HUNGARIAN **DEMOGRAPHIC** RESEARCH INSTITUTE

WORKING PAPERS

ON POPULATION, FAMILY AND WELFARE

Nº 33

GLOBAL AND LOCAL CORRELATIONS OF HAJNAL'S HISTORICAL HOUSEHOLD FORMATION MARKERS: A GEOGRAPHICALLY WEIGHTED APPROACH

by

Mikołaj Szołtysek – Bartosz Ogórek

demografia.hu

WORKING PAPERS

ON POPULATION, FAMILY AND WELFARE

Nº 33

GLOBAL AND LOCAL CORRELATIONS OF HAJNAL'S HISTORICAL HOUSEHOLD FORMATION MARKERS: A GEOGRAPHICALLY WEIGHTED APPROACH

by
Mikołaj Szołtysek and Bartosz Ogórek

Hungarian Demographic Research Institute
2020

ISSN 1588-3396
ISBN 978-963-9597-53-2

Director: Zsolt Spéder
Series editor: Róbert Iván Gál

Suggested citation:
Mikołaj Szołtysek and Bartosz Ogórek (2020): Global and local correlations of Hajnal's historical household formation markers: a geographically weighted approach
Working Papers on Population, Family and Welfare, No. 32
Hungarian Demographic Research Institute, Budapest.

Address: Hungarian Demographic Research Institute
Budapest, Buday László utca 1–3. 1024 Hungary
e-mail: gal@demografia.hu

HUNGARIAN **DEMOGRAPHIC**
RESEARCH INSTITUTE

ABSTRACT

Previous scholarship has hypothesised a global correlation between marriage age, premarital service in husbandry, the extent to which marriage coincided with the assumption of household headship, and the nuclear household structure. According to John Hajnal, these were the axial principles of historical household formation systems. However, whether such correlations apply universally across Europe remains uncertain. We test this possibility by applying both global and local (geographically weighted) measures of correlation to data for 256 rural populations from historic Europe. We demonstrate that the local correlations diverge considerably from the global results. The mutual associations between household formation markers exhibit considerable spatial drifts or important spatial gradients, and the numbers of the joint combinations of these associations far exceed those predicted by Hajnal. We conclude by arguing that the global relationship patterns that Hajnal promoted may lead to incorrect interpretations of historical family systems, and may detract from our understanding of their actual mechanics.

Keywords: spatial nonstationarity; household formation; family systems; census microdata; historical demography; spatial analysis; geographically weighted correlation

ACKNOWLEDGEMENTS

This research has been funded under the POLONEZ 3 scheme granted to Mikołaj Szołtysek by the National Science Centre, Poland (No. 2016/23/P/HS3/03984). The scheme has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 665778. We would like to thank participants in the seminars at the Department of History, University of Münster, and the Hungarian Demographic Research Institute in Budapest, for their comments, which were very helpful for refining the original manuscript.
Siegfried Gruber (Karl-Franzens-Universitaet Graz, Austria) is acknowledged for integrating Mosaic with the selection of NAPP data, and for deriving the rural dataset used for this research. Sebastian Klüsener (Max Planck Institute for Demographic Research in Rostock, Germany) is acknowledged for his support in deriving GIS-covariates for Mosaic and the NAPP data. We thank Miriam Hills for language editing.

AUTHORS

Mikołaj Szołtysek (corresponding author),
Senior Research Fellow, Hungarian Demographic Research Institute, Budapest, Hungary (mszoltis@gmail.com)

Bartosz Ogórek
Pedagogical University of Cracow, Institute of History and Archival Science, Podchorążych 2, 30-084 Cracow, Poland

DOI: 10.21543/WP.2020.33

CONTENTS

List of Figures **4**
List of Tables **4**

1. INTRODUCTION AND BACKGROUND **5**
2. EMPIRICAL METHODOLOGY **7**
2.1. Data scope and quality 7
2.1.1 Operationalisation of Hajnal's household formation markers 9
2.1.2. Analytical strategy 11
3. RESULTS **13**
3.1 Do the four household formation markers correlate globally? 13
3.2. Can these correlations be considered constant over Europe? 13
3.3 What were the local correlations? 14
3.4 What combinations of the relationships between the markers can be found across the study area? 17
4. SENSITIVITY TESTS **18**
5. DISCUSSION AND CONCLUSION **20**
REFERENCES **23**
FIGURES AND TABLES **28**
APPENDIX **36**
SUPPLEMENTAL MATERIAL **37**
A1. Building the rural dataset 37
A2. Systematic bias in smaller regional groupings 38
A3. Assessing "demographically traditional" features of the regional populations 40
A4. Defining the service household formation marker 41
A5. Illustration of the weighting matrices used in the analysis 44
A6. Regional datasets by selected features, including correlation coefficients of all local correlations 46
A7. Data references 59

List of Working Papers **62**

LIST OF FIGURES

Figure 1: Distribution of the combined NAPP/Mosaic data by time period and broad geographic groupings 28
Figure 2: Descriptive mapping of the four household formation markers (Jenks natural breaks) 29
Figure 3: GW Spearman's rho between male SMAM and premarital service (local N=24; tricube kernel) 31
Figure 4: GW Spearman's rho between male service and the CMHD (local N=24; tricube kernel) 31
Figure 5: GW Spearman's rho between male service and household nuclearity (local N=24; tricube kernel) 32
Figure 6: GW Spearman's rho between male SMAM and the CMHD (local N=24; tricube kernel) 32
Figure 7: GW Spearman's rho between male SMAM and household nuclearity (local N=24; tricube kernel) 33
Figure 8: GW Spearman's rho between the CMHD and household nuclearity (local N=24; tricube kernel) 33
Figure 9: Most frequent (regions ≥ 10) combinations of the relationships between the four household formation markers (local N=24; tricube kernel) 34
Figure 10: Changes in the distribution of correlation coefficients between the four variables (in form of the boxplots) as a function of number of nearest neighbours used as threshold for the adaptive bandwidth 35

LIST OF TABLES

Table 1: Global correlation matrix between household formation markers 30
Table 2: Comparison of the global and local correlations between household formation markers 30
Table 3: Differences between alternative and the baseline local (GW) correlations 34

1. INTRODUCTION AND BACKGROUND

Most physical processes are the same irrespective of their specific location, but human and social processes may not be constant over space (e.g., Fotheringham et al., 2002; Matthews and Parker 2013; Weeks 2004; Howell et al. 2016; Anselin and Rey 2010; Siordia and Matthews 2016). Such variance in statistical relationships as a function of geographical location is known as spatial non-stationarity. This concept refers to a situation in which associations among variables unfold variously across a study area, and are different in some places than they are in other places (Brunsdon et al. 1996; Fotheringham 1997; Siorda et al. 2012). Accordingly, any relationship that is not stationary over space will not be well-represented by a global statistic that could be very misleading locally (Fotheringham et al. 2002, 9; Lloyd 2011). Historical demographic literature has long been haunted by such a possibility.[1] Recently, there has been a wave of scholarship providing evidence of spatially dependent relationships between demographic variables in both contemporary and historical settings (Isik and Pinarcioglu 2006; Yang et al. 2013; Klüsener 2015; Vitali and Billari 2017; Szołtysek et al. 2019).

In this paper, we contribute to this agenda by providing the first spatially explicit exploration of John Hajnal's seminal household formation thesis (1982)[2]. Hajnal's model is one of the most pervasive of its kind, and nearly 40 years since its inception, it is still being used, applied, and discussed (e.g., Alter 2013; Wrigley 2014; Engelen and Wolf 2015; Van Zanden et al. 2019; Szołtysek and Ogórek 2020). Its major contribution was providing a general framework linking marriage, premarital service in husbandry, and the extent to which marriage coincided with the assumption of household headship by the husband; i.e., the axial principles of household formation throughout historic Eurasia. It was widely believed that these markers were interdependent, and that the mutual rise or fall in their values led to the emergence of the "simple household systems" in north-western Europe and Scandinavia, and the "joint (complex/grand) family systems" in eastern Europe and Asia (Hajnal 1982, 450, 453, 473, 478).

By furnishing a "necessary" link between service, marriage, headship, and household structure, Hajnal's paper laid the groundwork for a popular, if implicit, set of beliefs: namely, that the proportion of nuclear households inevitably increases as the age at marriage and the incidence of service rise and the time span between marriage and headship decreases; that the age at marriage tends to increase when the incidence of service increases; and that both the age at marriage and the incidence of service rise as the link between marriage and becoming a household head tightens (Smith 1981a, 1981b; Laslett 1983; Wrigley 2014, 18-21; Alter 2013; cf. Rowland 2002, 68-69; Hartman 2004; Engelen and Wolf 2005, 25-26).

The notion that all of Hajnal's household formation markers were generally well-correlated has been powerfully shaped by expectations derived from the well-researched preindustrial English case. Indeed, in that context, new households were formed neolocally; individuals tended to delay marriage by engaging in premarital service; high marital ages helped the service sector to flourish by supplying large numbers of unmarried teenagers and young adults to enter it; and the existence of a service labour market allowed nuclear households to adjust their labour needs without recourse to kin (Hajnal 1982, 455; Laslett 1977; Kussmaul 1981; Smith 1981a, 108; 1990; Macfarlane 1986;

1 Recall, for example, the seminal European Princeton Fertility Project, which found that historical fertility transitions occurred under remarkably diverse socio-economic and demographic conditions across European provinces (Coale and Watkins 1986).

2 An abridged, and slightly changed version of the 1982 paper was published under the same title in Wall and Robin (1983). For the sake of conceptual clarity, this paper sticks as closely as possible to Hajnal's original formulation of household formation rules (1982), and needs to be distinguished from continuing explorations of what has come to be known as the "European Marriage Pattern" (EMP) (Hajnal 1965; Dennison and Ogilvie 2014; Carmichael et al. 2016; Foreman-Peck and Zhou 2018). Despite some inevitable affinity between these ideas, Hajnal's marriage system theory (1965) referred to a neo-Malthusian model involving late marriage for women and high proportions of women never marrying, and as such needs to be treated separately.

Smith 1993; Woods 2004, 382; Hartman 2004; Wrigley 2014, 18-21; Hinde 1990, 86-87; Woods and Hinde 1985; Wrigley 1981, 219; also Klep 2005; Alter 2013)[3].

However, given that the data needed to prove the general validity of this model were patchy at the time when Hajnal wrote his paper, the question of whether the links that he posited would hold universally across Europe remained open. In this paper, we set out to explore this issue by seeking to answer five interrelated questions: Are the four characteristics of household formation – namely, marriage, service, headship, and nuclear household traits – indeed correlated across Europe? Can these correlations be considered constant over space? What are the local correlations; i.e., do these household formation markers display the same associations everywhere across the study area, or is there evidence that the relationships between them are spatially dependent? If spatial non-stationarity is at stake, what possible combinations of the relationships between household formation markers can be observed? And, finally, what implications would this heterogeneity have for our broader understanding of historical family systems?

To provide answers to these questions, we apply historical census microdata similar to those that Hajnal himself relied on, but in far greater quantities, and use them to operationalise the four markers of his household formation rules for 256 regional rural populations from Catalonia to western Siberia. Although scholars have become increasingly suspicious of Hajnal's household formation thesis, most of the empirical counter-evidence either remains confined to extensive case studies, or is derived from small-scale analyses (cf. Szołtysek & Ogórek 2020). While our own data are not without problems (see below), the spatially wide-ranging approach to re-examining Hajnal's thesis they facilitate is the first novelty of this study.

The second novelty is that we provide the first empirical assessment of the mutual correlations between the four household formation markers. Most of the existing Hajnal-inspired literature has focused on how particular populations fit his typology in terms of the combined presence (or levels) of the specific household formation traits (e.g., Smith 1993; Barbagli 1991; Szołtysek and Ogórek 2020). However, Hajnal's second crucial assumption that the four demographic traits would rise or fall together with global regularities has barely been tested. This claim is ripe for reconsideration with "big data" and rigorous methodologies, especially given that since shortly after it was first proposed, Hajnal's view has been challenged by some troubling counterexamples (esp. from Italy) showing that his markers may not always covary in the ways he posited (Barbagli 1991; Rowland 2002, 66–69; Kertzer 1991a, 1991b, 247; Kertzer and Brettel 1987; also Manfredini 2016). While it would be inadvisable to attempt to use these (oftentimes circumstantial) testimonies as the basis for discrediting the general validity of Hajnal's interdependence thesis, even this small amount of counterevidence presents us with one of historical demography's most intriguing puzzles, which merits exploration with more comprehensive data and methodology frameworks than have previously been applied.

