a lack of historical data that would give a fuller picture of the rich diversity of family settings, and from methodological shortcomings that have so far hindered the proper operationalisation of historical family systems and their potential effects on developmental outcomes. In this paper, we apply a recently developed multidimensional measure of historic familial organisation, the Patriarchy Index; and use spatially sensitive multivariate analyses to investigate its relationship with human capital levels, as approximated by numeracy across 115 populations of historic Europe. We find a strong negative association between the Patriarchy Index and regional numeracy patterns that remains significant even after controlling for a broad range of other important factors. Our observation that family-driven age- and gender-related inequalities, as captured by the index, are relevant for understanding variation in basic numeracy patterns in the past suggests that there are indeed important links between family organisation and human capital accumulation that merit further investigation.

[*] Max Planck Institute for Social Anthropology, Halle/Saale, Germany. Corresponding author (szoltysek@eth.mpg.de)

^ University of Białystok, Poland

^^ Max Planck Institute for Demographic Research, Rostock, Germany

[++] University of Graz, Austria.

[1] This paper is an elaborated version of arguments and observations presented in 2016 at the Economic and Social History Seminar at Utrecht University, the HPSS Seminar at the Cambridge Group for the History of Population and Social Structure, the European Social Science History Conference in Valencia, and the European Society of Historical Demography Conference in Leuven. We thank Jan Luiten Van Zanden, Auke Rijpma, Jan Kok and one anonymous reviewer for their comments on earlier versions of this draft. Any and all errors and omissions are our own. The paper is forthcoming in "Cliometrics and the Family: Global Patterns and their Impact on Diverging Development" edited by Claude Diebolt, Auke Rijpma, Sarah Carmichael, Selin Dilli, and Charlotte Störmer (Springer).

Introduction

There has been a growing interest in family patterns as potential determinants of differential growth and human development. The hypothesis that family systems could have an impact on wider societal outcomes represents a reversal of the more usual causal argument, which posits that economic development produces changes in dominant family patterns in societies around the world (Goode, 1963). While not new, the claim that family patterns can influence whether and how rapidly development occurs has been advanced much less frequently. Weber alluded to it when he argued that strong family values do not allow for the development of individual forms of entrepreneurship, which are fundamental to the formation of capitalist societies (Weber, 1904). As early as in the 1960s, Nimkoff stated in the context of his cross-cultural research that family organisation is not without influence on the social order, and suggested that there are links between certain variations in family structure and types of marriage on the one hand; and the strength of economic incentives, individual mobility, employment patterns, and political equality and participation on the other (Nimkoff, 1965, 61 ff).

More recently, E. Todd sought to use family systems to explain larger societal phenomena in *Explanation of Ideology* (1985) and *Causes of Progress* (1987). In *Causes of Progress*, Todd hypothesised that educational attainment is primarily determined by the parental authority vested in women. Specifically, he argued that the more power women have, the more educated the next generation will be. Thus, Todd linked the traits of family systems to key developments in the global economic and social histories of the 19th and 20th centuries.

Even more recently, New Institutional Economists have argued that family patterns can greatly influence regional inequalities through their effects on, for example, the status of women, investments in human capital, the persistence of specific cultural norms and values, labour relations, and the development of corporative institutions (e.g., Alesina & Giuliano, 2010; Duranton et al., 2009; Greif, 2006; also Carmichael et al., 2016). Since 2009, Alesina and Giuliano have been using a measure of “family ties” that is based on a set of responses collected from the six waves of the World Values Survey (1981-2010). In a series of regression models, they showed that strong family ties are negatively correlated with generalised trust, and that such ties are associated with more household production and lower levels of labour market participation among women, young adults, and the elderly. They also found that strong family ties are correlated with lower levels of interest and participation in political activities. In addition, such ties imply labour market regulation and welfare systems that are based on the family rather than on the market or the government (for a summary, see Alesina & Giuliano, 2014).

Similar perspectives can also be found in the historical demography and the economic history literature. In the 1980s, after Peter Laslett showed that the nuclear family structure had been the dominant family type in England long before any industrial development occurred in that country (between 1574-1821; Laslett, 1965), some scholars (including Laslett himself) argued that the dominance of the nuclear family was among the necessary preconditions for modernisation and industrialisation (Laslett, 1983; Macfarlane, 1987). More recently, a number of economic historians have asserted that certain systems of marriage and celibacy, individual life course trajectories, and patterns of household structure and formation were more conducive to economic growth than others (De Moor & Van Zanden, 2010; Foreman-Peck, 2011; also Greif, 2006; Duranton et al., 2009; Kick et al., 2000). In particular, these scholars have posited that the historical north-western European marriage and family pattern (based on late marriage, neolocality, and high levels of lifetime celibacy among women; henceforth EMP) was a key factor not only in the economic success of northern and western Europe relative to southern and eastern Europe, but also in the "great divergence" between Europe and the rest of the world. In these accounts, the family-growth nexus has often been presented in relation to the transformation of gender roles; i.e., it is posited that shifts in the balance of power between the generations and between the sexes have led to improvements in the human capital of (mainly female) household members ("girl power"), and, consequently, to a greater accumulation of human capital overall, which has in turn stimulated growth over the long run (De Moor & Van Zanden, 2010; Foreman-Peck, 2011; also Diebolt & Perrin, 2013).

Dennison and Ogilvie (2014; 2016) recently criticised the theoretical claims and empirical analyses in support of the view that historical family systems (and the EMP and its associated "girl power" in particular) played a role in economic growth in the early modern period. In their critique, they argued that family is interdependent with the wider framework of non-familial institutions, and that the impact on growth of any individual institution is constrained by the entire system in which it is embedded. Furthermore, they constructed a large meta-dataset of historical marriage and household patterns in 39 countries, and found no evidence that the EMP provided a foundation for the economic success of north-western Europe. In conclusion, they questioned the exogeneity of the EMP, pointing out that non-familial institutions (such as property rights and market factors) had far greater effects on cross-country differences in economic growth than family characteristics. However, the discussion of these issues continues to unfold. A recent contribution by Carmichael et al. (2016, 200) has suggested that this criticism of the influence of the EMP may be premature, and should be re-examined using newly available data from the recent expansion of public-use historical demographic databases, including those of Mosaic, NAPP, and IPUMS.

While we cannot consider all aspects of this debate in our study, we can make two important contributions to this discussion: namely, we demonstrate the importance of using better data and of properly operationalising historical family systems and their relationship to human capital formation. A recurring problem that all scholars who have attempted to study this issue have encountered is a lack of suitable and reliable historical data. In order to fully analyse the relationship between family and development, researchers need access to data that are both sufficiently detailed to account for the specific characteristics of family organisation, and sufficiently broad geographically to capture regional variation across Europe. In previous studies, however, the developmental statistics were linked to crude classifications of historical family systems, and the spatiality and multistrandedness of these classifications were considered only selectively (Todd, 1987; recently Duranton et al., 2009; Le Bris, 2016). Of the many elements of family organisation that may be linked to economic performance and human capital formation, only a few have been analysed, with most research focusing exclusively on marriage and celibacy patterns and on the prevalence of nuclear families (see De Moor and Van Zanden, 2010; Dennison and Ogilvie, 2014; Bertocchi & Bozzano, 2015; Baten, Szołtysek, & Campestrini, 2017). Moreover, most of these studies operated at a relatively high level of aggregation, using data at the level of administrative districts or countries (Rijpma & Carmichael, 2016); and had a spatial bias towards the high-performing economies of the North Atlantic seaboard (De Moor & Van Zanden, 2010; see, e.g., Bertocchi & Bozzano, 2015 for a rare focus on Italy).

Our intention in this paper is to move beyond these limitations by using data from Mosaic, one of the largest data infrastructure projects that have been undertaken in historical demography and family sociology (www.censusmosaic.org). Mosaic provides harmonised samples of census and census-like microdata for 115 regions of historic Europe, from Catalonia to central Russia. Because of its unprecedented scope, the information contained in the Mosaic database allows us to shift the discussion of the role of the family by making it possible for the first time to account for a wide variety of family patterns on the continent, and to represent family organisation patterns across major European demographic fault lines. Here, we apply Mosaic data to a recently developed measure of historic familial organisation (Gruber & Szołtysek, 2016). This multidimensional measure, which we call the Patriarchy Index (henceforth, PI), combines a range of variables related to familial behaviour, including nuptiality and age at marriage, living arrangements, post-marital residence, power relations within domestic groups, the position of the elderly, and the sex of the offspring. Thus, this measure is more comprehensive than previous assessment tools that examined cruder components of family systems. Moreover, by capturing the inner architecture of generational and gender relations at the domestic level across different family settings, the PI can be used to identify the channels that may have affected economic behaviour at the individual level. Furthermore, based on the theoretical predictions of the economic literature (e.g., Diebolt & Perrin, 2013), we shift the discussion from a focus on economic growth to a focus on one of its underlying forces: namely, human capital, which we approximate in our historical data by following the well-established age-heaping methodology.

We contribute to the ongoing debate on the relationship between family organisation and human capital accumulation by being the first – to the best of our knowledge – to empirically test this relationship using comprehensive, spatially sensitive family-demographic evidence at a European scale[2]. Our cross sectional results support the existence of a strong correlation between family organisation, as captured by our focal variable (the PI), and our measure of numeracy. This outcome remains even after controlling for many other relevant covariates. Based on the analyses we present in this paper, we cannot rule out the possibility of bidirectional or circular causation, or that the strong link between family structures and numeracy patterns is merely an outcome of a deeper underlying determinant. However, our identification strategy allows us to capture potential influences on human capital accumulation other than that of historical patriarchy. In addition, we offer a number of theoretical explanations for why familial institutions might have played an exogenous role. Generally, our results show that family-driven age- and gender-related inequalities, as captured by the PI, are relevant for understanding variation in basic numeracy patterns in historic populations. This suggests that there are indeed important links between family organisation and human capital accumulation that merit further investigation.

This chapter is organised as follows. First, we theorise how family organisation and numeracy patterns might be linked through microeconomic, cultural, and behavioural channels; both directly and indirectly. We then present our data and explain how they were used for the construction of our measure of family organisation. Next, we illustrate the index's application to data for 115 regions of historic Europe, and introduce our measure of numeracy, which includes a discussion of some potential biases involved in its interpretation. Finally, we present spatially sensitive regression estimates of the relationship between numeracy and a set of important covariates, including the PI and other controls that account for the broad variation in socioeconomic, institutional, and environmental conditions across the societies covered by our data. Finally, we conclude by summarising the main findings of the paper.

Background: the familial patriarchy-numeracy nexus

Human capital is arguably one of the most important determinants of economic growth (Galor & Weil, 2000; Hanuschek, 2013; Goldin, 2016; cf. Ogilvie & Küpker, 2015)[3]. Its development requires a significant amount of parental, social, and individual effort. As a component of human capital, numeracy can be "co-produced" by a number of different agents who (may) influence its potential development: family members (parents in particular), the government (bureaucracy), the community, educational institutions, other social services,

[2] Baten, Szołtysek, & Campestrini (2017) included in their set of predictors used to model historical numeracy patterns in east-central and eastern Europe one component of family related behaviour that they treated as indication of female autonomy; i.e., female marriage patterns.

[3] Broadly speaking, human capital can be defined as the knowledge, the skills, the competencies, and the attributes embodied in individuals that contribute to their "productivity"; or, more broadly, that facilitate the creation of personal, social, and economic well-being (see Goldin, 2016).

society as a whole, and – finally – the person who embodies human capital (Folbre 2012, 283; also Acemoglu 2002). Intuitively, family and households are likely to be primary agents in this process, especially in historical societies that lacked widespread schooling and formal educational institutions. Not only have households been an essential part of the functioning of preindustrial economy and society (Szołtysek, 2015a); they were also the most basic arena for building and enacting kinship ties, socialising individuals, and transmitting values, including values related to power and equality, justice and gender relations, age hierarchy, and the relationship between the individual and the authorities (Kok, *forthcoming*; Carmichael et al., 2016; Folbre, 1986; Malhotra et. al., 2002). Given the central role played by the family in building human capital, it is important that we understand variation of familial behaviour across historic Europe in order to address the questions of interest to us here.