The third novelty is that we take the possibility that there are place-specific relationships between Hajnalian markers into account by applying spatially sensitive methodology based on the Geographically Weighted Spearman's rank order correlation (GWrho) (Fotheringham et al. 2002; also Klüsener 2015). In contrast to the global measures of correlation, which assume that the relationships between variables are spatially constant, the geographically weighted approach provides a local version of the correlation statistics by computing correlation coefficients for each geographic location in the data (e.g., each region) based on the concept of geographical weighting. This provides us with significant new insights into the extent to which conventional (global) correlations hide important local variations in the relationships among the focal variables.

3 The joint family system was principally meant to represent a counterpoint to the "western"/English model, and the very scanty data from other parts of the world that Hajnal was able to assemble (1982, 455) seemed to conform to this view.

2. EMPIRICAL METHODOLOGY

2.1. DATA SCOPE AND QUALITY

In his paper, Hajnal declared his intention to deal with "the seventeenth and eighteenth centuries" (1982, 450) in order to capture populations that were – as the paper's title suggested – "preindustrial". It is, however, clear that his notion of "preindustrial" was more generic than specific, as he supported his model with data derived from societies widely separated in time: i.e., he compared 17th-18th-century Iceland, Denmark, the Low Countries and early modern England with late medieval Italy, 19th-century Russia, and even early 20th-century China and India (Hajnal 1982, 455, also 483-485). This made sense given that Hajnal considered his model to be applicable to societies that "were both demographically and economically 'traditional'": i.e., that had a "young" age composition ("something like 43 or more per cent of the population under age 20"), high fertility, and mortality rates; and that were based on households acting as "housekeeping" (production) units (Hajnal 1982, 450).

While Hajnal's hesitancy to identify the precise time frame he was using in his analysis was caused by the scanty and geographically dispersed data he had to rely on (1982, 483-485), it led other scholars to piggyback on non-standard chronology when testing his model[4]. Yet 40 years after Hajnal's paper appeared, obtaining large-scale European comparative data that would be unambiguously suitable for evaluating his thesis remains a challenge. Although massive quantities of historical microdata have been identified and digitised in recent years (Ruggles 2012; Szołtysek and Gruber, 2016), an ideal data structure for the reconstruction of historical household formation patterns for the whole of 17th- or 18th-century Europe does not yet exist, and is unlikely to become available in the future[5].

In this paper, we take stock of the largest public use collections of European historical census and census-like listings currently available, which come from the North Atlantic Population Project (NAPP) and the Mosaic project (Szołtysek and Gruber 2016; Szołtysek and Poniat 2018; Ruggles et al. 2011)[6]. Comprised of full-count national censuses, as well as regional fragments of censuses, list of parishioners, and tax and estate surveys, all of these listings are congruent in terms of their structure and the type of information they provide. They all group individuals into clearly defined residence units in conformity with Laslett's widely accepted notion of the "coresident domestic groups" (housefuls)[7], and provide a core set of common variables, including information about each individual's relationship to the household head, age, sex, and marital status. While there may be some slight normative differences between some of the censuses used, these discrepancies have almost no impact on the household formation markers as defined in this paper[8].

4 For example, in order to fit the North American colonial populations into Hajnal's dual typology, Smith (1993) compared medieval Tuscany, early 19th-century Denmark, and the US census of 1900. The Mediterranean scholarship that was believed to have "buried" Hajnal's household formation model (Kertzer 1991b, 247) merged information from the 18th, 19th, and even the early 20th centuries (see Barbagli 1991; Rowland 2002).

5 Laslett's seminal "English sample of 64 (100) settlements", which itself covered the1574-1821 period, is still not available in a digitised form (Wall et al. 2004).

6 See www.censusmosaic.org; https://www.nappdata.org/napp/. For convenience, we refer to our database as covering Europe, whereas in fact a very small number of data points from Asiatic Russia (i.e., in western Siberia) are also included.

7 Henceforth, the term "housefuls" is used interchangeably with "households""; although the latter are always understood in that former, broader sense.

8 Unlike in Mosaic, some lodgers in the British and the Swedish censuses may have occasionally been treated as separate units by enumerators (Schürer et al. 2018; https://international.ipums.org/international-action/sample_details/country/se). However, the marriage and service markers as defined here (see below) are largely robust to the potential ambiguity of household/houseful borders. The majority of the Mosaic data, as well the British census data, were drawn from de facto enumerations; but the NAPP listings of Denmark, Iceland, and Norway applied de iure criteria. Across all of our data, any units representing institutions rather than family domestic groups were excluded from the analysis.

Since Hajnal related his paper to rural populations (1982, 450-451), a rural dataset was derived[9]. In Mosaic, regions were divided by default into rural and urban categories based on the information provided in the listings. For the NAPP data, the official NAPP/IPUMS rules were followed. These rules identified rural populations based either on the legal status of a settlement (as in Denmark or Norway), or the settlement's population density (as in Great Britain; see more in section A1 of the supplementary material)[10]. Neither of those definitions prescribed that all of the individuals in such places must be engaged in the primary sector, although in the majority of the settlements identified in the Mosaic data, this was usually the case.

In order to conduct our analysis at the comparative meso-regional level, the microdata from 21,559 rural parishes, sub-parishes or communes in the NAPP were aggregated into 150 administrative units that were used in each respective census, and that were considered by the NAPP (generally counties). Accordingly, over 4500 separate Mosaic locations (settlements; parishes; estates) were agglomerated into 106 regions that correspond either to their respective administrative units (usually also counties), or to geographical clusters in the absence of applicable administrative units[11].

Overall, our datafile contains 256 such rural populations, all georeferenced[12]. This dataset captures a large share of the European variation in terms of familial, demographic, and kinship structures. It includes strictly nuclear/neolocal as well as joint family societies, populations with early and late marriage, as well as societies in which life-cycle service was both ubiquitous and rare. By incorporating historical Siberian data, it also covers both the "European" and the "non-European" patterns that Hajnal alluded to (Hajnal 1982). Altogether, our dataset represents a suitable testbed for exploring Hajnal's model of household formation.

Nevertheless, this dataset is not without problems. Only 58 of the 256 regional populations date from before 1800 (22.7 per cent), and their geographical distribution tends to be biased towards eastern and south-eastern Europe, as well Scandinavia. The other 17.6 per cent of the regional populations (N=45) are from the 1800-1850 period, while the remaining 59.8 per cent (mostly Great Britain) date from the post-1851 period (*see Figure 1*). Thus, the limited 17-18th century data used are geographically clustered, while a large share of the populations in the data from what Hajnal might term the "heartland" of the north-western European household formation pattern (esp. England, Wales, and Scotland) come from time periods when the industrial urban revolution was well underway.

However unfortunate this mixing of time periods may be, it is virtually inescapable given the availability of digitised data. Nonetheless, from the vantage point of testing Hajnal's model, what matters more than the temporal imbalances across our data is that the absolute majority of the data fit well within Hajnal's own notion of "demographically premodern" societies. First, most of the regions exhibited a "young age structure" as defined in the 1982 paper (1982, 450): i.e., in 197 out of the 256 regions, the share of the population under age 20 was 43 per cent or higher (and a further 20 regions fulfilled

9 In choosing the NAPP data, we gave preference to the oldest available censuses for Iceland, Denmark, Norway (18th to early 19th centuries), and England (with Wales) (1851); while the earliest NAPP data for Sweden came from the late 19th century (1880). The data for Scotland came from 1881 instead of 1851, because for the latter census it was impossible to derive a rural dataset. Except for England, for which we employed a 10-per cent sample, we used 100-per cent samples. All other data from Great Britain represented 100-per cent samples.

10 https://www.nappdata.org/napp-action/variables/URBAN#comparability_section.

11 An effect of combining full-count censuses (or samples of thereof) with local/regional listings from Mosaic is that there is a considerable range in unit sizes, with 20 per cent of the regions containing fewer than 1000 households. However, there is no evidence that for any of the variables we analyse here, these smaller units (N=58) differ substantially from the rest of the collection *(see more in section A2 of the supplementary material).* It should also be noted that three of the four markers employed in this paper are not household-level, but individual-level age-specific measures known to be less subject to stochastic variations than crude measures of household structure (Ruggles 2012, 431-32). Any remaining biases in this regard are generally mitigated by our reliance on Spearman's rank order correlation which is robust to the presence of strong outliers in the tails of the data distributions.

12 Altogether, our database comprises 11,784,850 persons living in 2,568,908 family households.

Hajnal's "criterion" chronologically)[13]. Second, in the overwhelming majority of our regions (240/256), the populations had not experienced the onset of a monotonic fertility decline. This was the case not only for the populations covered by the relatively late British and Swedish censuses, but also for the populations covered by the chronologically most recent listings (i.e., Albania in 1918; Siberia in 1926) (see more in section A3 of the supplementary material).

Admittedly, a similar generalisation about the "economically premodern" nature of all of our rural populations cannot be made (comp. Hajnal 1982, 477-79). While little can be done to escape this limitation, a partial solution to this problem is discussed in the *Sensitivity section*.

Although the scope of the Mosaic/NAPP data outweighs all preceding efforts to create a European family history dataset, some important areas are still missing, particularly in Italy and the Iberian peninsula (except Catalonia). However, given that Mediterranean scholars have long claimed that there was no straightforward association between any of the Hajnalian markers in those areas (e.g. Barbagli 1991; Rowland 2002), there are good reasons to expect that filling this gap would only strengthen our results[14].

2.1.1 Operationalisation of Hajnal's household formation markers

Hajnal's four household formation markers are operiationalised by means of variables pertaining to the incidence of premarital service, marital age, the marriage-headship nexus, and the share of nuclear households (comp. Smith 1993, 330; Smith 1981b, 600).

Service: Hajnal was interested in agriculture-related farm service in rural areas (1982, 470, 473). In our database, the male population reflects this type of activity better than the female population, most of who were domestic labourers[15]. First, we calculated the proportion of unmarried male servants among unmarried men in four consecutive quinquennial age groups (15-19, 20-24, 25-29, 30-34) representing a typical life-cycle distribution of service. In order to capture the pre-marital aspect of service (Hajnal 1982, 471), instead of choosing one particular age bracket for all populations, the value of the variable was derived for each region depending on that region's corresponding value of the male mean age at marriage[16]. The following principle was applied: when the value of the SMAM in a given region was higher than the midpoint of the five-year age-group (i.e. 17.5, 22.5, 27.5, 32.5 years), the share of servants from the very same age group was considered to compute the variable. When SMAM was lower than that midpoint value, the share of servants from the previous class was applied. Note, however, that because the male service variable is not without problems (see section 4 below; also section A4 of the supplementary material), alternative specifications of the correlations based on female rather than male variables are also explored (see the Sensitivity section).

Age at marriage: The age at marriage is calculated using Hajnal's formula for the singulate mean age at marriage (SMAM) (Hajnal 1953). Given that we computed the service variable for men, we did the same for the SMAM. While there are obviously

13 Obviously, this a very crude measure, which could, in some instances, heavily depend on life expectancy differentials.

14 The four French regions, while obviously small in number, represent census listings from 1846 and 1841 for 30 villages scattered across 28 départements of France. One-third of the villages in the data come from the collection developed by Louis Henry for the reconstruction of the population of France from 1670 to 1829 (all in northern France), and the remaining villages in the data are extensions into other areas.

15 Following the approach used by Hajnal and the Cambridge Group, "servants" were identified using the "relationship to head of household" variable, which was based on the official NAPP/Mosaic coding scheme that was applied uniformly to all our data. In the information on occupational terms – which was, however, available, for only a portion of our datafiles – the absolute majority of these men were just called "servants", or were denoted as farm workers and agricultural labourers *(see more in section A4 of the supplementary material).* For various contexts of female service employment in preindustrial Europe, see, e.g., Whittle (2017).

16 The measure of the mean age at marriage used for this purpose was the singulate mean age at marriage (SMAM; see below).

other aspects worth considering when studying nuptiality regimes besides the age at marriage (Schofield 1985; Engelen and Kok 2003), the SMAM remains the most widely used census-based measure of nuptiality, and it is the one Hajnal himself referred to when setting up his model (1982, 552, and ft. 6 p. 486; also Murphy 2014, 260; Schürer et al. 2018).

Marriage-headship nexus: Hajnal proposed measuring the relationship between *marriage* and *headship attainment* by a visual comparison of the age-specific curves of ever-married men, and the respective proportion of ever-married male heads in a population (Hajnal 1982, 464-466). To deal with the large number of populations being considered, Hajnal's idea was developed into a synthetic measure called the cumulative marriage headship difference (CMHD). Like Hajnal, we start from the two curves, with the area under the proportion of the ever-married curve representing the mean number of years lived as ever-married by an individual in a region, and the area under the proportion of the ever-married heads curve representing the mean number of years lived as the ever-married household head[17]. Strictly speaking, the computation of the CMHD proceeds by approximating the areas under the two curves by means of the trapezoidal rule, and then subtracting them from each other. The result can be understood as the average time that elapsed from getting married to the attainment of headship in the 20–40 age groups[18].

Proportion of simple families: To capture the distinction between simple and complex family households, we followed Hajnal (1982, 451) and used the regional share of nuclear households (the Hammel-Laslett classification). Ideally, allowances should be made to account for some types of stem families in Hajnal's typology (Hajnal 1982, 486, ft. 8; Engelen and Wolf 2005, 16–18; Szołtysek 2015). However, an empirical distinction between different forms of stem families was not made by Hajnal, and it is not workable with cross-sectional census data alone (Berkner 1972; Szołtysek 2016; also Wachter et al. 1978, 29–42)[19]. While our focus on nuclear families represents a necessary compromise on this issue, it is hardly far-fetched. Stem family formations have been characterised as a historical "nullity" in the heartland of the "Northwest European simple household system" (i.e., England), as well as in other parts of north-western Europe covered by our data (Laslett 1977; Wachter et al. 1978). Moreover, at least some of the nuclear families we have captured may represent domestic groups who had already passed the stage of multigenerational coresidence of a stem family type (i.e., the parents had already died).

Figure 2 presents a descriptive mapping of the four variables just described. At the most general level, a contrast can be seen in the relative occurrence of household formation markers between the broadly defined north-western and western parts of Europe, and the eastern and south-eastern portions of our data. However, a closer look reveals that the data for the eastern zone are far from uniform for all but one of the variables (service). For example, populations with marriage-headship patterns or levels of household simplification or late marriage similar to those of populations in north-western and western Europe are shown to be present in various areas of east-central, south-eastern, and eastern Europe; and even in Siberia.

17 Given that the marital ages among the men in our dataset ranged from the twenties to the mid-thirties, and that for the majority of our populations, male headship rates converged towards a plateau among men in their forties, we decided to consider all age groups between ages 20 and 40 for the computation of this measure.

18 The DescTools package (see Signorell et al. 2018) was used for the computation of the CMHD. We could also calculate the "singulate" mean age at household headship (SMAHH) and then subtract the SMAM figures from it (see Szołtysek 2015, 512-14, 529-531). However, this approach is less advantageous because of the quite rigid assumptions we would need to make when computing the SMAHH for populations with large differentials in age-specific headship prevalence.

19 From a census snapshot, we cannot ascertain, for example, whether a multigenerational household headed by a younger generation was formed when one son stayed at home and took over the household upon marriage and his father's retirement (a sequence that conforms with Hajnal's "Northwest European" rules), or whether it resulted from a belated headship transfer to a son who had previously married into the parental household (a sequence that is not compatible with such principles).

2.1.2. Analytical strategy

Our methodological approach proceeds in three steps. In the first step, we explore whether Hajnal's interdependence thesis linking service, marriage, entry into headship, and simple household structure holds for our dataset by computing global correlation coefficients (i.e., not spatially weighted) between these markers. Since all of our variables deviate significantly from the normal distribution (and show the presence of outliers)[20], we selected Spearman's rank order correlation (rho) coefficient (ρ) to derive the correlations. If Hajnalian household formation markers are indeed interrelated, we might expect to observe moderate-to-strong positive or negative correlations between all six pairs of them; i.e., when a given household formation marker steadily increases, the other variable should either increase or decrease[21].

However, such a "global" correlation dictates that from all of the spatial data in our collection, a single statistic must be computed that is essentially an average of the conditions that exist throughout the entire dataset. While this implicitly assumes that the relationships between the variables are spatially constant, they can, in fact, vary geographically. In order to test whether these relationships indeed vary geographically, we employ an explicit spatially sensitive approach based on the Geographically Weighted Spearman's rank order correlation (GWrho) coefficient (see Fotheringham et al. 2002; Klüsener 2015). Our exploration of the GWrho outcomes proceeds in two stages: first by looking at the relative performance of global and local correlations using summary statistics; and then by inspecting the spatial distribution of the local correlation coefficients using maps.

The basic idea behind using a geographically weighted correlation is to check whether the relationships under study vary across space. This is done by moving a search window from one point (i.e., the longitude and latitude coordinates of a region on the map) to the next point in a dataset, working through all of the points in sequence. As the search window rests on a sample point (say, region i), all of the other points that are around it and within the search window are identified. A correlation is then estimated for that subset of the data, giving most weight to the points that are closest to the one at the centre. In our case, for each of the 256 regions, a local Spearman's rank order correlation coefficient is obtained that includes values of the variables of interest for the region itself, and for the neighbouring regions around that observation for which local statistics are calculated. Given the number of markers, a total of 1536 (6*256) local bivariate correlations is obtained.

To select the area the search window covers each time (i.e., which of the neighbouring regions is included in each local correlation), it is necessary to determine the so-called *bandwidth*, either as a geographic distance that is assumed to be the same throughout the dataset (*fixed bandwidth*), or as a discrete number identifying the number of nearest neighbours to include in the local estimation (*adaptive bandwidth*). When an adaptive bandwidth is used, the spatial extent of the search window will vary in area from point to point in order to keep the number of observations within it constant: when the sample points are close together, the window will have less area; and when the points are sparse, it will fill a greater area. Given that the density of our observations varies over space, we determined that an adaptive bandwidth was the most appropriate choice, as it provides constant local information for each correlation depending on whether the data are dense or sparse (see Fotheringham et al. 2002, 57; Gollini et al. 2013).

20 We use the Shapiro-Wilk test with the following results: male SMAM w=0.97902***, male service w= 0.89601***, CMHD w=0.69095***, nuclearity w=0.91387***.

21 The presence of such relationships does not necessarily imply causation.

To determine what specific number of nearest neighbours should be used to establish correlations, we followed optimisation-based procedures implemented in the GW statistical approaches[22]. Six plausible bandwidth sizes corresponding to six pairs of correlations were suggested, with a median bandwidth value of 24. Based on these results, we decided for the sake of clarity to use the 24 nearest neighbours as the sample size for each local correlation (for the impact of bandwidth size on the GWrho results, see sensitivity tests below).

Within the GWrho framework, each local correlation is unique not only because it is based on a specific subset of data (as defined by the adaptive bandwidth), but because coefficients are allowed to vary locally as a result of spatial weighting that gives the neighbouring regions – which are closer to the region for which a local statistic is derived – more weight than the more distant regions (Fotheringham et al. 2002; Tobler 1970). Options for deriving these weights are available through spatial kernel density functions (Silverman 1986; Fotheringham et al. 2002, section 8.4.2), which can be used to determine the amount of weight attributed to each location in the correlation depending on its distance to the point under analysis. Given that our data can be very dispersed, and that we are interested in identifying potential small-scale deviations of the correlation outcomes over space, we decided to use the weighting scheme based on the *tricube* function. This function assigns a nonzero weight that decays with distance to points within the threshold number of neighbours (with the highest weight, equal to unity, situated at the observational point); and a weight of zero to points that are beyond the specified bandwidth, thus nullifying their impact on the local correlation estimate (Wheeler 2019; Bidanset & Lombard 2017, 109; also Klüsener 2015). Compared to other functions, the tricube weighting scheme gives the near neighbours more importance, while reducing the impact of the more distant neighbours. While using such a function differs little from using other similar density functions (like *bi*-square) in regions with dense data, its application is more suitable in places that are relatively isolated (see also section A5 of the supplementary material).

Finally, for the estimation of the significance levels of the obtained local correlation coefficients, we follow the Monte Carlo (randomisation) procedure proposed by Fotheringham et al. (2002, 64, 93-94). To assess the probabilities that the obtained local correlation coefficient differs from a local coefficient found by chance, at each region *i*, the one true GW correlation is compared with the 99 simulated local correlations obtained from randomly rearranging the attribute values across data coordinates within a given bandwidth (Lu et al. 2014)[23].

By concentrating on the relationship between the markers within a determined search area centred on an observational point and its 24 neighbours, in most of the cases, the local correlations include populations from either the same census date (for example, most of the populations in England and Wales) or relatively adjacent time periods. Iceland in 1703, with its spatially and temporally isolated position in our data structure, is perhaps the most prominent exception to this general rule.

22 Since we analyse the bivariate correlations between household formation markers, we base our choice of the bandwidth on the goodness of fit (corrected Akaike Information Criterion) calculated for the analogous bivariate geographically weighted regression models. The procedure is implemented in the GWmodel R package (Lu et al. 2014; Gollini et al. 2015).

23 Command gwss.montecarlo within GWmodel R package (Lu et al. 2014; Gollini et al. 2015). As our aim was not to make claims about causality, but simply to detect correlational associations, these significance levels are, ultimately, of secondary importance.

3. RESULTS

3.1 DO THE FOUR HOUSEHOLD FORMATION MARKERS CORRELATE GLOBALLY?

The results of the global correlation among the four variables of interest – i.e., without regard to their spatial location and structure – are presented in *Table 1*.

The results of the global model are intuitively appealing. For all six possible pairs of variables, there is an apparent moderate-to-strong significant association, and the global bivariate correlations are all in the expected direction. The global pattern suggests that the rankings of the regional values for the incidence of male service are positively associated with the rankings of the male SMAM, but are inversely associated with the marriage-headship time span. The volume of service among men also has a moderate positive correlation with the regional shares of simple households. All three of these features are in line with the predictions of Hajnal's model: the higher the incidence of service that is observed at the regional level, the higher the average age at marriage is, the shorter the average length of time between marriage and headship is, and the greater the incidence of simple households in a population is.

While the mean age at marriage among males is negatively associated with the average length of time between marriage and headship, it also has a weak but significant positive relationship with the prevalence of nuclear households. Moreover, these relationships are in line with those observed by Hajnal and his followers. Within our data, an increase in the average age at marriage coincides with closing the CMHD gap. In populations in which men tend to marry later, a greater incidence of simple family households is also observed – although this relationship seems to be much weaker than expected. Finally, our finding that there is an inverse association between the CMHD and nuclear households suggests – again, in line with Hajnal – that once the association between marriage and headship weakens in a region (the CMHD increases), the share of nuclear households tends to decrease.

3.2. CAN THESE CORRELATIONS BE CONSIDERED CONSTANT OVER EUROPE?

At the level of the global correlation, the basic premises of Hajnal's thesis seem to be confirmed, as the focal variables generally cohere well, and they rise or fall together with the aggregate regularities that Hajnal predicted. But how much realism is discarded in such a global account? Given the spatial variability of our data set, is it possible that the relationships between Hajnalian traits were not the same across space, as the global account would seem to suggest?

In order to examine the potential non-stationarity in these relationships, we assessed the local correlation coefficients for all six pairs of binary associations using the geographically weighted Spearman's correlation (see section A6 of the supplementary material for a list of all place-specific local correlations). *Table 2* repeats some of the information from *Table 1* in order to summarise the relative performance of global and local correlations, and to provide order-based summary statistics of the GWrho, as well as the number of significant local estimates.

As expected, we find that the local correlations diverge considerably from the global results. For all pairs of variables, the relationships between particular household formation markers differ when assessed in separate spatial locations of the dataset. These differences include changes in magnitude, as well in directionality. The local Spearman's coefficients generally have a wide range of values stretching from highly negative to highly positive, indicating spatial non-stationarity. The most striking case in

Table 2 appears to be the association between service and household nuclearity markers. The upper and lower tails of the distribution of the coefficients are nearly equally split on both sides of zero (129 regions show a positive relationship between the two markers; 127 regions show a negative relationship). In addition, for both tails of the distribution, regions can be identified in which the Spearman's ρ is close to one, which indicates the presence of nearly perfect negative or perfect positive associations of ranks for different portions of the same dataset[24]. At the same time, there are regions in which almost no relationships between the ranks of the two variables are observed (i.e., the two variables clearly do not vary together). Albeit to a lesser extent, the other correlation pairs presented in *Table 2* basically display the same general tendency.

Taken together, these results provide a first-level indication that the apparent global bivariate associations between Hajnal's household formation markers can be misleading. In reality, these global correlations may fail to provide a true description of the relationship at any given site, as they offer only an average impression of the relationship over the study area. Instead, it is clear that there are important spatial inequalities in the covariance of the Hajnalian markers across our dataset. In light of these findings, a visualisation with explicit spatial techniques is needed to check whether the observed multidirectional relationships between Hajnal's household formation markers follow any specific spatial patterns.