There are a wide range of angles from which family systems may be analysed. A number of scholars have suggested ways to measure family systems across time and space. However, most of these approaches have tended to favour one structural aspect of the family system (e.g., household structure or marriage patterns), while neglecting others (for a discussion, see Gruber & Szołtysek, 2016; Szołtysek, 2015a). A heuristically attractive feature of focusing on family systems – albeit one that becomes apparent only when seen through a holistic lens – is that it is clear that different family types have varying effects on the status of women, the aged, and the young; and, thus, that family organisation can systematically enhance or diminish the agency of (specific) family members (Kok, *forthcoming*). We can therefore assume that family systems have a bearing on family members' access to human capital investment opportunities (cf. Acemoglu, 2002). In this context, the concept of familial patriarchy, which refers to the levels of sex- and age-related social inequality in familial settings (Gruber & Szołtysek, 2016; Szołtysek et al., 2017), seems particularly useful for providing a theoretical basis for the more general nexus between family and numeracy (and human capital more broadly)[4].

There are various channels through which patriarchy may negatively affect the accumulation of human capital, including numerical skills. First, patriarchy may hinder the formation of human capital because of the inherent interference of older and male family members in the life course decisions of younger and female family members. Thus, patriarchy can place powerful constraints on individual agency. In the male- and adult-centred patriarchal social hierarchy, human capital acquisition by younger and female family members may be discouraged by familial pressures (Chow & Zhao, 1996) or by the expressed preferences of other family members. While it cannot be ruled out that the fathers and/or the

[4] Following Therborn and others, patriarchy is here understood as having two basic intrinsic dimensions: "the rule of the father and the rule of the husband, in that order" (Therborn, 2004, 13-14; also Halpern et al., 1996; Gruber & Szołtysek, 2016). As such, patriarchy refers to generational and conjugal family relations, or, more clearly, to generational and gender relations, thereby encompassing both the stratification by sex in social attainment, and the domination of men over each other based on the seniority principle.

elders in a patriarchal family might seek to increase the reproductive fitness of their own family or lineage by investing in the education and the skills of their successors (see Todd, 1987; recently Le Bris, 2016), it seems unlikely that these male elders would want to invest in the education of the women in the family. These choices in turn have implications for the ability of women to provide information, education, and cognitive skills to their offspring (Kambhampati & Rajan, 2008; Grogan, 2007). However, the lack of interest in investing in education and training may be a more general tenet of patriarchal families.

The senior household heads in patriarchal societies – who are usually also the net users of household resources – have limited incentives to allocate resources to the education of children (often grandchildren or grand-nephews) because the heads are unlikely to be alive when returns to such investments are realised. Moreover, the senior heads may have little inclination to prioritise the acquisition of human capital of the youngest generation because the heads are more distantly related to these children, and therefore feel less responsible for helping the children build their future economic capacities than their biological fathers, who may have a weaker bargaining position in the household (Grogan, 2007, 687).

Viewed from an even broader perspective, it is possible that education and human capital formation of the young, and of females in particular, is considered a threat to parental or spousal authority in settings characterised by strong patriarchy, and is therefore rejected or opposed. Dildar (2015) detected a significant inverse relationship between patriarchy (measured by the internalisation of patriarchal norms by women based on their responses to various statements) and years of schooling in the recent Turkish Demographic and Health Survey. By using a scale of family conservatism in pre-adult socialisation as an instrument for patriarchal norms, she provided support for the view that this relationship might be causal. Furthermore, evidence from traditional family- and kinship-based agrarian economies has shown that the internal underlying patriarchal authority structures are often consciously safeguarded by child-rearing practices that oppose fostering or even allowing competitiveness and individual initiative in children during their upbringing (Caldwell, 1981, 15).

Because in most patriarchal settings children tend to be the net producers (rather than the consumers) of wealth, the mobility of sons in particular may be prohibited or strongly limited (Caldwell, 1982). Other than the land itself, a farmer's sons are his basic resources. If the father's goal is to encourage his sons to stay on the patriarchal farm (and to marry early and patrilocally, for economic reasons), he may discourage them from acquiring knowledge or skills through training or apprenticeship that would enable them to enter into labour and contractual (often also monetary) relationships in the public sphere (see Caldwell 1982, 169; Kwon & Dae-Bong, 2009).

Related to the limited incentives to invest in the education of daughters, in most patriarchal societies parents are afraid of exposing girls to outside influences (Gruber & Szołtysek, 2016). Thus, girls have very limited access to education, employment, and training; and are instead encouraged to become proficient in performing household chores (Kambhampati & Rajan 2008; Grogan, 2007). The human capital acquisition opportunities of women are further constrained through the indirect effects of early marriage. In many patriarchal societies, the family's honour depends on the sexual purity of women, and girls tend to marry at very young ages, what further contributes to suppression of activities of women outside home (Gruber & Szołtysek, 2016; Caldwell, 1981, 10-11; Feldman, 2001, 1099; Szołtysek, 2015b).

Because patriarchal family structures are characterised by customs and attitudes that collectively serve to maximise fertility (e.g., Dyson & Moore, 1983), they create incentives for women to remain subordinate in exchange for receiving support in raising their children. Thus, women have relatively little bargaining power. Pre-adult socialisation practices reinforce the idea that a woman's primary roles consist of being a good wife and a good mother (Caldwell, 1981; Dildar ,2015; Xiao, 1999; De Baca et al., 2014[5]), which in turn may inhibit women from exercising agency in acquiring human capital outside of the domestic sphere. Women therefore become "overspecialised" in reproductive, child-rearing, and domestic work at the cost of accumulating other forms of human capital. This lack of broader knowledge further discourages investments in child quality (Galor & Klemp, 2014).

Another important element of patriarchy is its potential to restrict interactions between the family and the public sphere. Because of its strong emphasis on loyalty to family, lineage, and kin (familism), a patriarchal family structure discourages family members from forming cooperative relationships with non-relatives, and thus limits potentially stimulating "peer group effects" on human capital acquisition (Acemoglu, 2002; also Whyte, 1996, 3-4). Because patriarchal families and societies place a high premium on family loyalty, filial piety, and reverence for ancestors (i.e., a collectivist mindset), they may be less prone to encouraging their members to engage in the types of entrepreneurship, collaboration with non-kin, risk-taking, and innovation that are prerequisites for success in modern societies (Whyte, 1996, 4; Xiao, 1999, 642; Triandis, 2001; also Sinha, 2014, ch. 2).

Naturally, in our analysis we have to take into account potential influences on numeracy and on human capital in general outside of the familial realm. Although research into potential determinants of numeracy in the historical context is still relatively new, scholars have already identified various factors that might explain the observed regional differences in basic quantitative reasoning across Europe, and beyond. Among these potential factors are the role of religion, the extent of formal schooling, the quality of the institutions, levels of land inequality (via the political economy of landlords opposing primary schooling),

[5] For example, cross-cultural research (De Baca et al., 2014) has shown that the presence of patrilineal kin in childhood has a positive effect on the development of patriarchal values later in life. Dildar found that in Turkey "women who have a traditional [patriarchal] mindset (…) put lower value on education and have a preference for family formation instead of having an individual career" (Dildar, 2015, 49).

the degree of integration into markets, geography and natural geographical conditions, and even nutritional advantages (Tollnek & Baten, 2016; also Crayen & Baten, 2010). However, only a few of the existing economic history analyses of this topic have included the elements of family systems when modelling numeracy (see Baten, Szołtysek, & Campestrini, 2017, for a recent example).

Data

In constructing the measures of familial patriarchy and numeracy, we relied on census and census-like microdata from the Mosaic project. The advantages of these data are that they are abundant across historic Europe, they are available in the form of machine-readable, harmonised microdata samples, and they are relatively easy to process (Szołtysek & Gruber, 2016). Table A (Appendix 1) and Figure 1 below show the distribution of the Mosaic data across Europe. The Mosaic database currently covers continental Europe from Catalonia to central Russia. In addition to full-count national censuses, it includes a wide range of historical census-like materials (e.g., church lists of parishioners, tax lists, local estate inventories, local fragments of censuses) that go back to 1700 or even earlier. The Mosaic microdata samples are very similar in their structure and organisation, and in the types of information they provide. Each of these samples contains information on the characteristics of all of the individuals in a settlement or area grouped into households (co-resident domestic groups), and on the relationships between co-resident individuals. Virtually all of the datasets have a core set of variables, including variables on the relationship of each individual to the household head, and on each inhabitant's age, sex, and marital status. As all of the variables are harmonised across space and time using international standards established by leading census microdata initiatives, such as IPUMS-International and the North Atlantic Population Project (NAPP)[6], spatially sensitive accounts of historical localised gender and generational indicators can be generated across multiple locations (Gruber & Szołtysek, 2015; Szołtysek & Gruber, 2016). Because all of the Mosaic data are geo-referenced, they can be linked to a range of detailed GIS-derived covariates.

Because our approach is situated at the meso level of comparative analysis, the units of analysis are "regions". These "regions" are either administrative units used in the respective census or geographical clusters in the absence of applicable administrative units. As a rough guideline, one "region" should have at least 2000 inhabitants, and should include only urban or rural settlements. These regions have been grouped into five larger European territorial clusters designed to capture the varying institutional and socioeconomic characteristics of societies in early modern Europe: "Germany" (German-dominated areas outside of the Habsburg territories), "West" (western and south-western Germany),

[6] https://www.ipums.org/index.shtml; https://www.nappdata.org/napp/

“Habsburg”, “East” (east-central and eastern Europe, including the former Polish-Lithuanian Commonwealth and Russia), and “Balkans” (areas south and/or east of Croatia and Hungary). In total, there are 115 such regions (see Table A in the appendix).

Figure 1: somewhere here

These regions span a large share of the European landmass, and run across many – though not all[7] - important fault lines in the European geography of family and demographic regimes (Szołtysek, 2015a). Furthermore, the regions in the database have a wide variety of geographical features, populations, cultures, and socioeconomic geographies: i.e., plains, mountains, and coastal areas; free and un-free peasantries; a range of ethnicities and religions; and a variety of regional patterns of economic growth in the early modern and modern eras. About two fifths of the 115 datasets contain data collected after 1850, including data from the early 20th century (41.7 per cent); 40.9 per cent of the datasets cover the period 1800-1850, while 17.4 per cent predate 1800. The collection contains both rural and urban regions, although rural regions clearly predominate.

The Patriarchy Index

To account for historical family variation, we apply the Patriarchy Index developed by Gruber and Szołtysek (2016; also Szołtysek et al., 2017). The index combines a wide range of variables related to familial behaviour, such as patterns of marriage, post-marital residence, and headship; as well as to living arrangements in the life course and the sex of the offspring. Table 1 provides the list of the considered components, showing how they are defined and measured, and indicating the expected direction of their relationship with familial patriarchy levels (+/-) (more in Gruber & Szołtysek, 2016).