3.3 WHAT WERE THE LOCAL CORRELATIONS?

To this end, the GWrho results were projected onto geographic coordinates of the regional populations (Fotheringham et al. 2002, 7). The resulting *Figures 3–8* show a continuous surface of correlation coefficients for all six associations from *Table 2*. These maps should be interpreted as follows. The upper panel shows the Spearman rho coefficient for each local correlation; i.e., the direction and the magnitude of the association centred on this specific location and its 24 nearest neighbours, with its weight decaying with increasing distance. The lower panel shows the p-value of the association. The coloured point means that the significant correlation is observed in the specific place and in its neighbours, while the white points indicate the locations where the association (irrespective of its strength and direction) is statistically non-significant. Note that the observed non-significance can imply that the correlation is very weak or non-existent; that the association between the variables in a specific area is non-monotonic; that the significant but opposite relationships of the variables are cancelled out; or that the density of observations in the particular area is too low to capture the local specificity of the correlation, and that the strength of the association is reduced when more distant observations from areas representing considerably different contexts are included. Hence, even the non-significant results might point to places where the relationship between the variables seems to call into question the usual interpretation, especially when the patterns are visually striking (Kievit et al. 2013, 8).

Figure 3 shows the GW correlation between the incidence of male service and the marriage variables. The "global" interpretation of the relationship (i.e., as the incidence of male service rises, the age at marriage increases) is confirmed for the majority of the locations (177/256 regions), and especially for those in England and Wales; much of Scotland; northern and western Scandinavia; and much of central, east-central, and eastern Europe, including Siberia. In a number of these populations – and especially those in northern Denmark, Wales, western and southern Norway, Hungary, Slovakia, and eastern Poland – the local correlations are much stronger than the global correlation

24 Given that Spearman's rho limits the outlier to the value of its rank, this data pattern cannot be attributed solely to the presence of a few very peculiar observations.

(0.545). These findings may be combined with evidence of clear spatial drifts in the direction of the relationships observed in northern Scotland and in parts of Denmark and the German areas; as well as in southern Sweden, where the most extreme reversed (negative) correlations are found. The central Ukrainian and the Balkan regions also belong to this otherwise westward-leaning group (79 locations in total). We can see that both of these broad groupings cover populations that are widely dispersed across the study area. Although the correlations are found to be statistically non-significant for the large majority of the regional populations, significant associations (in both directions) are observed in 24 locations spread over parts of England, Sweden, Germany, and Poland.

The spatial distribution of the local correlation coefficients between the incidence of male service and the CMHD is even more scattered (*Figure 4*). Again, the expected negative direction of the association is confirmed for the majority of cases (152/256), which indicates that the incidence of male service and the marriage-headship nexus increase and decrease in tandem. While populations from nearly all major parts of our dataset converge on this trend without any geographic regularity, these correlations are generally modest in magnitude, and are only rarely stronger than they are in the global account (in 45 out of 152 regions). Furthermore, pockets of spatial non-stationarity in the relationship between the incidence of male service and the CMHD are clearly discernible, as evidenced in the weak-to-moderate positive correlations observed in highland Scotland, France, north-western Germany, around the Baltic Sea, the Balkans, and the Ukraine. It is noteworthy that the strength of the correlations found in some of those areas, especially in Scotland, is far greater than that of the global coefficient. Again, examples of significant negative as well as positive correlations can be observed (23 instances in total).

As anticipated (*see Table 2*), the results indicate that the spatial distribution of the local coefficient correlations between the incidence of male service and the nuclearity variables is more bifurcated (*Figure 5*). Populations are strongly polarised between those with a positive (expected) association between the markers, and those with a counterintuitive (negative) relationship between the markers. In both cases, the coefficients can be considerably higher than they are in the global correlation. The strongest support for Hajnal's view on the service-nuclearity relationship can be found in Iceland, Denmark, northern Sweden, and Norway, but also in western Siberia and Poland. Further instances of a similar, albeit weaker, relationship are found in Flanders, parts of the Netherlands, as well as various parts of east-central Europe and the Balkans. In all of these areas, the regions with higher values for the incidence of male service than those of a given data point generally also have higher proportions of nuclear households. Surprisingly, the findings show that most of England and Wales, together with Scotland (77 out of 87 all British regions covered here), have an inverse relationship between household simplification and the incidence of service. This pattern was found to be particularly strong in parts of western Scotland (e.g., Renfrewshire, Argyll, Bute, and Lanarkshire counties, and their surroundings) and in several regions of Wales (especially those centred around Rednorshire and Carmarthenshire), where negative coefficients of 0.7 or more were found. Most of the German areas, France, Ukraine, and the areas around Moscow, as well as parts of Romania, also conform to this pattern, albeit with less intensity.

Local correlation outcomes for the male SMAM and the CMHD variables are shown in *Figure 6*. The type of relationship suggested by Hajnal – i.e., that the association between marriage and headship tightens (CMHD decrease) as the mean age at marriage increases in a population – and supported by the global correlation can be found in 152 out of the 256 regions. However, this negative association is not only detected in populations that are widely dispersed across space; it varies greatly in strength across different locations. It spreads from England and Wales (which now, unlike Scotland, fits Hajnal's model reasonably well) through northern Scandinavia and east-central Europe, and then continues farther east towards Siberia. The number of locations in which the coefficients' values are higher than they are in the global correlation (–0.567) is small (23); and the places where the negative values are highest are dispersed across Iceland, Austria and Hungary, and north-western Siberia[25]. While the relationship Hajnal suggested clearly exists in these regions, it is highly ambivalent among other locations from this group (e.g., 84 per cent of the regions with an inverse association have Spearman's ρ smaller than the global correlation coefficient).

Furthermore, we again see that the association between the SMAM and the CMHD shifts considerably across space, as a positive correlation between the variables is observed in 41 per cent of all of the locations. Three main hot spots of this pattern can be identified in Scotland, Denmark, and Romania. Other areas with similar, but weaker, relationships can be found in much of central and northern Germany, and in the Balkans. Apparently, all these locations seem to lack a common economic or sociocultural denominator.

The relationship between male marriage and household nuclearity (*Figure 7*) is particularly eye-catching, as most of the local correlations (181/256) are negative, contradicting Hajnal's theory. Except in the English, Welsh, and Scottish regions, which followed this pattern unambiguously, the spatial structure of the negative correlations is patchy. It includes southern France and Catalonia, large parts of Austria, Germany, and Denmark, as well as much of Norway and southern Sweden, but it also extends much further into parts of Poland and Slovakia and towards south-eastern Europe. Overall, however, this spatial distribution highlights quite a number of areas for which the expected positive relationship was habitually modelled in the literature (Hajnal 1982; Imhof 1976). The major hot spots of this inverse association are found in north-eastern England, Scotland, and Wales, as well as in northern Germany and Switzerland.

The 68 cases for which the positive association postulated by Hajnal is found represent equally diverse populations. These cases include regions in northern France, north-western Germany, most of Denmark and Sweden, and southern Norway. Spatially, however, the most coherent case that provides support for Hajnal's view is in east-central Europe and Russia. It is worth noting that in 69 per cent of these positive correlations, the absolute value of the local coefficients is considerably stronger than the global coefficient correlation of 0,192. As before, a small number of locations with significant coefficients for both types of associations can be identified in the data.

The most stationary of all six correlations appears to be that between the marriage-headship nexus (CMHD) and household nuclearity (*Figure 8*). The expected negative association prevails in 213 out of 256 local correlations spread over nearly all major areas covered by our data. While the distribution of the coefficients on these territories is unidirectional, considerable variation in the intensity of the association can still be observed. Regions in which the intensity of this feature is particularly strong intersect our data in a half-ring structure, linking Iceland with south-central Sweden and the Baltics, then with central Ukrainian and the Balkan areas to the south, and with an extension into the Urals. In these areas, the local coefficients are considerably stronger than they are in the global correlation; and in some of these areas, particularly in Iceland, Latvia, and Albania, they basically indicate a perfect inverse association between the markers. Still, in slightly over 50 per cent of the regions in which this negative correlation is identified, the coefficients are lower than the value of the global correlation. Furthermore, a spatial drift in the direction of the relationship is still observable, although it affects a relatively small number of regions, mainly in south-west England and Wales, Austria, and Slovakia.

25 Of all the UK regions, only Herefordshire and Worcestershire have local coefficients stronger than those in the global correlation.

For both types of the association, locations in which the correlation is statistically significant have been identified. Taken together, these findings provide strong support for Hajnal's reasoning that the close co-occurrence of entry into marriage and into headship is intrinsically related to the formation of nuclear households.

3.4 WHAT COMBINATIONS OF THE RELATIONSHIPS BETWEEN THE MARKERS CAN BE FOUND ACROSS THE STUDY AREA?

In the previous section, we showed that mutual associations between pairs of household formation markers vary strongly over space. Our approach was to focus on mapping the results separately for each of the six pairs of bivariate correlations in order to illustrate this heterogeneity. In fact, however, at each observational point, various combinations of these bivariate associations can be overlaid to produce a complex texture of multi-layered correlations. In this section, we take another step towards encoding such combinations from our results.

The single most important finding from this exercise is the identification of 33 (!) region-specific constellations of the concomitant pairwise associations between the Hajnalian markers. While it is possible that some of this evidence can be attributable to mere circumstantial occurrences, or to associations that are too weak to be truly meaningful, the sheer number of these combinations is truly stunning, and no doubt provides ample evidence of linkages between household formation traits that go far beyond those identified by Hajnal or his critics.

Only one of those combinations can be interpreted as fitting a "pure" Hajnalian type as identified in his 1982 paper: i.e., the pattern in which an increase in the mean age at marriage among males is accompanied by an increase in the incidence of service, rising household nuclearity, and a tighter connection between marriage and headship; an increased incidence of service is similarly associated with a rise in the values of the latter two traits; and a longer period of time between marriage and headship is negatively linked with the share of nuclear households (*see Table A* in the Appendix; also black dots in *Figure 9*). Other than that, our regional populations are prone to stark larger or smaller departures from this pattern. The positive relationship between the age at marriage and the incidence of service on the one hand and the CMHD on the other, paired with the inverse relationship between marriage/service and simple household structure; or the negative association of SMAM with the service and nuclearity variables, concomitant with the positive relationship between marriage/service and the CMHD; are just few examples of the multi-linkages that have been found, but that should not, theoretically, exist. Although a complete reversal of the relationships suggested by Hajnal is not found in our data, in a number of regions, such a mirror image is very close to being observed; e.g., in the southern Netherlands and in Albania (see the regions dotted pink in *Figure 9*; also *Table A* in the Appendix).

We should also note that although the "Hajnalian" pattern (see above) is the most represented among the combinations, it features in only a very small fraction of all of the regions (33 out of 256) dispersed over the large areas. It is especially noteworthy that this group does not include a single region from Great Britain or the Netherlands, but instead converges mostly among east-central and eastern European regions, with extensions over Iceland and parts of Scandinavia. Other relationship patterns are more regionally clustered. In particular, the Danish-southern Swedish conglomerate, Wales with the West Midlands, and the Scottish and the Norwegian regions each displays a specific intermingling of bivariate patterns not observed elsewhere in the study area. In general, however, the spatial structure of the combinations does not follow any meaningful geographical order.

4. SENSITIVITY TESTS

There are three potential criticisms of our results. First, they may be sensitive to alternative measurements of some of the household formation markers or other perturbations in the dataset (cf. Neumayer and Plümper 2017, 26). In particular, one could argue that focusing on male rather than female service (and its related marriage variables) has a specific impact on our results, even though it yields the best approximation of Hajnal's concept of the "agricultural farm service" (see above). The risk that some of the correlations may not perform as Hajnal predicted is especially high in Victorian Britain, where the incidence of male service was much lower than it was in the early modern era (Kussmaul 1981; Humphries 2004, 250–253)[26]. The second major criticism is that whatever the "premodern" demographic nature of the British rural regions covered by our data (see the *Data section*), because their populations do not fully conform to Hajnal's notion of a "traditional" rural economy (e.g., Overton 1996; Gritt 2002), they may not fulfil the conditions under which the Hajnalian linkages were to be made visible.

In order to check how much impact these potential issues have on our results, the stability of the baseline local correlations was tested against two alternative runs of GWrho: one in which the male service and marriage variables were replaced with their female equivalents[27]; and one in which the baseline set of male variables was used, but the British data were removed from the correlations.

However, even after those changes were introduced, the major thrust of our results remained basically the same. The order-based summary statistics of the two alternative correlations (*see Tables B* and *C* in the Appendix) are very similar to those shown in *Table 2* above, and reflect the familiar pattern of strong spatial non-stationarity for all pairs of variables. These findings are further confirmed in *Table 3,* which shows that the median absolute difference in the values of six pairwise associations is moderate at best between the baseline and the female-based correlations (0.013–0.237), but is entirely negligible when Great Britain is removed (0.046–0.075). More importantly, the overall stability in the directionality of the local coefficients is retained in the alternative correlations. Depending on the type of test applied, between 77 and 96 per cent (female variables), or even 92 and 99 per cent (omission of Great Britain) of all alternative correlations had the same sign as in the baseline results. Furthermore, the set of combinations derived based on the female variables has a very similar composition (available upon request), with the "Hajnalian" pattern again dominating, but covering a similarly small fraction of all regions (25 out of 256) in absolute terms. Thus, we conclude that for the absolute majority of cases, changing some of the markers or curtailing the dataset has little impact on the outcomes of local relationships.

The third potential problem is that our main results may be too susceptible to changes in the value of the bandwidth with which we are working (Fotheringham et al., 2002)[28]. To check the general impact of the bandwidth size changes on our baseline results, changes in the distribution of correlation coefficients between the six pairs of markers (in the form of the boxplots) are shown in the figure below as a function of the number of nearest neighbours used as the adaptive bandwidth threshold (*Figure 10;*

26 While this concern may be crucial in the context of the north-western "heartland", for the remainder of our data, it is largely irrelevant.

27 Unlike male service, female service was much less responsive to the changes associated with the transition to a capitalist/industrial economy, and the demand for female domestic servants remained high throughout the 19th century in Great Britain (Cooper 2004, 287). Note also that the measures of male and female premarital service are strongly (and significantly) correlated with each other (Spearman's rho= 0,846).

28 To some extent, this is just a property of the geographically weighted correlation, which inevitably changes as the size of a bandwidth changes. Recall also that our choice of a bandwidth resulted from the optimisation procedures based on the corrected Akaike Information Criterion.

comp. Klüsener 2015). The results indicate that up to very large bandwidth sizes, the correlation coefficients come from both sides of zero in all of the cases. The chosen number of nearest neighbours proposed by the optimisation procedure explained above (24 NN – marked with a solid vertical line) provides a good representation of the overall pattern. While the small numbers of NN result in very high variation of the correlation coefficients (from -1 to one in all cases), and would certainly pose a problem of statistical significance (a bandwidth of five NN would result in a local regression with a mere six observations), the large ones (close to 100) produce converging sets of coefficients that approach the global statistic, and that ignore the local specificity of the regions.

Taken together, the outcomes of these tests suggest that the spatially varying relationships between household formation markers as captured in the baseline correlations cannot be taken as artificially resulting from the specific operationalisation of certain focal variables, from the presence of highly specific and clustered British data, or from the chosen size of the bandwidth. Although we cannot rule out the possibility that if our dataset had a different chronological composition, it would have produced correlations with different spatial structures, the ability of such data permutations to seriously undermine our findings remains doubtful, at least within the limits of currently available data[29].

29 For example, the baseline geographically weighted approach has been applied to a dataset in which the NAPP input data were changed so that they could, in some cases, replace the oldest censuses used in the baseline correlation with later censuses (i.e., Iceland 1701 was changed into 1801; Denmark 1787 was changed into 1801; Norway 1801 was changed into 1865; England and Wales 1851 was changed into 1881; and Sweden 1880 was changed into 1890). Despite such aggressive changes being introduced, the new local correlations have produced patterns that are very similar to the baseline results: on average, more than 80 per cent of the regions retained the direction of the relationship (detailed results are available upon request). It is also noteworthy that the largest changes were observed for the correlations between headship attainment and household nuclearity; i.e., for the association that is likely to be the most prone to corresponding shifts related to the agricultural modernisation of the societies.

5. DISCUSSION AND CONCLUSION

While scholars have long been calling for a re-specification of Hajnal's hypothesis (Barbagli 1991; Kertzer 1991a; Rowland 2002; also Engelen and Wolf 2005; Szołtysek 2008), this task has not yet been successfully performed on a European-wide scale due to the unsystematic character of the historical evidence and the dearth of more advanced methodologies. In this article, we attempted to move beyond these limitations by applying an explicitly spatial approach to an unprecedented family history database. Our data and methodology framework allowed us to provide the first empirical assessment of the meso-level correlations between marital age, life-cycle service, the marriage-headship nexus, and household nuclearity – which Hajnal thought of as the axial principles of household formation across Eurasia. The empirical evidence presented here provides support for the suggestion made in some prior studies that Hajnal's model is a gross simplification of historical reality that should be approached with caution.

While our results build upon those of previous family historians taking a critical stance towards Hajnal's thesis, we believe that adding the finer point of spatial variability places the evaluation of his household formation model on a new footing. The spatial distributions of historical family systems have interested scientists for well over 100 years. Nonetheless, the theoretical framework for interpreting spatial or place-based patterns in historical family organisation is still incompletely developed, and is thus poorly "spatialised". This is also the case for Hajnal's discussion of the relationship between the markers of household formation systems, which largely ignored the possibility of the existence of spatial dependence and variation.

What, then, can be learned from the spatial analysis of Hajnal's seminal model that would not have otherwise been gleaned from the non-spatial "classical" approach? The single most important finding is that of the complexity and heterogeneity of the "world we have lost". Our key results show that whereas Hajnalian model associations are derived in a global correlational framework, when mutual associations between nuclear household structure, service, headship, and marriage are examined locally, their relationships turn out to be highly spatially variable. In addition to finding links that run in the opposite direction of those put forward by Hajnal and his followers, we provided a first-hand demonstration that the relationships between these variables are prone to considerable spatial drifts (sign reversals) or important spatial gradients (magnitude changes). Taking these arguments a little further, we suggested that a "pure" Hajnalian type of association could be detected in only a very tiny part of our European database, and that this pattern was largely absent in areas commonly seen as the "heartland" of the "north-western European simple household system". Finally, one of the main contributions of this paper is that we uncovered associations and regularities that were "hidden" in global correlations; and, hence, that we found relationships between different household formation traits that go beyond those identified by Hajnal and his critics. Notably, we also found that simple geographic determinism does not seem to explain the ways in which the four key parameters covary across historic Europe and western Asia. Apparently, neither the "lines" nor the "zones" that earlier scholars have suggested to summarise the heterogeneity of these historic regions are useful for doing so (e.g., Hajnal 1965; Laslett 1983).

This unexpected finding of spatial variability lends support to the view that Hajnal's proposition cannot be upheld. Although it may prove useful as a crude heuristic device for understanding preindustrial family systems (Smith 1993), it is certainly doubtful that a generalised version of Hajnal's household formation thesis can contribute much to the actual analysis of particular societies. Indeed, relating Hajnal's universally claimed household formation tendencies to the features of particular populations (or regions) within Europe (and perhaps within Eurasia, as well) may turn out to be highly unproductive. Placing too much reliance on the global relationship patterns that Hajnal hypothesised is likely to result in incorrect interpretations of historical family systems, and to detract from our understanding of their actual mechanics.

By challenging the proposition that the four Hajnalian traits have unambiguously coevolved across preindustrial Europe, our findings make the quest to develop a generalised model of household formation largely untenable. Clearly, when spatial phenomena are being examined in historical demographic research, no firm conclusions can be drawn until local statistics have been taken into account (cf. Matthews and Parker 2013). But including these local statistics comes at a price. While the application of geographically weighted approach to Hajnal's seminal model may enable us to come to terms with the spatially unique aspects of the relationships between household formation markers, whether it could be used to produce a general body of knowledge about these relationships remains uncertain.

If non-stationarity exists, then it is possible that different processes are at work within a study region. Thus, in our case, investigating which processes are at work becomes the most pressing question. While our results converge on what has long been the major thrust of anthropological demography – i.e., that beneath the average (global) performance of human populations, there is always the high degree of contextual (spatio-temporal) variability of demographic phenomena (e.g., Kertzer and Fricke 1997) – to extend our analysis beyond this general point, we would need additional data and different methodologies. Future comparative research will have to take a stance on various aspects of the nuanced geography revealed here, and use appropriate tools to nest the variable interdependence of historical household formation markers within their local contexts. It is, however, clear that in order to understand the observed variability, complex explanations that take into account a combination of demographic, political-economic, legal, and ecological factors will be needed. It has long been known that family regimes are adaptive outgrowths formed in response to variability in social, cultural, and environmental contexts (e.g., Berkner and Mendels 1978; Kertzer 1991b; Rudolph 1992; Mitterauer 1996; Szołtysek 2015). Accordingly, these contexts are, in turn, likely to influence the interdependence of the household formation markers.

Clearly, the various ways in which the interdependencies between household formation markers manifest themselves in a given society are related to the spatial context in which that society exists. Human social behaviour can vary intrinsically over space (Thrift 2009). Moreover, a combination of various contextual political-economic, institutional, and/or environmental characteristics may influence household formation strategies differently in different places, either by shaping variations in people's attitudes or preferences regarding entry into service, marriage, or headship; or by having different effects on local administrative, political, or other contextual constraints on how these strategies can be executed in different places. This complex tangle of spatially dependent contextual factors may foster relationship patterns that differ from those posited by Hajnal. For example, these interdependencies could make it possible to delay marriage without being bound to entry premarital service; to allow post-marital lodging to serve as an alternative to household formation; or to make it acceptable to acquire a household headship at marriage by taking over the parental farm without recourse to strict neolocality.

Acknowledging such complexities prompts us to challenge Hajnal on yet another premise. Given the sheer variability of the patterns revealed in this paper, it seems doubtful that the emergence of two (or, in fact, any number of) household formation systems can be attributed to a single global relationship pattern between their markers, as Hajnal seemed to believe (Hajnal 1982, 449). Our findings also suggest that we should be wary of expecting to observe a strict spatial congruence between the typology and the cartography of European household formation patterns based on the *levels* of the Hajnalian traits (Smith 1993; Barbagli 1991; Szołtysek and Ogórek 2020) and the patterns of their mutual correlations. While it appears that geographically clustered groupings of similar levels of household formation markers can be identified in historic Europe (Szołtysek and Ogórek 2020; also Faragó 1998), given the inconsistencies in the observed relationships between the markers, identifying spatially coherent correlational patterns that underlie these clusters would be very hard, if not impossible. Thus, it is conceivable that some important community- or meso-level non-familial institutions had more influence on the constitution of household formation "systems" than the mutual ups and downs in the values of their markers.

These discrepancies appear puzzling, but they seem to be pointing to a crucial aspect of the problem; namely, the possible equifinality in the operation of historical household formation traits. The idea that different contextual and demographic processes (including household behavioural rules) are able to produce similar structural outcomes resonates well with Wrigley's classic notion of the preindustrial family and household organisation pattern as a "repertoire of adaptable systems" (Wrigley 1981, 182). Acknowledging this possibility will have important implications for further investigations of the geography of European household formation systems. Scholars may need to decide, for example, which of the approaches are better suited to advancing their work: comparing populations solely on the basis of observed levels of household formation traits, or assessing populations based on the relationships between these traits.

Moreover, scholars seeking to explore this problem further should give serious consideration to potential hermeneutic limits of the measures that Hajnal proposed. It is probable that his four household formation markers do not actually cover all of the axial principles of the family systems in historic Europe (and even less so beyond it), and that there are other elements that may need to be taken into account when investigating the differences and the commonalities among regional family systems within Europe (e.g., Gruber and Szołtysek 2016). For instance, it might help if scholars used a less categorical approach in the operationalisation of the marriage marker. It could, for example, be argued that the age at marriage per se may be insufficient to adopt as a measure of nuptiality, and that more attention should be paid to proportions never-married, pre-marital births, and extra-marital fertility – which may not always move together in a unified manner either (Wrigley et al. 1997; Engelen and Kok 2003). Future scholars should also seek to find new ways of accommodating the stem family in broader assessments of the correlation between household formation markers (cf. Engelen and Wolf 2005), and to try to link Hajnal's hypothesis with the variable patterning of the living arrangements of the aged (Szołtysek et al. 2019).

To this end, the spatial distribution of the relationships between the four household formation markers envisaged here should serve future micro-oriented studies aimed at providing a better understanding of the local intermingling of various demographic variables, and of how they change over time. Having a better understanding of these micro- and meso-level intricacies can make it easier for scholars to develop a new mid-range theory of household formation behaviours. Future research will show whether we are bound to either a very general (and misleading) global theory, or a unique description. It is to be hoped that this paper will add to the current momentum of research on historical household formation systems, and will help historical demographers fulfil their duty to provide meaningful generalisations about human behaviour in the past.

REFERENCES

Alter, G. (2013). Generation to Generation: Life Course, Family, and Community, *Social Science History* 37(1), 1–26.

Anselin, L., and Rey, S. (2010). *Perspectives on Spatial Data Analysis*. Berlin: Springer.

Barbagli, M. (1991). Three Household Formation Systems in Eighteenth- and Nineteenth-Century Italy. In Kertzer, D. I. and Saller, R. P. (eds): *The Family in Italy from Antiquity to the Present*. New Haven– London: Yale University Press, 255–269.

Berkner, L.K. (1972). The stem family and the developmental cycle of the peasant household: an eighteenth-century Austrian example. *The American Historical Review* 77(2), 398–418.