The components have been chosen with the aim of capturing the essential aspects of the four major dimensions of familial patriarchy: domination of men over women, domination of the older generation over the younger generation, patrilocality, and preference for sons; while taking into account existing data constraints (Szołtysek et al., 2017; Therborn, 2004, 13-14). Most of the component variables directly capture various forms of gender and generational biases at the household level. Other variables, like patrilocality, have been chosen to proxy behavioural patterns that could not be derived directly from our data (in this case, inheritance practices). We preferred applying individual-level age-specific measures to using household-level variables because the former tend to minimise the effect of variation in demographic conditions on indicators of family structure (Ruggles, 2012). This is, for

[7] The current scope of the Mosaic project does not cover the main Iberian and Mediterranean countries, such as Portugal, Spain (except for Catalonia), Italy, and Greece. Data for these countries could serve to delineate the north-south division of European family systems, following the suggestion made by Reher (1998).

example, the reason why in the domain *generational domination* we have chosen the generational patterns of headship, age-specific household formation, and residential patterns of the aged instead of the incidence of three-generation-households[8]. Accordingly, the proportion of the elderly living with a married son (another common demographic measure) was not considered because without locating headship, it is a poor measure of the presence of patriarchal behaviour in the domestic group.

Table 1: somewhere here

Taken together, our components are intended to reflect the extent to which women, the aged, and the young attained certain socially valued resources (positions or statuses) within the familial domestic realm; and thus to represent absolute (not relative) measures of gender and age inequality at the societal level (Szołtysek et al., 2017). Following a strategy detailed elsewhere (Gruber & Szołtysek, 2016), these variables are combined into a single composite measure that allows us to compare the "intensity" of patriarchy across time and space. Regional patterns of patriarchy can be interpreted as indications of societal differences in the extent to which the power, the capabilities, the prestige, and the autonomy of family members varied according to gender and age (cf. Niraula & Morgan, 1996). The distribution of the PI across space is presented in two ways. Figure 2 shows the complete scale of index points arranged according to regional membership and time period, while Figure 3 charts the data geographically.

Figure 2: somewhere here

Figure 3: somewhere here

The observed values of the PI range from 8 to 35 points. In the context of the data we used, it means that while all the regional populations had at least some patriarchal features, as defined above, none of the regional populations could be characterized as fully patriarchal (maximum PI: 40 points). The PI has a rather smooth continuum from very low to very high levels of patriarchy. At the most general level, our ranking of regions is broadly consistent with perceptions in the historical demographic and sociological literature, and seems to confirm the existence of the well-known east-west pattern (Therborn, 2004; also Hajnal, 1965). When we look at the map, we see that to the east and the south of the Danube after it passes Vienna, patriarchal features are much more pronounced than elsewhere on the

[8] Including this former variable in the index would not increase its usability because it is also highly correlated to other components in this domain.

continent; especially in the western Balkans and east of the Bug River in Poland (for a more detailed discussion, see Szołtysek et al., 2017).

Age-heaping-based indicators of human capital

To arrive at a proxy measure of human capital, we follow an established practice of relying on techniques developed around the phenomenon of age-heaping. Baten and his collaborators have long been arguing that the tendency of people to round off their ages to a number ending with a five or a zero can serve as a proxy for the degree to which people could count and calculate (*basic numeracy*); hence, age-heaping can be treated as a measure of human capital in historic periods (Tollnek & Baten, 2016; A'Hearn et al., 2009)[9]. The Mosaic database allows us to use the age-heaping methodology to scrutinise the numeracy patterns of 500,000 men and women across our sample of 115 regions of historic Europe between 1680-1918. However, unlike Baten and other scholars, in assessing age-heaping patterns we apply the Total Modified Whipple's Index developed by Spoorenberg (henceforth W_{tot}; see Spoorenberg, 2007) instead of the more popular Whipple's Index or its linear transformation known as the ABCC Index (A'Hearn, Baten & Crayen, 2009)[10]. Our motivation for choosing W_{tot} is that it detects a wider range of age-heaping patterns than measurements based only on rounding one's age to a number ending with a five or a zero. Thus, our method is well-suited for analysing our data, and for future applications in cross-cultural research[11].

Furthermore, we decided not to apply the cohort approach advocated by Baten and his colleagues (A'Hearn et al., 2009), in which numeracy values are broken down by birth cohort to estimate decadal trends. This is because in our context such an approach would require us to make parsimonious assumptions about the stability of the corresponding patriarchy patterns. Moreover, having both numeracy and family variables derived from data at the same point in time seems preferable for cross-sectional regressions. We are assuming that the W_{tot}

[9] Numeracy is the basic competency of quantitative reasoning; namely, the ability to count, to keep records of one's counting, and to make calculations (Emigh, 2002, 653; A'Hearn et al., 2009, 785). Some scholars have claimed that evidence regarding age-heaping not only provides an additional indicator of human capital, but that given the strong correlations observed between age-heaping and literacy, it has the potential to extend our knowledge of human capital *as such* to times and places for which data on literacy are entirely absent or extremely scarce (A'Hearn et al., 2009, 805–806). Studies of the early introduction of schooling in some historical societies have revealed that arithmetic was taught in schools along with writing skills (See Ogilvie & Kupker, 2015, 8).

[10] Like for the other age-heaping indicators just mentioned, the W_{tot} is computed over the 23-62 age interval in order to limit spurious effects that can influence age awareness at specific moments in life. If there is no age preference, then $W_{tot} = 0$. The theoretical maximum value of the index is 16, which suggests massive heaping. Such value could be reached when all persons report their age on only one similar given age digit (for example, all ages ending in 4 (24, 34, 44, etc.).

[11] Some of our datasets (e.g., from the German territories) had rather unusual patterns of digit preference, which led us to turn to a more sensitive measurement instrument. Overall, the two measures are highly (inversely) correlated when computed for all 115 Mosaic regions (r=-0.975, p<0.01).

is an indicator of the collective human capital of the populations under study. Since in historical populations human capital levels could vary significantly by gender, we calculate the W_{tot} for men and women separately.

Figure 4: somewhere here

Figure 5: somewhere here

Figures 4 and 5 present the distribution of numeracy data across the Mosaic locations by regional membership and time period, for men and women respectively. Across the 115 Mosaic regions, the average W_{tot} is 3.09 among men and 3.65 among women, which indicates a substantial female disadvantage in numeracy (see also Table 2 below). Overall, however, male and female numeracy patterns are highly correlated (r=0.891, p=0.000). When we look below these general patterns, we see considerable variation within countries and across the macro-regions of Europe, including those we have defined. At the most general level, these regularities are reminiscent of the numeracy patterns established in earlier studies (Hippe & Baten 2012)[12], though, however, it is important to keep in mind the unequal distribution of our data across time periods and regions. We observe differences in the numeracy levels of populations in western regions and in central and eastern regions, with the German, the Dutch-Belgian, and the French[13] populations exhibiting higher average numeracy levels across time periods. On the other hand, the populations in the east of Europe were much more diverse than the populations in the west. Although the eastern populations were generally situated at the lower end of the numeracy scale, we also see populations with exceptionally low levels of age-heaping (high numeracy) in parts of 18th-century eastern Poland and northern Ukraine, and in the 19th-century neighbourhood of Moscow. The hot spots with the highest levels of age-heaping were located in areas of modern-day Belarus, southern Romania, and Albania. Again, these patterns generally correspond to the broad spatial patterns shown in previous research (Hippe & Baten, 2012).

A question that most numeracy studies have attempted to address is whether the age-heaping found in the sources reflects the numeracy of the responding individual or, rather, the numeracy of another household resident, such as the registered head of household, or maybe the diligence of the reporting personnel who wrote down the statements (e.g., Tollnek & Baten, 2016, 136; Szołtysek, 2015c)[14]. Although the age data of the relevant age groups were

---

[12] Hippe and Baten used data on 550 regions in Europe between 1790 and 1880. In their data, numeracy ranged between 25 and 100 ABCC points; in our data (if these are converted into ABCC format), the respective range is from 22 to 100 index points.

[13] This applies to the populations in northern France in particular.

[14] It should be noted that several other scholars working on data similar to the Mosaic data came to the reassuring conclusion that their data provide a reliable basis for estimating numeracy levels (see Tollnek &

usually reported by the respondent himself or herself, there is still considerable scope for bias[15].

In order to explore these issues, we performed several tests following suggestions made in the literature (Manzel et al., 2012; also Tollnek & Baten, 2016). First, we compared the numeracy levels of heads with those of other co-resident domestic group members (Manzel et al., 2012), and found only negligible differences – except, perhaps, among some Balkan populations (Figure 6). Second, we compared the numeracy levels of female household heads with those of wives of male household heads (Földvári et al., 2012), and found for all of the Mosaic regions that female heads were no more likely than wives of heads to report their age in a consistent manner (Figure 7).

Figure 6: somewhere here

Figure 7: somewhere here

Finally, we employed logistic regression to assess the probability of reporting an age ending in a zero or a five[16] for individuals aged 23 and older (see Table B, Appendix 2). To do so, we used mixed effects models with age, sex, marital status, and relationship to the head of the household as predictors; while controlling for the random effect of the regional population to which an individual belonged. The models show that there was a strong and robust association of digit preference with age, and that the propensity for age-rounding among spouses, children, and other relatives (except for the parental generation) did not differ from that of household heads. Furthermore, we see a clear downward trend in the propensity for age-heaping from smaller to larger households. This trend corresponds to basic social stratification patterns, at least in rural societies where bigger household sizes usually reflected higher socioeconomic status.

These findings suggest that the potential analytical challenge of heads having reported the ages of everyone in the household may only be relevant for a small fraction of the Mosaic regional populations (which are mostly located in the Balkan regions). Thus, we have confidence that, in general, the data at our disposal can be used for the analysis of numeracy patterns. Overall, our numeracy estimates can be taken as indicative of the degree to which a

Baten 2016). Since, however, the Mosaic collection is relatively new, we decided to perform some of the robustness tests independently.

[15] It is important to note that the diligence of enumerators as well as the details of the questionnaire design must have differed across time and space, especially in early modern times; and these factors may have created spurious variation in numeracy estimates. But as Baten and Crayen (2009, 794-5), made clear, "this criticism applies with equal force to other indicators of human capital, in particular literacy estimates based on signature rates. Some clergy, in some times and places, insisted on signatures in the marriage registry, while others did not, and the choice itself was probably not random.".

[16] For the sake of convenience in the regression specification, we ignored in these calculations other types of digit preference, focusing on the predominant patterns only.

given population had acquired basic quantitative reasoning skills, which is sufficient to pursue our goals.

Results

Patriarchy-numeracy correlation

Figure 8 shows a relatively strong and positive relationship between the PI and the W_{tot}, indicating a negative relationship between our measure of familial organisation and numeracy: across the Mosaic populations, higher PI values tend to be associated with lower levels of numeracy, and the relationship seems particularly steep and strong among women. An increase in the PI by one index point is associated with a 13 per cent higher W_{tot} in for women and 11 per cent higher W_{tot} for men. Among both men and women, the most outlying cases generally come from the Balkans or the East; e.g., several Romanian populations had moderate levels of patriarchy, but very high levels of age-heaping; while some localities in Albania and Bulgaria combined relatively high numeracy values with otherwise strong patriarchal features. Women were particularly disadvantaged compared to men in the Balkan regions.

Figure 8: somewhere here

While Figure 8 depicts an inverse relationship between the PI and numeracy, the former is clearly not the only possible factor associated with quantitative literacy in a given society (see above). To investigate the omitted variable problem, i.e. the question of whether the relationship between patriarchy and numeracy levels is mediated by other observable factors, we added other covariates to our model that could be relevant for understanding variation in numeracy based on theoretical considerations and/or empirical findings. Accordingly, we calculated a series of robust linear regression models[17] with the W_{tot} as our dependent variable, separately for men and women. In addition to the PI, we controlled for a number of socioeconomic, environmental, and institutional characteristics, as well as for the time period. To meet the regression assumptions, we decided to log-transform our W_{tot} measure. In all cases, we employed regressions with regional weights[18] that help to reduce the influence of the populations that are overrepresented in our dataset.

[17] We used the MM-type regression estimator described by Yohai (1987) and Koller and Stahel (2011), which is implemented in the R library *robustbase* (http://projecteuclid.org/euclid.aos/1176350366). Robust regression is less affected by violations of linear regression assumptions, such as those caused by the presence of outliers. Linear regression alone can produce misleading results if unusual cases go undetected; even a single case can have a significant impact on the fit of the regression surface. Robust regression can provide results that are largely unaffected by these problems.

[18] The number of populations from each macro region in our database divided by the number of all researched populations (e.g., number of populations from "Germany" divided by 115).

As spatial data are used in these models, the model estimates may be distorted by spatial autocorrelation (Anselin, 1988). One of the underlying assumptions of a linear regression model is that the sample consists of independently drawn observations. This assumption is often violated in spatial analyses of regional data, as nearby spatial units are likely to share many similarities. Nevertheless, standard regression models treat these adjacent observations as independent, which could lead to biases in coefficient estimates and derived significance levels (see Bivand et al., 2013). In order to assess the presence of spatial autocorrelation in the dependent variables and the model outcomes, we performed a Moran's *I* test[19]. As our regressions include regional weights, we decided not to derive the Moran's *I* for the dependent variable, but instead to calculate a base model that includes the dependent variable, the intercept, and the weights. For the residuals of the base model, we then generated the Moran's *I* on the residuals. The obtained Moran's *I* for the model residuals amounts to 0.56 and 0.70 for male and female numeracy, respectively (p= 0.000); which is indicative of quite high positive spatial autocorrelation (especially in the data for women). This finding provides confirmation that it is important to control for spatial autocorrelation in our model diagnostics. To determine whether the model is able to account for the spatial autocorrelation pattern present in the dependent variable, we decided to perform for each model Moran's *I* tests on the unexplained model residuals. If these tests report insignificant results, this provides reassurance that the specific model estimates are not substantially biased by spatial autocorrelation.

The independent variables used in the regressions are divided into two groups: the first group includes only the PI, since our primary goal is to investigate the robustness of its association with numeracy patterns; while the second group represents control variables aimed at capturing the broad variation in historical living standards, in socioeconomic and institutional frameworks, and in topographic features. The purpose of including these variables in the model is not to increase the model's goodness-of-it, but rather to examine how the estimates of the PI are changing while controlling for other factors.

The child-woman ratio (CWR; see Willigan & Lynch, 1982, 102-104) is a net fertility measure used to account for a possible "quantity-quality trade-off" in human capital investments, and as an indication of the constraints on women's mobility based on the sexual division of labour in the household. It is measured by dividing the number of children under

[19] The Moran's *I* index is very similar to Pearson's product moment correlation coefficient, except that instead of assessing the correlation between the values of two variables *x* and *y* by each unit *i*, it measures the correlation between the values of a variable *x* in each region *i*, with the (weighted) mean value of the same variable *x* in the regions *j* that are adjacent to region *i*. In calculating the Moran's *I*, we considered the five nearest neighbouring regions, derived by calculating the spherical distances between the regions' coordinates. As the regions' coordinates for the Mosaic dataset, we used the population-weighted coordinates obtained from our 1692 Mosaic locations. The Moran's *I* Index can take on values from –1 (strong negative spatial autocorrelation) through zero (no spatial autocorrelation) to +1 (strong positive spatial autocorrelation).

age five by the number of women aged 15-49[20]. We assume that the CWR is positively related with the strength of age-heaping (and is hence negatively related to numeracy). The proportion of the elderly (aged 65+) in each regional population was chosen as a crude approximation of the living standards (Rosset, 1964, 209-210, 231). We thus expect this share to be negatively related with the dependent variable. To control for potentially "de-patriarchalising" factors associated with urban industrial life, the regions were distinguished as being either urban or rural, and were further classified according to the date of census-taking. Based on the previous research, we expect to find that the age-heaping levels were lower in the cities and in the small towns. Furthermore, we distinguished between regional populations who were or were not subjected to serfdom. This approach enabled us to take into account the various channels through which these factors could hinder human capital formation[21].

Two spatial control variables were included following suggestions made in recent economic geography studies, which argued that an unfavourable geographic location may represent a penalty that provides disincentives to invest in human capital (Diebolt & Hippe, 2016; Lopez-Rodriguez et al., 2007). The first of these covariates is terrain ruggedness (Wilson et al., 2007)[22]. Rugged terrain limits a population's opportunities to intensify agricultural activities that might provide an impetus for human development, such as the production of grain and other marketable crops. Populations living in areas with rugged terrain may also face constraints in accessing educational institutions. Moreover, rugged topography frequently represents an obstacle to the construction of transportation

[20] In the child–woman ratio (CWR), the relationship between the number of children and the number of potential mothers is usually multiplied by 1,000. But to avoid small coefficient values in our regression results, we decided to use this ratio without such a multiplication.

[21] The second serfdom hypothesis is commonly cited in the economic history literature as a determinant of slow development (e.g., Kula, 1976). Large landowners prevented the establishment of tax-financed public schooling, as they saw no need for serfs to be educated (and perhaps to learn how to demand political rights) in schools financed by the taxes of the rich. Until the 19th century, the influence of large landowners on east-central European governments ensured that national educational policies provided for comparatively low levels of education. Serfs therefore had relatively few incentives or opportunities to invest in the kind of basic education that would have enabled them to grasp the numeracy concept applied in this study. Finally, because of the serf-based economy was heavy reliant on coerced labour with draught animals (*corvée*), it created structural conditions that led to the acute devaluation of female labour, which in turn negatively affected women's status and agency (see Alesina et al., 2013).

[22] Information on the ruggedness of the terrain has been derived from elevation data from the GTOPO30 dataset, which is a global digital elevation model (DEM) with a horizontal grid spacing of 30 arc seconds (downloaded 30 and 31 August 2016 from http://earthexplorer.usgs.gov/; files: gt30e020n40, gt30e020n90, gt30w020n40, gt30w020n90, gt30w060n90). We use the Terrain Ruggedness Index as applied by Wilson et al. (2007) by employing the focal function in the R package *raster* (formula provided in the help function of "terrain" in the raster package). The calculation is performed separately for each of the 1692 Mosaic locations that form our 115 Mosaic regional populations. Around each location we included all raster points within a diameter of 7.5km centred on the location coordinates for obtaining our ruggedness measure. Based on the data for the 1692 locations, we derived population-weighted values for our 115 Mosaic regions.

infrastructure. This issue was of particular relevance in the period between 1850 and 1950, when railways were the dominant means of transport, and regional access to railways was an important determinant of development prospects. Moreover, as areas with rugged topography may have been more likely than other regions to have maintained their cultural anomalies due to constraints on congregation, communication, and interaction; investments in human capital and skill acquisition may have been inhibited in these areas (see Jimenez-Ayora & Ulubaşoğlu, 2015). On the other hand, places with rugged terrain might have provided the population with opportunities for early proto-industrialisation because they had access to water energy and/or mineral deposits (in mountain areas); which may in turn have fostered human development (Medick, 1976). However, across our pan-European sample, we rather expect to find a positive association between the ruggednes of the terrain and age-heaping levels.

The second geographic variable is population potential (see Stewart, 1942), which accounts for the centrality and the accessibility of a region by determining the size of the population living close to the location of a region. To calculate this variable, we applied spatial weights that give the population living near a given location more weight than a population living farther away[23]. During the period of observation, transport costs were still an important determinant of market access, and thus likely affected the economic growth potential in peripheral regions located far away from important population centres (Redding & Scott, 2003). Furthermore, in pre-modern periods, living in close proximity facilitated the diffusion and maintenance of knowledge and skills, as information and innovative ideas could spread more easily in denser populations (Goldin, 2016, 59). While we might be able to use terrain ruggedness (above) to control for this factor in mountainous areas, there are other areas in our analysis that were peripheral for other reasons, such as the marsh areas in today's southern Belarus. Thus, we believe that the population potential measure could serve as a

[23] To calculate the population potential, we used population data derived from the History Database of the Global Environment (HYDE), Version 3.2. These are available in 10-year intervals from 1700-2000, and we took the data for 1800: http://themasites.pbl.nl/tridion/en/themasites/hyde/index.html. In obtaining the population potential, we restricted ourselves to areas located between a longitude of 60° west and 60° east, and a latitude of 20° and 80° north. We calculated the population potential using the stewart-command in the R library *SpatialPosition* with the following specifications: span=100,000; b=2; typefct= exponential. This operation was done for each coordinate of our 1692 Mosaic locations that form our 115 Mosaic regions. From these data, we then obtained a population-weighted value for the 115 Mosaic regions. It is important to note that the HYDE population data are estimates that are based on historical population estimates and official census counts for countries and sub-regions, present-day information on urban and built-up areas, and a number of assumptions related to historical urban density development (Goldewijk et al., 2010). We did consistency checks by contrasting data from HYDE for 1910 with a polygonal dataset of more than 5000 regions and locations across Europe (see Klüsener et al., 2014 for details on this dataset). Our findings suggest that the HYDE dataset faces some challenges, especially in central and eastern Europe. But we consider the data to be of sufficient quality to allow us to estimate at a European scale whether a Mosaic regional population was located in close proximity to important population centres or in a rather peripheral location.

proxy for centrality in both mountainous and non-mountainous areas. We expect to find a negative relationship between our population potential variable and the degree of age heaping.

Finally, we include dummies for time and region in an attempt to control for other characteristics that cannot be measured directly. To account for the fact that the data for our 115 regions was collected at different points in time, we control for the period in which all or most of the data of a regional population were collected. To do so, the following categories are considered: pre-1800, 1800-1850, and after 1850 (reference category). We assume that age-heaping would decline over time (Hippe & Baten, 2012). In addition, we applied a set of macro-regional dummies intended to capture aspects such as the efficiency of bureaucracy, the role of labour markets, and the role of legal systems. These regional dummies (REGIONS) are Germany, West, East, Habsburg lands, and the Balkans[24]. Descriptive statistics of the variables used in the models are presented in Table 2.

Table 2: somewhere here

Table 3: somewhere here

Table 3 reveals that some of the control variables remain moderately to strongly correlated with the main variable of interest (the PI); and in some cases even more strongly than with the outcome variable (numeracy). While this is not surprising given that the PI is designed to capture a multidimensional phenomenon, it may indicate that many of the candidate control variables are potentially endogenous to it. This suggests that multicollinearity may present a challenge in modelling the PI-numeracy relationship. However, as we show in the results section, our regressions do not appear to be substantially affected by this issue.

Regression results

The regression results are presented in Tables 4-5, separately for females and males. The most important finding is that the association between our focal variable, the PI, and age-heaping levels is highly significant for both women and men in all our models in which we included additional covariates in a stepwise manner. The beta-estimate is in the expected positive direction, and tends to get a bit more attenuated as we move towards the full models with all covariates (from 0.13 to 0.08 among females, and from 0.11 to 0.08 among males). These results support the view that the PI and the age-heaping levels are indeed highly related: the

[24] We are using these regions as general umbrella terms that allow us to control for some other factors for which detailed historical and place-specific information are hard to get or completely unobtainable. We are aware that these macro-regional dummies are quite crude measures, but we consider this approach justified as we use them simply as controls to explore how their introduction affects the association between the PI and our dependent variable.

greater the "patriarchal bias" in the patterning of family organisation at the regional level, the lower the levels of numeracy – and hence, of human capital.