Berkner, L.K., and Mendels, F. (1978). Inheritance systems, family structure, and demographic patterns in Western Europe, 1700–1900. iIn Tilly, C. (ed.),: *Historical studies of changing fertility*. New Jersey: Princeton University Press, 209–225.

Bidanset, P., and Lombard, J. R. (2017). Optimal Ker-nel and Bandwidth Specifications for Geographically Weighted Regression: An Evaluation Using Automated Valuation Models (AVMs) for Mass Real Estate Appraisal. In Lombard, J., Stern, E. and Clarke, G. (eds): *Applied Spatial Modelling and Planning*. Abingdon: Routledge, 107–120.

Brunsdon, C., Fotheringham, S., and Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis* 28, 281–298.

Carmichael, S. G., De Pleijt, A., Van Zanden, J. L., and De Moor, T. (2016). The European Marriage Pattern and Its Measurement. *The Journal of Economic History* 76(1), 196–204.

Coale, A. J., and Watkins, S. C. (eds.) (1986). *The Decline of Fertility in Europe.* Princeton: Princeton University Press.

Cooper S. M. (2004). From Family Member to Employee*:* Aspects of Continuity and Discontinuity in English Domestic Service, 1600–2000. In Fauve-Chamoux, A. (ed.),: *Domestic service and the formation of European identity. Understanding the globalization of domestic work, 16th-21st centuries*. Bern: Peter Lang, 277–296.

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

Engelen, T. and Kok, J. (2003). Permanent Celibacy and Late Marriage in the Netherlands, 1890-1960, *Population* 58(1), 67–95.

Engelen, T. and Wolf, A. P. (2005). Introduction. In Engelen, T. and Wolf, A. P. (eds): *Marriage and the Family in Eurasia*: *Perspectives on the Hajnal Hypothesis*. Amsterdam: Aksant, 15–34.

Faragó, T. (1998). Different household formation systems in Hungary at the end of the 18th century: variations on John Hajnal's thesis. *Historical Social Research* 1–2, 83–111.

Foreman-Peck, J. and Zhou, P. (2018). Late marriage as a contributor to the industrial revolution in England. *The Economic History Review* 71:, 1073–1099. doi:10.1111/ehr.12651

Fotheringham, A.S. (1997). Trends in Quantitative Methods I: Stressing the Local. *Progress in Human Geography* 21, 88–96.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression: the analysis of spatially varying relationships*. Chichester: Wiley & Sons.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2013). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *arXiv* preprint arXiv:1306.0413:5

Gritt, A. J. (2002). The "Survival" of Service in the English Agricultural Labour Force: Lessons from Lancashire, c. 1650—1851. *Agricultural History Review* 50(1), 25–50.

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

Hajnal, J. (1953). Age at Marriage and Proportions Marrying. *Population Studies* 7(2), 111–136.

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

Hajnal, J. (1982). Two kinds of preindustrial household formation system. *Population and Development Review* 8, 449–494.

Hartman, M. S. (2004). *The household and the making of history. A subversive view of the western past*. Cambridge: Cambridge University Press.

Hinde, A. P. R. (1990). Resources and the fertility transition in the countryside of England and Wales. In Landers, J. and Reynolds, V. (eds): *Fertility and resources*. Cambridge: Cambridge University Press, 76–91.

Howell, F. M., Porter, J. R. and Matthews, S. A. (eds) (2016). *Recapturing Space: New Middle-Range Theory in Spatial Demography*. Dordrecht: Springer.

Humphries, J. (2004). Household economy. In Floud, R. and Johnson, P. (eds): *The Cambridge Economic History of Modern Britain*. Cambridge: Cambridge University Press. 238–267.

Imhof, A. E. (1976). Ländliche Familienstrukturen an einem hessischen Beispiel: Heuchelheim 1690-1900. In Conze, W. (ed.): *Sozialgeschichte der Familie in der Neuzeit Europas*. Stuttgart: Klett, 121–135.

Işik, O. and Pinarcioğlu, M. M. (2007). Geographies of a silent transition: a geographically weighted regression approach to regional fertility differences in Turkey. *European Journal of Population* 22, 399–421.

Kertzer, D. I., and Brettell, C. (1987). Advances in Italian and Iberian Family History. *Journal of Family History* 12(1-3), 87–120.

Kertzer, D. I. 1991a. Household history and sociological theory. *Annual Review of Sociology* 17, 155–179.

Kertzer, D. I. 1991b. Introduction to Part Three. In Kertzer, D. I. and Saller, R. P. (eds): *The Family in Italy from Antiquity to the Present*. New Haven- London: Yale University Press, 247–249.

Kertzer, D.I. and Fricke, T. (1997). *Anthropological* demography. *Toward a new synthesis*. Chicago: The University of Chicago Press.

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., and Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in psychology* 4(513), doi:10.3389/fpsyg.2013.00513.

Klep, P. M. M. (2005). An adult life before marriage: Children and the Hajnal hypothesis. In Engelen, T. and Wolf, A. P. (eds): *Marriage and the family in Eurasia. Perspectives on the Hajnal hypothesis*. Amsterdam: Aksant, 241–270.

Klüsener, S. (2015). Spatial variation in non-marital fertility across Europe in the twentieth and twenty-first centuries: recent trends, persistence of the past, and potential future pathways. *The History of the Family* 20(4), 593–628.

Kussmaul, A. (1981). *Servants in Husbandry in Early Modern England*. Cambridge: Cambridge University Press.

Laslett, P. (1977). Characteristics of the Western family considered over time. In Laslett, P.: *Family life and illicit love in earlier generations. Essays in historical sociology*. Cambridge: Cambridge University Press, 12–49.

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

Lloyd, C. D. (2019). *Local Models* for *Spatial Analysis (2nd Edition)*. CRC Press.

Lu, B., Harris, P., Charlton, M. and Brunsdon, C. (2014). The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science* 17(2), 85–101.

Macfarlane, A. (1986). *Marriage and Love in England*: *Modes of Reproduction*, *1300–1840*. New York: Basil Blackwell.

Manfredini, M. (2016). Family Forms in Later Periods of the Mediterranean. In Huebner, S. R. and Nathan, G. (eds): *Mediterranean Families in Antiquity*. Oxford: Wiley, 310–323.

Matthews, S. A., and Parker, D. M. (2013). Progress in Spatial Demography. *Demographic Research* 28, 271–312.

Mitterauer, M. (1992). Peasant and non-peasant family forms in relation to the physical environment and the local economy. *Journal of Family History* 17, 139–159.

Murphy, M. (2014). John Hajnal (1924-2008). *Biographical Memoirs of Fellows of the British Academy* 13, 251–269.

Neumayer, E., and Plümper, T. (2017). *Robustness Tests for Quantitative Research* (Methodological Tools in the Social Sciences). Cambridge: Cambridge University Press. doi:10.1017/9781108233590.

Overton, M. (1996). *Agricultural Revolution in England: The Transformation of the Agrarian Economy 1500–1850*. Cambridge: Cambridge University Press.

Rowland, R. (2002). Household and Family in the Iberian Peninsula, *Portuguese Journal of Social Science* 1(1), 62–75.

Rudolph, R. L. (1992). The European Peasant Family and Economy: Central Themes and Issues. *Journal of Family History* 17(2), 119–138.

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

Ruggles, S., Roberts, E., Sarkar, S., and Sobek, M. (2011). The North Atlantic Population Project: Progress and prospects. *Historical Methods* 44, 1–6.

Schofield, R. S. (1985). English Marriage Patterns Revisited. *Journal of Family History* 10, 2–10.

Schürer, K., Garret, E. M., Hannaliis, J., and Reid, A. (2018). Household and family structure in England and Wales (1851–1911): continuities and change. *Continuity and Change* 33, 365–411.

Signorell, A. et mult. al. (2018). DescTools: Tools for descriptive statistics. R package version 0.99.26. Available at https://cran.r-project.org/web/packages/DescTools/DescTools.pdf

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Siordia, C., Saenz, J., and Tom, S. E. (2012). An introduction to macro-level spatial nonstationarity: a geographically weighted regression analysis of diabetes and poverty. *Human Geographies – Journal of Studies and Research in Human Geography* 6(2), 5–13.

Siordia, C. and Matthews, S. A. (2016). Extending the boundaries of place. In Howell, F. M., Porter, J. R. and Matthews, S. A. (eds): *Recapturing Space: New Middle-Range Theory in Spatial Demography*. Dordrecht: Springer, 37–56.

Smith, R. M. (1981a). The People of Tuscany and Their Families in the Fifteenth Century: Medieval or Mediterranean? *Journal of Family History* 6, 107–128.

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

Smith, R. M. (1990). Monogamy, landed property and demographic regimes in pre-industrial Europe: Regional contrasts and temporal stabilities. In Landers, J. and Reynolds, V. (eds): *Fertility and Resources*. Cambridge: Cambridge University Press, 164–188.

Smith, D. S. (1993). American Family and Demographic Patterns and the Northwest European Model. *Continuity and Change* 8(3), 389–415.

Szołtysek, M. (2008). Three kinds of preindustrial household formation system in historical Eastern Europe: A challenge to spatial patterns of the European family*, The History of the Family* 13(3), 223–257

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

Szołtysek, M. (2016). A stem-family society without the stem-family ideology? The case of eighteenth-century Poland, *The History of the Family* 21(4), 502–530.

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

Szołtysek, M. and Poniat, R. (2018). Historical family systems and contemporary developmental outcomes: what is to be gained from the historical census microdata revolution? *The History of the Family* 23(3), 466–492.

Szołtysek, M. and Ogórek, B. (2020). How Many Household Formation Systems Were There in Historic Europe? A View Across 256 Regions Using Partitioning Clustering Methods. *Historical Methods* 53(1), 53–76.

Szołtysek, M., Ogórek, B., Poniat, R., & Gruber, S. (2019). "One size does not fit all": Spatial nonstationarity in the determinants of elderly residential isolation in historical Europe, *Population, Space and Place* 25(6). doi:10.1002/psp.2242

Thrift, N. (2009). Space: The fundamental stuff of geography. In Clifford, N. J., Holloway, S. L., Rice, S. P. and Valentine, G. (eds): *Key Concepts in Geography* (2nd edition). London: SAGE, 85–96.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46(2), 234-–240.

Van Zanden, J. L., Carmichael, S., and De Moor, T. (2019). *Capital Women. The European Marriage Pattern, Female Empowerment and Economic Development in Western Europe 1300–1800*. Oxford: OUP.

Vitali, A., and Billari, F. C. (2017). Changing Determinants of Low Fertility and Diffusion: a Spatial Analysis for Italy. *Population, Space and Place* 23, doi: 10.1002/psp.1998.

Wachter, K. W., Hammel, E. A., and Laslett, P. (1978). *Statistical Studies of Historical Social Structure*. New York: Academic Press.

Wall, R., Laslett, P. and J. Robin (eds) (1983). *Family forms in historic Europe*. Cambridge: Cambridge University Press.

Wall, R., Woollard, M., and Moring, B. (2004). *Census schedules and listings, 1801–1831: an introduction and guide.* Colchester: University of Essex, Department of History. https://www.academia.edu/619532/Census_Schedules_and_Listings_1801-1831_An_Introduction_and_Guide, accessed 1 August 2019.

Weeks, J. R. (2004). The Role of Spatial Analysis in Demographic Research. In Goodchild, M. F. and Janelle, D. G. (eds.): *Spatially Integrated Social Science: Examples in Best Practice*. New York: Oxford University Press, 381–399.

Wheeler, D. C. (2019). Geographically weighted regression. In *Handbook of regional science*, 1–27.

Whittle, J. (ed) (2017). *Servants in Rural Europe 1400-1900.* Woodbridge: Boydell and Brewer.

Woods, R. I. (2004). *The Demography of Victorian England and Wales*. Cambridge: Cambridge University Press.

Woods, R. I., and P. R. A. Hinde. (1985). Nuptiality and Age At Marriage in Nineteenth-Century England. *Journal of Family History* 10(2), 119–144.