Table 4: somewhere here

Table 5: somewhere here

The outcomes for the other covariates are not the immediate focus of our analysis, but we still would like to briefly discuss them. Among the socio-demographic controls, the child-woman ratio is significant in all of our models, which provides confirmation (if only indirectly) of the "quality-quantity trade-off" hypothesis. The only unexpected significant outcome we obtained was for the serfdom variable, whereby the models for females returned negative coefficients that are significant at the 0.05 level. In order to explore this issue further, we implemented sensitivity checks in which we dropped the dummy serfdom from the full model for females. After doing so, the coefficient for the PI variable was not attenuated and remained highly significant. Because some notable changes occurred simultaneously in the coefficients for some other covariates, including the CWR and the proportion of the elderly, we believe that multicollinearity may have affected the outcomes for the serfdom variable. Our regional dummies indicate that, other things being equal, both women and men in the Balkans and in the Habsburg lands had considerably lower numeracy levels than their counterparts in Germany. Finally, in line with our expectations, age-heaping levels in later time periods were significantly lower than in the period before 1800 (our reference category).

If we turn to the VIF values to explore potential bias due to multicollinearity, only the estimates for serfdom and for the PI appear to require further investigation. However, the elevated VIF values for these two variables seem to be mostly driven by the regional dummies, as they are substantially lower in Model 4 (females) and Model 9 (males) that do not include the regional dummy variable (at values far below three). The outcomes of the Moran's *I* test on the residuals indicate that the estimates for the models in which we just control for the PI (Model 1 for females and Model 6 for males) might be biased due to positive spatial autocorrelation, as the test is in these cases positive and significant. But for all of the other models the results are insignificant[25]. These findings reassure us that the outcomes for the other models are not greatly affected by spatial autocorrelation.

[25] The Moran's *I* on the residuals of Model 5 is significant if the 0.1 level is considered. But as the Moran's *I* of Model 5 indicates negative spatial autocorrelation, this outcome is of less concern, since negative spatial autocorrelation tends to decrease the obtained significance levels.

Conclusions

Recent advancements in New Institutional Economics and its "sister" discipline of New Institutional Economic History have led scholars to take an increasing interest in the family as the potential instigator of economic and developmental change. However, there have so far been relatively few empirical investigations of the links between family organisation and economic growth, or of the forces that underlie this relationship (like human capital formation) across pre-modern European societies; and those findings that exist have not been unequivocal. This paper contributes to this debate by testing the supposed link between family patterns and human capital formation (as proxied by age-heaping methods) using a novel multidimensional measure of familial organisation (the Patriarchy Index) derived from historical microdata that cover a large area of Europe.

Our most important finding is the detection of a significant positive association between the Patriarchy Index and regional age-heaping patterns. This association remains statistically significant even after accounting for other factors in our multivariate regressions. This outcome suggests that the greater the "patriarchal bias" was in the patterning of family organisation at the regional level, the lower were the respective levels of numeracy – and, hence, the levels of human capital[26]. We also provided theoretical considerations in support of the view that this relationship could be exogenous, as patriarchal familial institutions might have affected human capital formation by imposing constraints on individual agency (especially female life course choices and sexuality), and by upholding parental and spousal authority.

However, as soon as we acknowledge that familial behaviour interacts with human capital formation, we have to recognise the possibility that the causality may have run not just from family patterns to numeracy and human capital development, but also from human capital accumulation to familial behaviour; and that both numeracy and patriarchy may have been influenced by some set of underlying factors.

First, we could argue that the higher the level of human capital was, the more likely it was that the presence of this capital led to a reduction in the level of patriarchy, since better educated and skilled women and men would have been influenced by people other than their fathers and husbands, and exposed to new ideas that did not fit within the patriarchal mindset of their forebears. Indeed, evidence from kinship-based agrarian economies shows that the introduction of schooling in patriarchal societies undermined the ability of the older members to claim that they had greater wisdom in the familial and public sphere (Caldwell 1981, 12), and often led to an erosion of strong family ties.

Equally, it is possible that both family structures and numeracy patterns are merely an outcome of a deeper underlying factor, such as a general level of societal development. In a

[26] We also obtained some evidence that the negative association between patriarchy and human capital formation was somewhat stronger among women than among men across our populations.

crude rendering of developmental theory, we could assert that an underdeveloped society is characterised by high levels of patriarchy and low levels of human capital (including numeracy). As this society develops, the level of patriarchy decreases and the level of human capital increases due to growing investments in education. Finally, this society has high levels of human capital and low levels of patriarchy (a kind of ideal version of a "modern western society").

Because it is possible that there is circular causation, or that both family structures and numeracy patterns are merely an outcome of a deeper underlying influence, our regression estimates should be read as indicating an association, rather than a "causal influence" of family organisation on numeracy. Although we cannot address the endogeneity question explicitly at this stage[27], we have nevertheless provided a number of theoretical considerations in support of the view that such causal influences might indeed exist. Clearly identifying the direction of the dominating relationship – if it is feasible at all – would require more detailed data analysis and further testing; and perhaps testing that focuses simultaneously on variation within and across locations or on spatial transition zones with substantial shifts in either patriarchy or numeracy levels. This is a task for the future.

Nevertheless, even if we cannot make any strong claims about the direction of causality at this point, our analysis suggests that family organisation is a promising candidate among the possible explanations of varying human capital levels in early modern Europe. Thus, our findings provide strong support for the view that future research should seek to further improve our understanding of the relationship between family structures and economic development.

[27] A major challenge in this regard would be to identify an appropriate instrumentation strategy based on the introduction of an external source of variation in our focal explanatory variable which could be argued to be exogenous with respect to the dependent variable (numeracy).

Literature

Acemoglu, D. (2002). The Theory of Human Capital Investments. Lecture Notes for Graduate Labor Economics, 14.662, Part 1, Chapter 1, MIT.

A'Hearn, B., Crayen, D., & Baten, J. (2009). Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital. *Journal of Economic History* 69, 783–808.

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

Alesina, A., & Giuliano, P. (2010). The power of the family. *Journal of Economic Growth,* 15, 93–125.

Alesina, A. & Giuliano, P.(2014). Family Ties. In: Handbook of Economic Growth, Volume 2, Chapter 4, pp. 177–215. Elsevier.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Dordrecht et al., Kluwer Academic Publications.

Banfield, E. C. (1958). The *Moral Basis of a Backward Society*. Free Press.

Baten, J., Szołtysek, M., & Campestrini, M. (2017). Girl power in eastern Europe? The human capital development of central-eastern and eastern Europe in the seventeenth to nineteenth centuries and its determinants. *European Review of Economic History* 21(1), 29-63.

Bertocchi, G., & Bozzano, M. (2015). Family Structure and the Education Gender Gap: Evidence from Italian Provinces. *CESifo Econ Stud* 61(1), 263–300.

Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R* (2nd ed.). New York, NY: Springer.

Caldwell, J. C. (1981). The Mechanisms of Demographic Change in Historical Perspective. *Population Studies* 35(1), 5–27.

Caldwell, J. (1982). *Theory of Fertility Decline*. London and New York: Academic Press.

Carmichael, S. G., Dilli, S., & van Zanden, J. L. (2016). Introduction: Family Systems and Economic Development. *Economic History of Developing Regions* 31(1), 1–9. doi: 10.1080/20780389.2015.1132625.

Chow, E. N., & Zhao, S. M. (1996). The one-child policy and parent-child relationships: a comparison of one-child with multiple-child families in China. *International Journal of Sociology and Social Policy* 16(12), 35–62.

Crayen, D., & Baten, J. (2010). Global trends in numeracy 1820–1949 and its implications for long-term growth. *Explorations in Economic History* 47, 82–99.

Das Gupta, M. (1997). Kinship systems and demographic regimes. In D.I. Kertzer & T. Fricke (eds), *Anthropological demography. Toward a new synthesis*, pp. 36–52. Chicago: The University of Chicago Press.

De Baca, T. C., Sotomayor-Peterson, M., Smith-Castro, V., & José Figueredo, A. (2014). Contributions of Matrilineal and Patrilineal Kin Alloparental Effort to the

Development of Life History Strategies and Patriarchal Values: A Cross-Cultural Life History Approach. *Journal of Cross-Cultural Psychology* 45, 534–554, doi:10.1177/0022022113513068.

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. *The Econ Hist Review*, 63: 1–33.

Dennison, T, & Ogilvie, S. (2014). Does the European Marriage Pattern Explain Economic Growth? *The Journal of Economic History* 74, 651–693.

Dennison, T. K., & Ogilvie, S. (2016). Institutions, Demography, and Economic Growth. *The Journal of Economic History* 76(1), 205–217. doi: 10.1017/S0022050716000486.

Diebolt, C, & Perrin, F. (2013). From Stagnation to Sustained Growth: The Role of Female Empowerment. *The American Economic Review* 103(3), 545–549.

Diebolt, C., & Hippe, R. (2016). Remoteness equals backwardness? Human capital and market access in the European regions: insights from the long run. Working Papers of BETA 2016-32, Bureau d'Economie Théorique et Appliquée, UDS, Strasbourg.

Dildar, Y. (2015). Patriarchal Norms, Religion, and Female Labor Supply: Evidence from Turkey. *World Development* 76, 40–61.

Duranton, G., Rodríguez-Pose, A., & Sandall, R. (2009). Family types and the persistence of regional disparities in Europe. *Economic Geography* 85(1), 23-47.

Dyson, T., & Moore, M. (1983). 'On Kinship Structure, Female Autonomy, and Demographic Behavior in India'. *Population and Development Review* 9(1), 35–60.

Emigh, R. J. (2002). Numeracy or enumeration? The uses of numbers by states and society. *Social Science History* 26, 653–698.

Feldman, S. (2001). Exploring Theories of Patriarchy: A Perspective from Contemporary Bangladesh. *Signs* 26(4), 1097–1127

Folbre, N. (1986). Cleaning House: New Perspectives on Households and Economic Development. *Journal of Development Economics* 22(1), 5–40.

Folbre, N. (2012). The Political Economy of Human Capital. *Review of Radical Political Economics* 44(3), 281–292. doi:10.1177/0486613412440240

Földvári, P., van Leeuwen, B., & van Leeuwen-Li, J. (2012). How did women count? A note on gender specific age heaping differences in the 16th-19th century. *The Economic History Review* 65, 304–313.

Foreman-Peck, J. (2011). The Western European marriage pattern and economic development. *Explorations in Economic History* 48(2), 292–309.

Galor, O., & Weil, D. N. (2000). Population, technology and growth. From the Malthusian regime to the demographic transition. *American Economic Review* 90, 806–828.

Galor, O., & Klemp, M. (2014). The Biocultural Origins of Human Capital Formation. NBER Working Paper No. 20474, September 2014.

Goldewijk, K. 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. *The Holocene 20*(4), 565–573.

Goldin, C. (2016). Human Capital. In: Handbook of Cliometrics. Heidelberg, Germany: Springer Verlag, 55–86.

Goode, W. J. (1963). *World Revolution and Family Patterns*. New York: Free Press

Greif, A. (2006). Family Structure, institutions, and growth: The Origins and implications of Western corporations. *American Economic Review* 96(2), 308–312.

Grogan, L. A. (2007). Patrilocality and Human Capital Accumulation: Evidence from Central Asia. *Economics of Transition* 15(4), 685–705.

Gruber, S., & Szołtysek, M. (2016).The patriarchy index: a comparative study of power relations across historical Europe. *The History of the Family* 21(2), 133–174, doi: 10.1080/1081602X.2014.1001769.

Hajnal, J. (1965). European marriage patterns in perspective. In D. V. Glass and D. E. C. Eversley (eds), *Population in history. Essays in historical demography*, pp. 101–143. London: Edward Arnold.

Halpern, J. M., Kaser, K., & Wagner, R. A. (1996). Patriarchy in the Balkans: Temporal and Cross-Cultural Approaches. *The History of the Family* 1(4), 425–442.

Hippe, R., & Baten, J. (2012). Regional inequality in human capital formation in Europe 1790-1880. *Scandinavian Economic History Review* 60, 254–289.

Hanuschek, E. A. (2013). Economic Growth in Developing Countries: The Role of Human Capital. *Economics of Education Review* 37, 204–212.

Jimenez-Ayora, P., & Ulubaşoğlu, M. A. (2015). What underlies weak states? The role of terrain ruggedness. *European Journal of Political Economy* 39, 167–183.

Kambhampati U, & Rajan R. (2008). The 'Nowhere' Children: Patriarchy and the Role of Girls in India's Rural Economy. *Journal Of Development Studies* 44(9), 1309–1341.

Kick, E., Davis, B., Lehtinen, M., & Wang, L. (2000). Family and economic growth: A world-system approach and a cross-national analysis. *International Journal of Comparative Sociology* 41, 225–244.

Klüsener, S., Devos, I., Ekamper, P., Gregory, I., Gruber, S., Martí-Henneberg, J., van Poppel, F., da Silveira, L.E., & Solli, A. (2014). Spatial inequalities in infant survival at an early stage of the longevity revolution: A pan-European view across 5000+ regions and localities in 1910. *Demographic Research* 30, 1849–1864.

Kok, J. (*forthcoming*). Women’s Agency in Historical Family Systems. To be published in JL van Zanden, A. Rijpma & J. Kok (Eds.), Agency, Gender, and Economic Development in the World Economy 1850-2000: Testing the Sen hypothesis.

Koller, M., & Stahel, W. A. (2011). Sharpening Wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis* 55(8), 2504–2515

Kula, W. (1976). *An economic theory of the feudal system. Towards a model of the Polish economy, 1500-1800*. London: NLB [trans. by Lawrence Garner].

Kwon, D.-B. (2009). Human capital and its measurement. Paper presented at the 3rd OECD World Forum on Statistics, Knowledge and Policy *Charting Progress, Building Visions, Improving* Life, Busan, Korea - 27-30 October 2009.

Laslett, P. (1965). *The world we have lost (1st edition)*. London: Methuen.

Laslett, P. (1983). Family and household as work group and kin group: areas of traditional Europe compared. In R. Wall & J. Robin (Eds), *Family forms in historic Europe*, pp. 513–563. Cambridge: Cambridge University Press.

Le Bris, D. (2016). Family Characteristics and Economic Development. Paper presented at the Utrecht Economic History Seminar, May 2016.

Lopez-Rodriguez, J, Faina, J, & Lopez-Rodriguez, J. (2007). Human Capital Accumulation and Geography: Empirical Evidence from the European Union. *Regional Studies* 41(2), 217–234

Macfarlane, A. (1987). *Marriage and Love in England: Modes of Reproduction 1300-1840*. London: Basis Blackwell.

Malhotra, A., Schuler, S. R., & Boender, C. (2002). *Measuring Women's Empowerment as a Variable in International Development*. International Center for Research on Women and the Gender and Development Group of the World Bank. The World Bank, Washington, DC.

Manzel, K., Baten, J., & Stolz, Y. (2012). Convergence and divergence of numeracy: the development of age heaping in Latin America from the seventeenth to the twentieth century. *Economic History Review* 65, 932–960

Medick, H. (1976). The Proto-Industrial Family Economy: The Structural Function of Household and Family during the Transition from Peasant Society to Industrial Capitalism. *Social History* 1, 291–315.

Nimkoff, M. F. (Ed.) (1965). *Comparative Family Systems*. Boston: Houghton & Mifiin.

Niraula, B., & Morgan, S. P. (1996). Marriage formation, post-marital contact with natal kin and the autonomy of women: Evidence from two Nepali settings. *Population Studies* 50(1), 35–50.

Ogilvie, S., & Küpker, M. (2015). Human Capital Investment in a Late-Developing Economy: Evidence from Württemberg, c. 1600 – c. 1900. *Cambridge Working Papers in Economics* 1528.

Redding, S., & Schott, P. K. (2003). Distance, skill deepening and development: will peripheral countries ever get rich? *Journal of Development Economics* 72(2), 515–541.

Reher, D. S. (1998). Family ties in Western Europe: persistent contrasts. *Population and Development Review* 4(2), 203–234.

Rijpma, A., & Carmichael, S. G. (2016). Testing Todd and Matching Murdock: Global Data on Historical Family Characteristics. *Economic History of Developing Regions* 31(1), 10–46. doi: 10.1080/20780389.2015.1114415

Rosset, E. (1964). *Aging Process of Population*. New York: The Macmillan Company.

Ruggles, S. (2012). The Future of Historical Family Demography. *Annual Review of Sociology* 38, 423–441.

Sinha, J. B. P. (2014). *Psycho-Social Analysis of the Indian Mindset*, Springer-India.

Skinner, G. W. (1997). Family systems and demographic processes. D. I. Kertzer & T. Fricke (eds), *Anthropological demography. Toward a new synthesis*, pp. 53–95. Chicago: The University of Chicago Press.

Spoorenberg, T. (2007). Quality of age reporting: Extension and application of the modified Whipples index. *Population-E*, 62(4), 729–742.

Stewart, J. Q. (1942). A measure of the influence of a population at a distance. *Sociometry* 5(1), 63–71.

Szołtysek, M. (2015a). Households and family systems in early modern Europe. In H. Scott (Ed.), *The Oxford Handbook of Early Modern European History, 1350-1750, Volume I: Peoples and Place*. Oxford: Oxford University Press, 313–341.

Szołtysek, M. (2015b). *Rethinking East-central Europe: family systems and co-residence in the Polish-Lithuanian Commonwealth*, 2 vols. Bern: Peter Lang

Szołtysek, M. (2015c). Age heaping and digit preference in eighteenth-century Poland-Lithuania: who was rounding off their age, and why? In P. Guzowski & C. Kuklo (Eds.), Studies on family and household in preindustrial Poland, Białystok: Institute for Research of European Cultural Heritage, pp.163–195.

Szołtysek, M., & Gruber, S. (2016). Mosaic: recovering surviving census records and reconstructing the familial history of Europe. *The History of the Family* 21(1), 38–60

Szołtysek, M., Klüsener, S., Poniat, R., & Gruber, S. (2017). The Patriarchy Index: a new measure of gender and generational inequalities in the past. *Cross-Cultural Research*, advanced online publication; doi: 10.1177/1069397117697666.

Therborn, G. (2004). *Between Sex and Power: Family in the World, 1900-2000*. London, New York, NY: Routledge.

Todd, E. (1987). The *Causes of Progress*: Culture, Authority and Change. Translated by Richard Boulin. New York: Basil.

Tollnek, F., & Baten, J. (2016). Age-Heaping-Based Human Capital Estimates. In C. Diebolt & M. Haupert (Eds.), Handbook of Cliometrics (Springer), pp. 1–20 doi: 10.1007/978-3-642-40406-1_24

Triandis, H. C. (2001). Individualism-Collectivism and Personality. *Journal of Personality* 69, 907–924. doi: 10.1111/1467-6494.696169

Yohai, V. J. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *Ann. Statist.* 15(2), 642–656. doi:10.1214/aos/1176350366

Whyte, M. K. (1996). The Chinese Family and Economic Development: Obstacle or Engine. *Economic Development and Cultural Change* 45(1), 1–30.

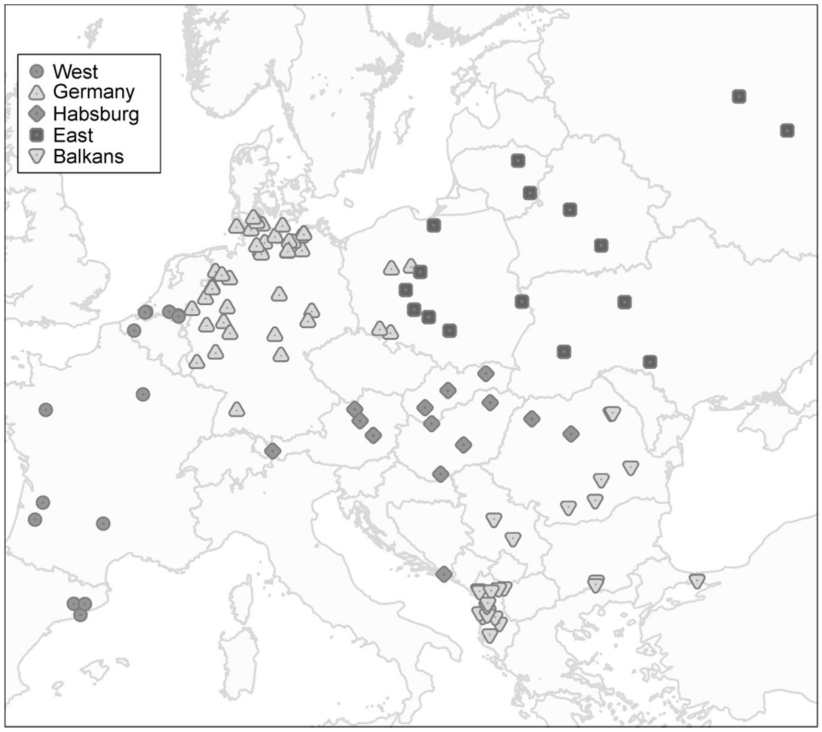
Willigan, J. D., & Lynch, K. A. (1982). *Sources and methods of historical demography*. New York: Academic Press.

Wilson, M.F.J., O'Connell, B., Brown, C., Guinan, J.C., & Grehan, A.J. (2007). Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy* 30, 3–35

Xiao, H. (1999). Independence and Obedience: An Analysis of Child Socialization Values in the United States and China. *Journal of Comparative Family Studies* 30(4), 641–657.

Tables and figures

Figure 1: Spatial distribution of Mosaic data by European region



Source: Mosaic.

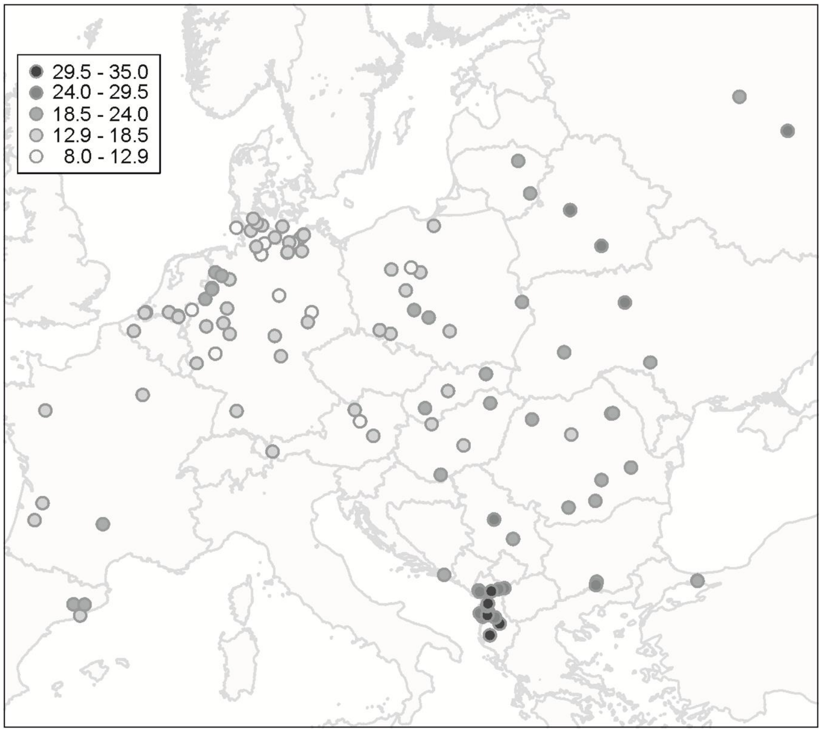
Table 1: Components of the Patriarchy Index