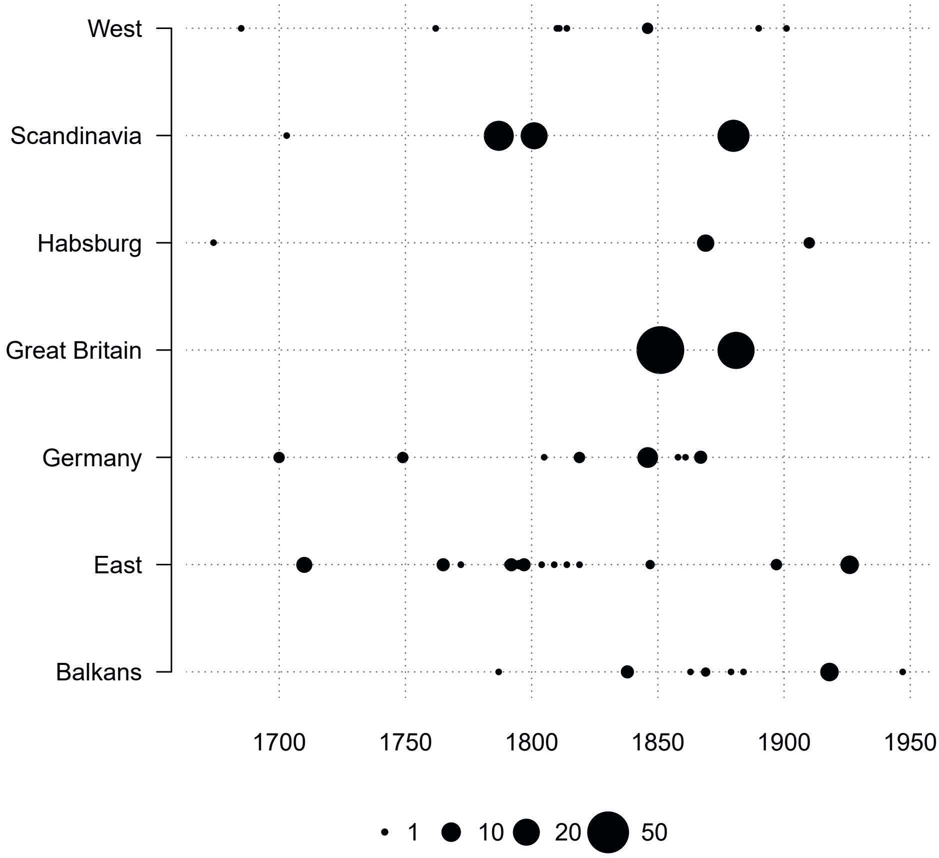
Wrigley, E. A. (1981). Marriage, fertility and population growth in eighteenth century England'. In Outhwaite, R. B. (ed.): *Marriage and society: Studies in the Social History of Marriage*, 17–34. London: St. Martin's Press, 137–185.

Wrigley, T. (2014). European marriage pattern and their implications: John Hajnal's essay and historical demography during the last half-century. In Briggs, C. D., Kitson, P. and Thompson, S. (eds.): *Population, welfare and economic change in Britain, 1290-–1834*. Woodbridge: Boydell Press, 15–41.

Wrigley, E. A., Davies, R. S., Oeppen, J. E. and Schofield, R. S. (1997). *English Population History from Family Reconstitution 1580–1837*. New York: Cambridge University Press.

Yang, T. C., Shoff, C., Matthews, S. A. (2013). Examining the spatially non-stationary associations between the second demographic transition and infant mortality: A Poisson GWR approach. *Spatial demography* 1, 17–40.

FIGURES AND TABLES



Note: Seven bigger territorial groupings on the right-side panel of the figure followed major institutional and socioeconomic distinctions across historic Europe. "Great Britain": England, Wales, and Scotland; "Scandinavia": Danish, Swedish, and Norwegian data, as well as Iceland; "Germany": German-dominated areas other than the Habsburg territories; "West": areas west and south-west of Germany; "Habsburg": Austrian, Hungarian, Croatian, as well as Slovakian data; "East": east-central and eastern Europe, including the former Polish-Lithuanian Commonwealth and Russia; "Balkans": areas south and/or east of Croatia and Hungary.

Source: Mosaic/NAPP data. For primary sources of the Mosaic and NAPP data and the full list of items, see section A7 of the Supplementary Material.

Figure 1

Distribution of the combined NAPP/Mosaic data by time period and broad geographic groupings

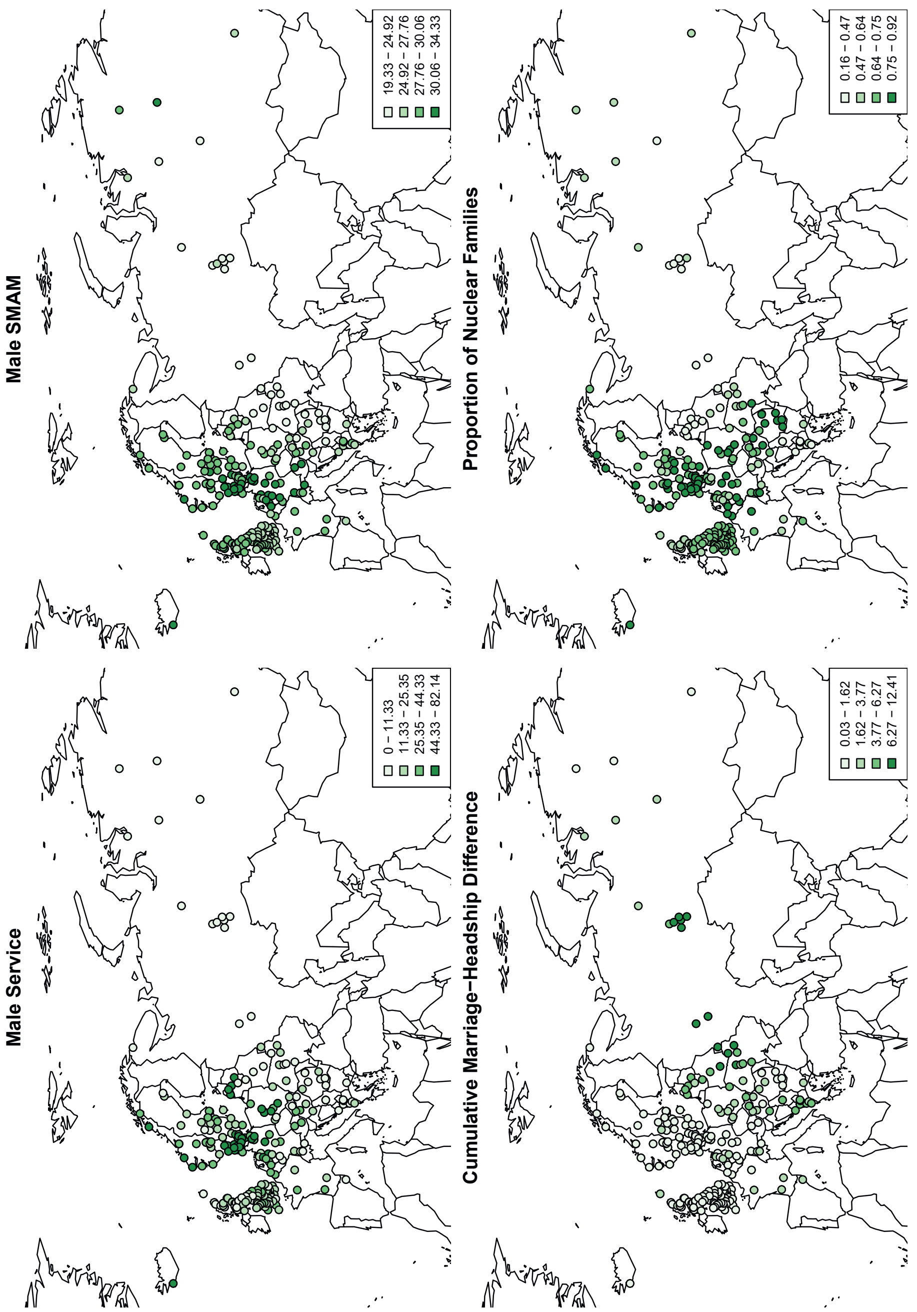


Figure 2
Descriptive mapping of the four household formation markers (Jenks natural breaks)

Table 1

Global correlation matrix between household formation markers

Variables	Male service	Male SMAM	CMHD	Proportion nuclear households
Male service	1.000	0.545**	−0.434**	0.404**
Male SMAM	0.545**	1.000	−0.567**	0.197**
CMHD	−0.434**	−0.567**	1.000	−0.580**
Proportion nuclear households	0.404**	0.197**	−0.580**	1.000

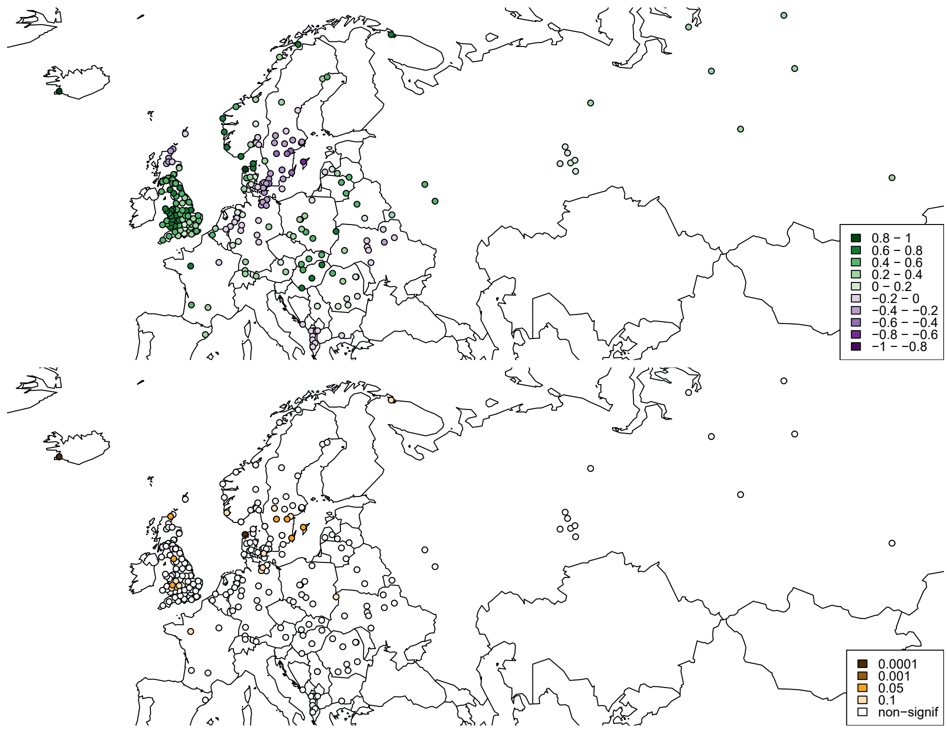
Note: Correlation is significant at the 0.01 level (2-tailed).
Source: Mosaic/NAPP data.

Table 2

Comparison of the global and local correlations between household formation markers

Relationships	Global *r*	Geographically weighted *r*						
		Min.	1st Q	Median	Mean	3rd Q	Max.	N sign. (%)
SMAM-Service	0.545	−0.759	−0.074	0.265	0.215	0.489	0.893	24 (9.38)
Service-CMHD	−0.434	−0.957	−0.338	−0.069	−0.093	0.131	0.709	23 (8.98)
Service-Nuclearity	0.404	−0.918	−0.372	0.004	0.047	0.568	0.973	21 (8.20)
SMAM-CMHD	−0.567	−0.897	−0.423	−0.174	−0.102	0.202	0.704	20 (7.81)
SMAM-Nuclearity	0.197	−0.816	−0.532	−0.206	−0.192	0.110	0.822	23 (8.98)
CMHD-Nuclearity	−0.580	−0.955	−0.735	−0.364	−0.377	−0.133	0.747	23 (8.98)

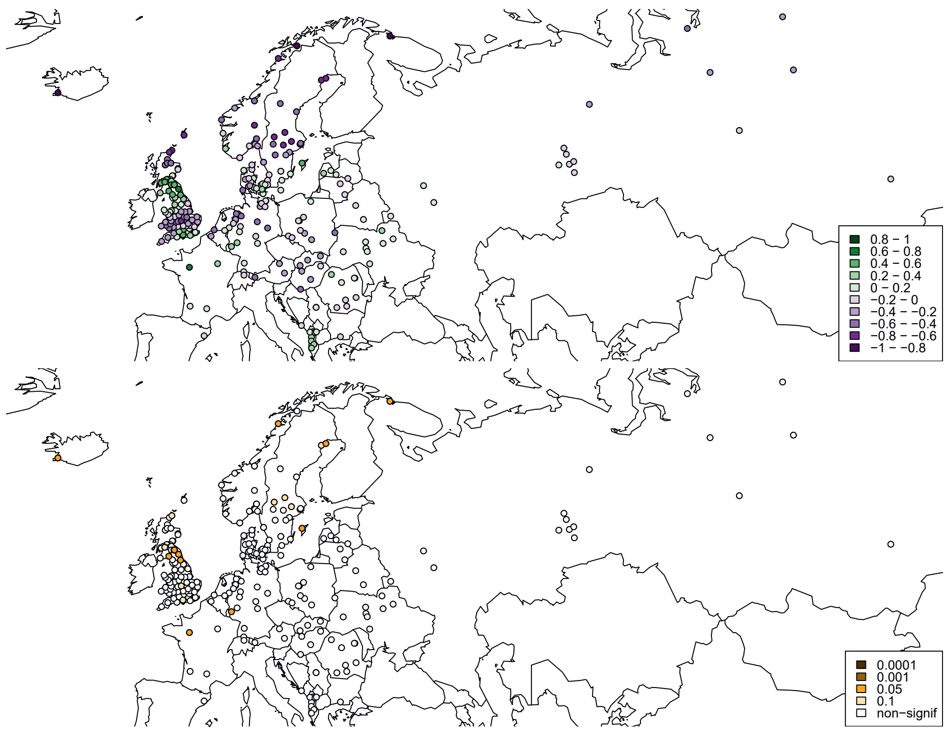
Source: Mosaic/NAPP data. For local correlations: local N=24; *tricube* kernel. For SMAM and Service: male variables.