Domain	Component	Abbreviation	Definition/measurement	Relationship with patriarchy	Specification
Male domination	Proportion of female household heads	*Female heads*	The proportion of all female household heads (20+ years) among all adult heads of family households	Negative	Age-standardized
	Proportion of young brides	*Young brides*	The proportion of ever-married women in the age group 15-19 years	positive	
	Proportion of wives who are older than their husbands	*Older wives*	The proportion of all wives who are older than their husbands among all couples for whom the ages of both spouses are known	Negative	Age-standardized
	Proportion of young women living as non-kin	*Female non-kin*	The proportion of women aged 20-34 years who live as nonkin, usually as lodgers or servants	Negative	Age-standardized
Generational domination	Proportion of elderly men coresiding with a younger household head	*Younger household head*	The proportion of men aged 65+ years living in a household headed by a male household head of a younger generation	Negative	Only family households;the elderly men must be relatives of the household head
	Proportion of neolocal residence among young men	*Neolocal*	The proportion of male household heads living without any relatives except spouses/children among ever-married men aged 20-29 years	Negative	Only family households; age-standardized
	Proportion of elderly people living with lateral relatives	*Lateral*	The proportion of people aged 65+ years living with at least one lateral relative in the household	Positive	Only family households
Patrilocality	Proportion of elderly people living with married daughters	*Married daughter*	The proportion of people aged 65+ years living with at least one married daughter in the same household among those elderly people who live with at least one married child in the same household	Negative	Only family households
Son preference	Proportion of boys among the last child	*Boy as last child*	The proportion of boys among the last children (if the last child is one of a set of siblings of both sexes, he or she will be excluded from the analysis).	Positive	Only children (aged 10-14 years) of household heads; family households
	Sex ratio of youngest age group	*Sex ratio*	The sex ratio (boys to 100 girls) in the youngest age group (0-4 years old)	Positive	Only family households

Source: Szołtysek et. al. (2017).

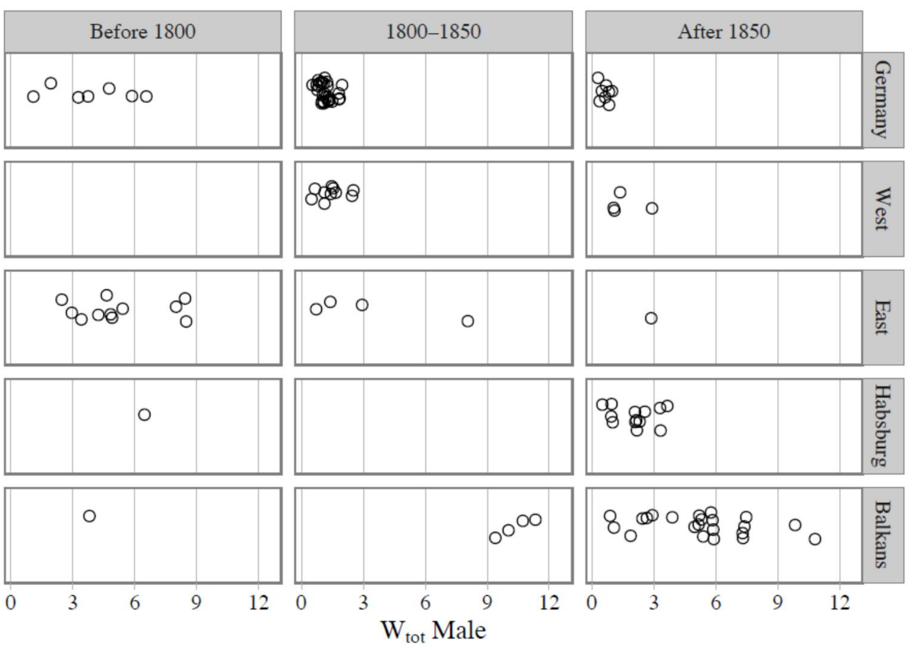
Figure 2: The Patriarchy Index by European regions and time period

Figure 3: Spatial distribution of the Patriarchy Index across Mosaic data (115 regions)



Note. The map is based on a standard deviation categorization centered on the mean of 18.5

Figure 4: Distribution of numeracy by European regions and time period (men)



Note: high W_{tot} implies low numeracy levels.
Source: Mosaic data

Figure 5: Distribution of numeracy by European regions and time period (women)

Note: high W_{tot} implies low numeracy levels.
Source: Mosaic data

Figure 6: Age-heaping patterns of household heads vs. other household members

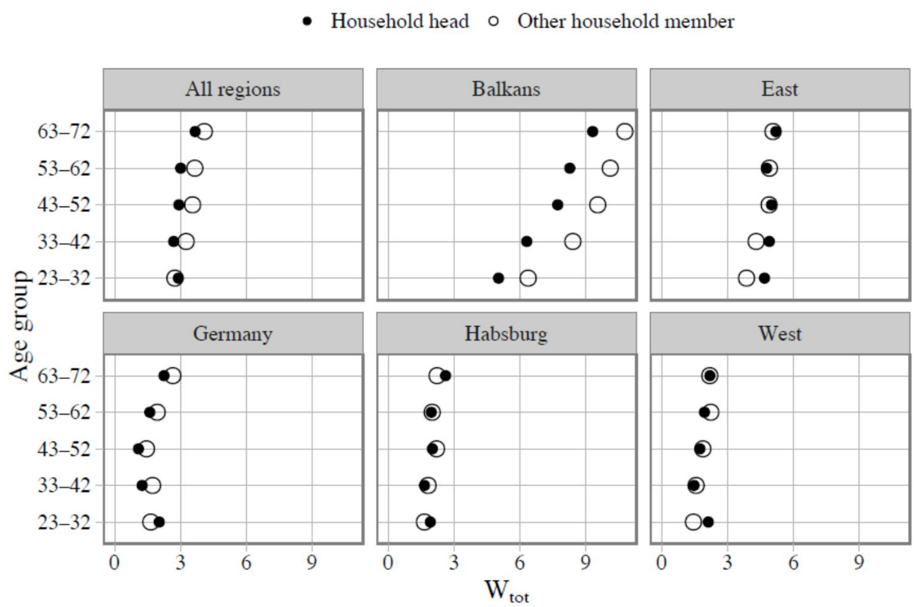


Figure 7: Age-heaping patterns of household heads vs. spouses of the heads (females only)

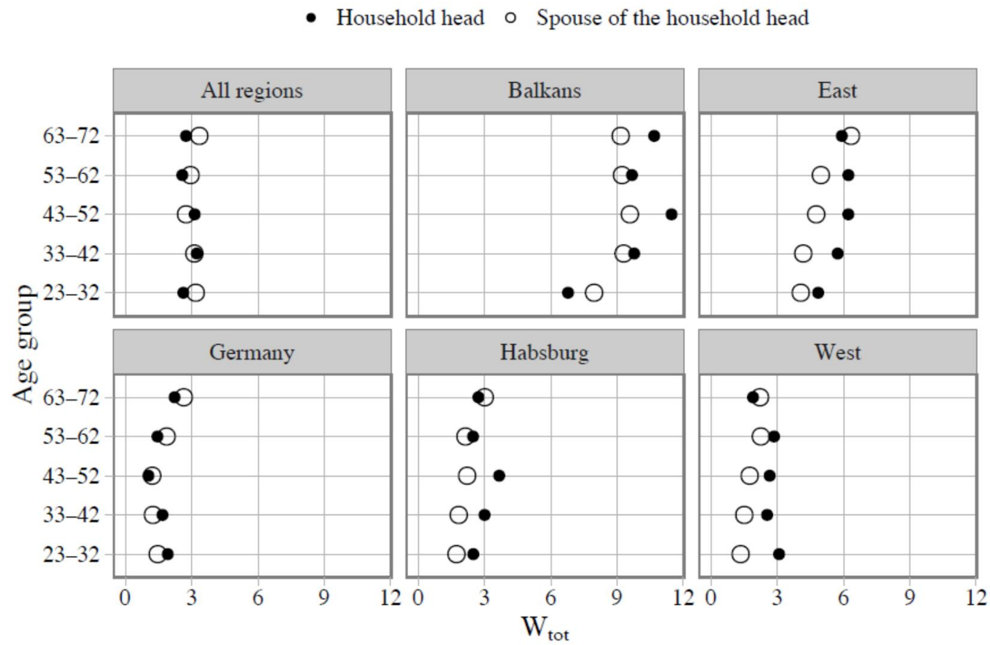


Figure 8: Relationship between the Patriarchy Index and age-heaping (W_{tot})

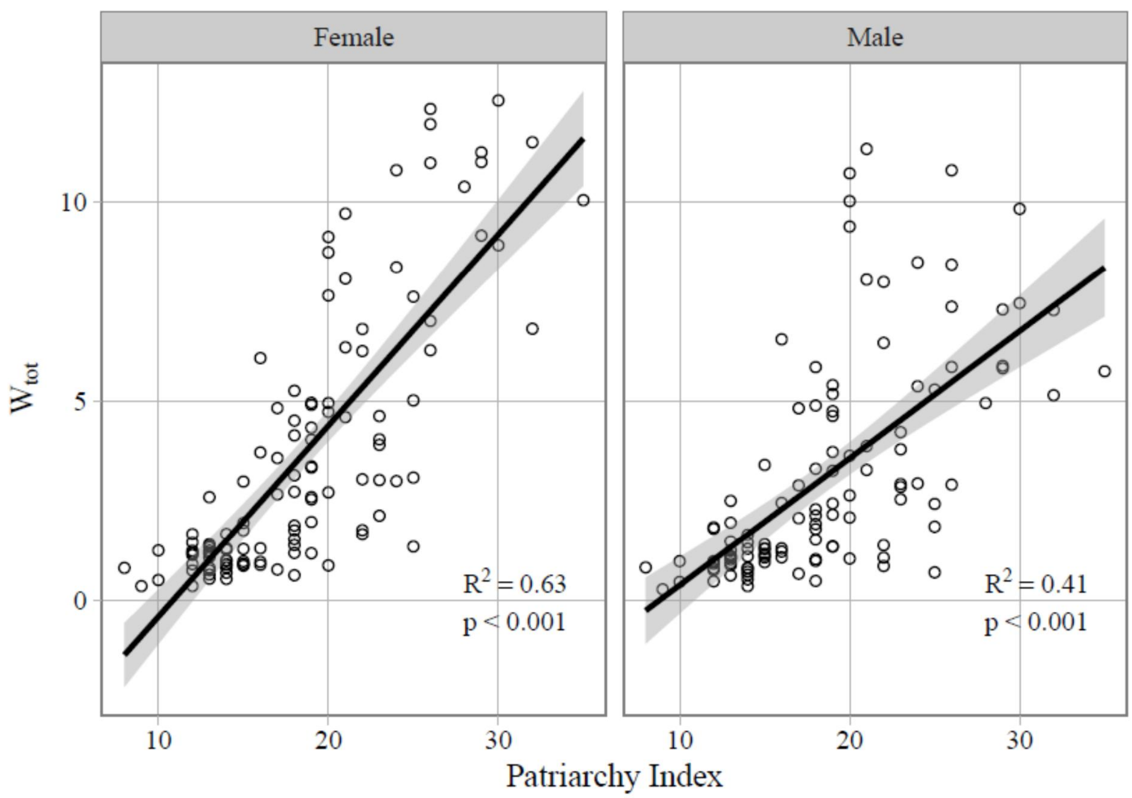


Table 2: Descriptive statistics for all variables

Variable	N	Mean	Median	Std. dev.	Minimum	Maximum
Wtot female	115	3.65	2.14	3.33	0.37	12.54
Wtot male	115	3.09	1.93	2.74	0.29	11.33
Patriarchy Index	115	18.45	18	5.54	8	35
Child-woman ratio	115	0.50	0.49	0.13	0.26	0.92
Elderly population	115	4.52	4.77	1.81	0.58	11.81
Rural (dummy)	115	0.77	1	0.42	0	1
Serfdom (dummy)	115	0.16	0	0.36	0	1
Terrain ruggedness	115	24.44	9.50	35.30	0.15	219.88
Population potential	115	1653000	1315000	1040959	317900	4607000
Region: Germany (dummy)	115	0.38	0	0.49	0	1
Region: West (dummy)	115	0.12	0	0.33	0	1
Region: East (dummy)	115	0.14	0	0.35	0	1
Region: Habsburg (dummy)	115	0.12	0	0.33	0	1
Region: Balkans (dummy)	115	0.23	0	0.43	0	1
Period: before 1800 (dummy)	115	0.17	0	0.38	0	1
Period: 1800–1850 (dummy)	115	0.41	0	0.49	0	1
Period: after 1850 (dummy)	115	0.42	0	0.50	0	1

Table 3: Correlation matrix for variables used in the model

Variable	Patriarchy Index	Wtot female	Wtot male
Patriarchy Index		0.79***	0.65***
Wtot female	0.62***		0.89***
Wtot male	0.65***	0.89***	
Child-woman ratio	0.38***	0.49***	0.60***
Elderly population	-0.19*	-0.15	-0.15
Rural	0.08	0.00	0.13
Serfdom	0.11	0.03	0.16
Terrain ruggedness	0.44***	0.35***	0.21*
Population potential (1800)	-0.56***	-0.47***	-0,36***
Region: Germany	-0.61***	-0.52***	-0.45***
Region: West	-0.18	-0.21*	-0.22*
Region: East	0.19*	0.11	0.22*
Region: Habsburg	-0.01	-0.09	-0.10
Region: Balkans	0.69***	0.74***	0.58***
Period: before 1800	0.10	0.10	0.28**
Period: 1800–1850	-0.49***	-0.37***	-0.27**
Period: after 1850	0.41***	0.29***	0.05

Table 4: Regression results (females)

	Model 1			Model 2			Model 3			Model 4			Model 5			
	β	s.e.	p	β	s.e.	p	β	s.e.	p	β	s.e.	p	β	s.e.	p	VIF
Intercept	-1.60	0.20	0.000	-1.90	0.32	0.000	-1.46	0.34	0.000	0.07	1.63	0.964	0.52	1.54	0.732	
Patriarchy Index	0.13	0.01	0.000	0.12	0.01	0.000	0.12	0.01	0.000	0.12	0.01	0.000	0.08	0.01	0.000	3.84
Child-woman ratio				1.84	0.44	0.000	1.98	0.43	0.000	1.98	0.44	0.000	1.19	0.43	0.006	1.75
Elderly population				-0.01	0.03	0.678	-0.05	0.03	0.073	-0.06	0.03	0.054	-0.07	0.03	0.027	2.04
Rural							-0.22	0.11	0.065	-0.20	0.12	0.085	-0.10	0.11	0.352	1.40
Serfdom							-0.35	0.16	0.035	-0.42	0.17	0.019	-0.41	0.19	0.034	3.72
Terrain ruggedness (ln)										0.00	0.04	0.997	0.00	0.03	0.884	1.84
Population potential (ln)										-0.09	0.10	0.340	0.08	0.09	0.416	2.42
Region (ref. Germany)																
West													0.25	0.13	0.061	1.62
East													0.42	0.21	0.052	4.44
Habsburg													0.50	0.16	0.002	2.42
Balkan													0.92	0.21	0.000	4.38
Period (ref. before 1800)																
1800–1850				-0.31	0.14	0.025	-0.46	0.14	0.001	-0.47	0.14	0.001	-0.49	0.14	0.000	3.21
after 1850				-0.49	0.13	0.000	-0.78	0.16	0.000	-0.82	0.17	0.000	-0.96	0.17	0.000	4.81
N	115			115			115			115			115			
Adjusted R-squared	0.64			0.72			0.75			0.75			0.80			
Moran's *I* Residuals of base model	0.70		0.000	0.70		0.000	0.70		0.000	0.70		0.000	0.70		0.000	
Moran's *I* Residuals	0.14		0.000	0.01		0.1468	-0.04		0.551	-0.04		0.470	-0.01		0.076	

Table 5: Regression results (males)

	Model 6			Model 7			Model 8			Model 9			Model 10			
	β	s.e.	p	β	s.e.	p	β	s.e.	p	β	s.e.	p	β	s.e.	p	VIF
Intercept	-1.34	0.22	0.000	-1.78	0.30	0.000	-1.58	0.33	0.000	-2.54	1.58	0.111	-2.41	1.49	0.109	
Patriarchy Index	0.11	0.01	0.000	0.09	0.01	0.000	0.09	0.01	0.000	0.10	0.01	0.000	0.08	0.01	0.000	3.73
Child-woman ratio				2.53	0.41	0.000	2.66	0.41	0.000	2.62	0.42	0.000	2.02	0.41	0.000	1.73
Elderly population				0.02	0.02	0.438	0.00	0.03	0.853	0.01	0.03	0.837	0.00	0.03	0.848	2.05
Rural							-0.17	0.11	0.116	-0.18	0.11	0.105	-0.13	0.11	0.247	1.40
Serfdom							-0.11	0.16	0.473	-0.07	0.17	0.669	-0.05	0.19	0.781	3.74
Terrain ruggedness (ln)										0.02	0.04	0.518	0.00	0.03	0.915	1.78
Population potential (ln)										0.06	0.09	0.520	0.09	0.09	0.326	2.39
Region (ref. Germany)																
West													0.10	0.13	0.412	1.60
East													0.30	0.20	0.153	4.45
Habsburg													0.52	0.15	0.000	2.42
Balkan													0.79	0.20	0.000	4.21
Period (ref. before 1800)																
1800–1850				-0.59	0.13	0.000	-0.65	0.14	0.000	-0.64	0.14	0.000	-0.64	0.14	0.000	3.25
after 1850				-0.82	0.12	0.000	-0.95	0.15	0.000	-0.95	0.17	0.000	-1.11	0.17	0.000	4.85
N	115			115			115			115			115			
Adjusted R-squared	0.50			0.72			0.72			0.72			0.77			
Moran's *I* Residuals of base model	0.56		0.000	0.56		0.000	0.56		0.000	0.56		0.000	0.56		0.000	
Moran's *I* Residuals	0.16		0.000	-0.04		0.595	-0.06		0.713	-0.06		0.643	-0.04		0.219	

Appendix 1

Table A: Mosaic data used for analysis:

census	regions	N (=pop.)
Albania, 1918 census	8 rural regions, 6 cities	140,611
Austria-Hungary, 1869 census	9 rural regions from Hungary, Romania, Slovakia	31,406
Austria-Hungary, 1910 census	3 rural regions and 1 city from Austria	20,036
Belgium 1814 census	1 rural region from Western Flanders	13,666
Bulgaria, 1877-1947 household registers	1 rural region and 1 city from the Rhodope area	8,373
Dubrovnik, 1674 status animarum	1 rural region from Dalmatia	1,88
Denmark, 1803 census	9 rural regions and 2 urban regions from Schleswig and Holstein	107,861
France, 1846 census	3 rural regions	16,967
France, 1831-1901 census	1 rural region from South-Western France	5,109
France, 1846-1856 census	1 city from South-Western France	5,669
German Customs Union, 1846 census	10 rural regions and 4 urban regions	36,76
German Customs Union, 1858 census	1 rural region from the East	3,468
German Customs Union, 1861 census	1 rural region from the Southwest	6,541
German Customs Union, 1867 census	4 rural regions and 1 city in Mecklenburg-Schwerin	66,938
Germany, 1900 census	1 city	55,705
Mecklenburg-Schwerin, 1819 census	3 rural regions and 1 city	37,332
Münster, around 1700 status animarum	3 rural regions in North-Western Germany	23,01
Münster, 1749 status animarum	3 rural regions in North-Western Germany	34,169
Netherlands, census 1810-1811	2 rural regions and 3 cities in the south	40,037
Poland-Lithuania, 1768-1804 listings	12 rural regions	155,818
Moldavia, 1781-1879 status animarum	2 rural regions	5,291
Wallachia, 1838 census	4 rural regions	21,546
Russia, 1795 revision lists	1 rural region in Ukraine	8,05
Russia, 1814 private enumeration	1 region in Central Russia	2,955
Russia, 1847 enumeration	2 rural regions in Lithuania and Belarus	19,917
Russia, 1897 census	1 rural region around Moscow	11,559
Serbia, 1863 census	1 rural region and 1 city	9,746
Serbia, 1884 census	1 rural region	9,434
Spain, 1880-1890 local census	1 rural and 2 urban regions in Catalonia	23,997
Ottoman Empire, 1885 census	Istanbul	3,408
Ottoman Empire, 1907 census	Istanbul	4,946
Mosaic data overall	115 regions (89 rural and 26 urban)	932,205

Appendix 2

Table B: Logistic regression on the probability of reporting an age ending in zero or five, on selected characteristics, Mosaic data

	Model 1					Model 2					
	β	s.e.	Wald	p	exp(β)	β	s.e.	Wald	p	exp(β)	Group
Intercept	-1.207	0.068	-17.84	0.0000	0.299	-1.335	0.069	-19.21	0.0000	0.263	Fixed
Sex: Male (ref.)											
Sex: Female	0.166	0.011	15.31	0.0000	1.181	0.193	0.013	15.17	0.0000	1.213	Fixed
Age: 23–34 (ref.)											
Age: 35–44	0.380	0.009	41.27	0.0000	1.463	0.386	0.011	34.09	0.0000	1.472	Fixed
Age: 45–54	0.519	0.010	50.63	0.0000	1.681	0.498	0.013	37.24	0.0000	1.645	Fixed
Age: 55–64	0.595	0.012	50.23	0.0000	1.813	0.571	0.016	35.44	0.0000	1.769	Fixed
Age: 65–74	0.749	0.015	49.43	0.0000	2.115	0.697	0.021	33.03	0.0000	2.008	Fixed
With spouse (ref.)											
No spouse	0.031	0.011	2.97	0.0030	1.032	0.125	0.015	8.53	0.0000	1.133	Fixed
Head age multiples by 5						0.356	0.010	37.28	0.0000	1.427	Fixed
Relate: head (ref.)											
Relate: spouse	0.008	0.014	0.57	0.5656	1.008						
Relate: child	0.008	0.015	0.53	0.5965	1.008	-0.081	0.020	-3.99	0.0001	0.922	Fixed
Relate: parent	0.279	0.021	13.48	0.0000	1.322	0.204	0.024	8.39	0.0000	1.227	Fixed
Relate: other	0.011	0.014	0.79	0.4307	1.011	-0.084	0.017	-5.04	0.0000	0.920	Fixed
Relate: non relative	0.184	0.015	12.43	0.0000	1.203	0.134	0.020	6.81	0.0000	1.144	Fixed
HH size:1-4 (ref.)											
HH size:5-6	-0.065	0.012	-5.65	0.0000	0.937	-0.057	0.013	-4.37	0.0000	0.944	Fixed
HH size:7-8	-0.107	0.021	-5.08	0.0000	0.898	-0.117	0.023	-5.07	0.0000	0.890	Fixed
HH size:9+	-0.181	0.024	-7.40	0.0000	0.834	-0.199	0.026	-7.62	0.0000	0.819	Fixed
Sd (Intercept) Region	0.786	NA	NA	NA	2.194	0.735	NA	NA	NA	2.086	Region
	Sigma	logLik	AIC	BIC	df.residual	Sigma	logLik	AIC	BIC	df.residual	
	1	-254939.9	509911.9	510088.2	449057	1	-156847.3	313726.6	313895.2	278077	