Source: Mosaic/NAPP data.

Figure 3

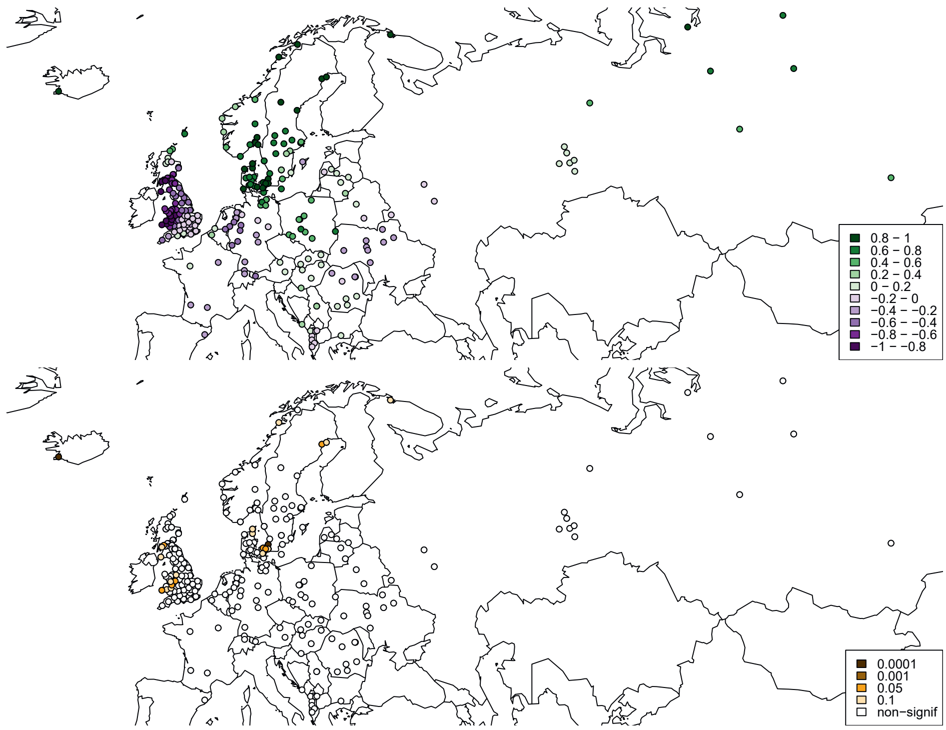
GW Spearman's rho between male SMAM and premarital service (local N=24; tricube kernel)



Source: Mosaic/NAPP data.

Figure 4

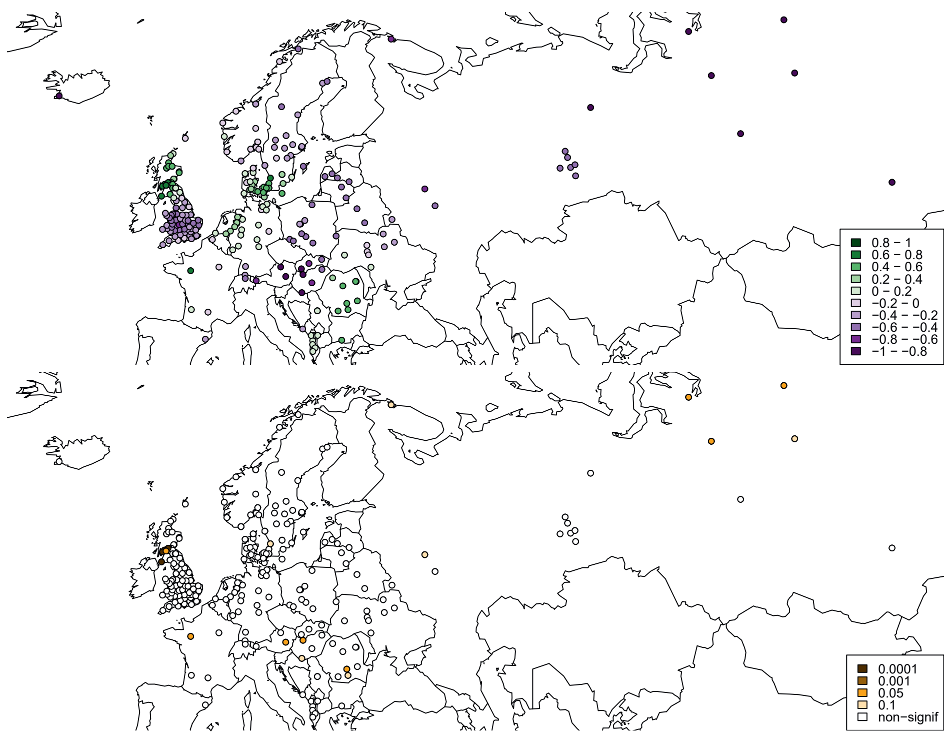
GW Spearman's rho between male service and the CMHD (local N=24; tricube kernel)



Source: Mosaic/NAPP data.

Figure 5

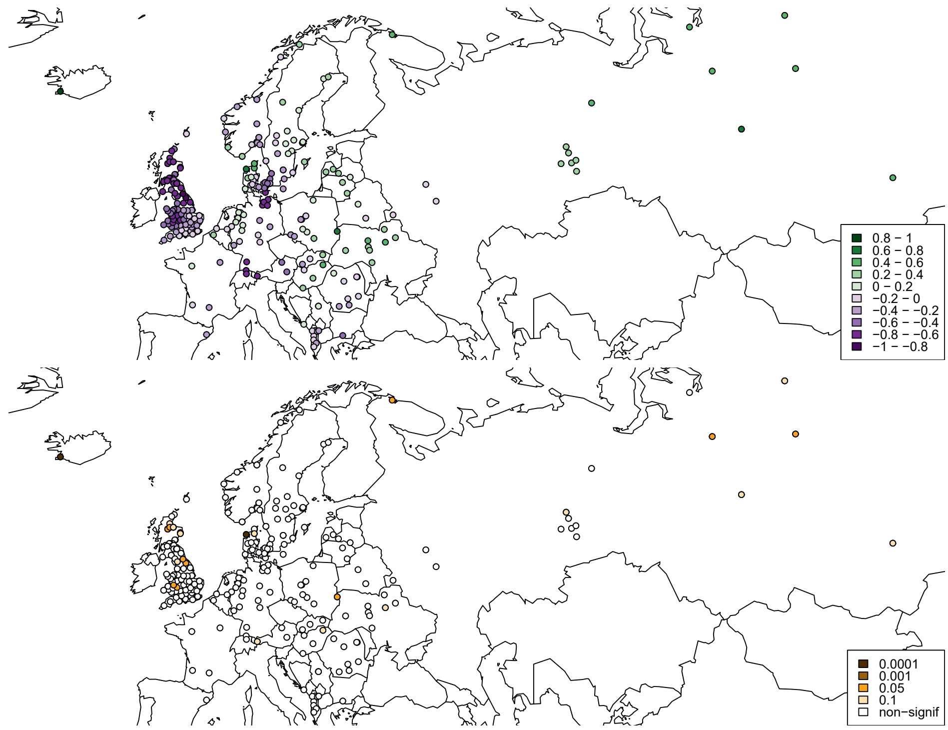
GW Spearman's rho between male service and household nuclearity (local N=24; tricube kernel)



Source: Mosaic/NAPP data.

Figure 6

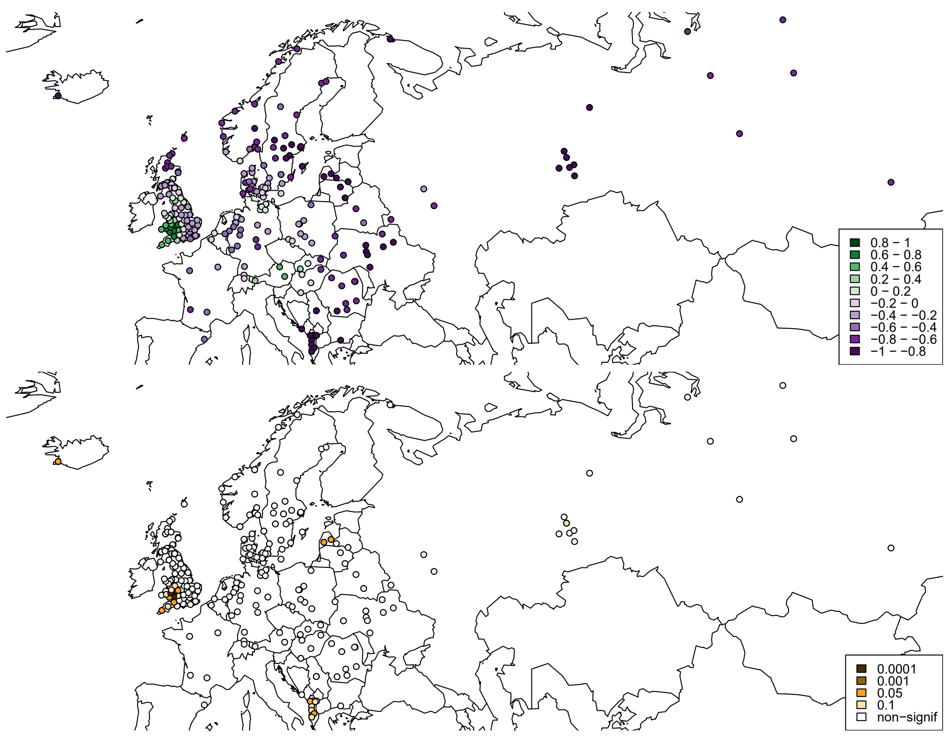
GW Spearman's rho between male SMAM and the CMHD (local N=24; tricube kernel)



Source: Mosaic/NAPP data.

Figure 7

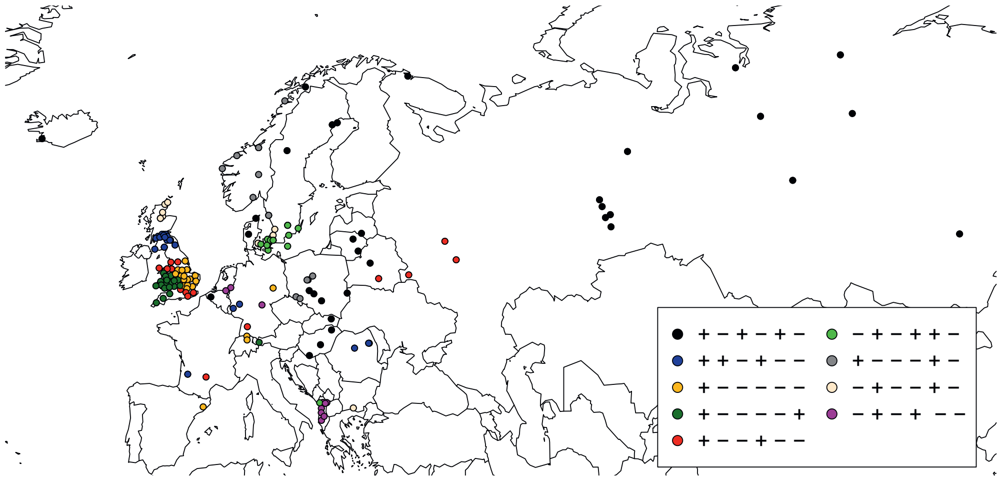
GW Spearman's rho between male SMAM and household nuclearity (local N=24; tricube kernel)



Source: Mosaic/NAPP data.

Figure 8

GW Spearman's rho between the CMHD and household nuclearity (local N=24; tricube kernel)



Note: for each of the labelled pattern, the signs of the coefficients (+/-) refer respectively to correlations between: male SMAM vs service; male SMAM vs CMHD; male SMAM vs nuclearity; male service vs CMHD; male service vs nuclearity; CMHD vs nuclearity.
Only combinations featured in at least ten regions are considered.
Source: Mosaic/NAPP data.

Figure 9

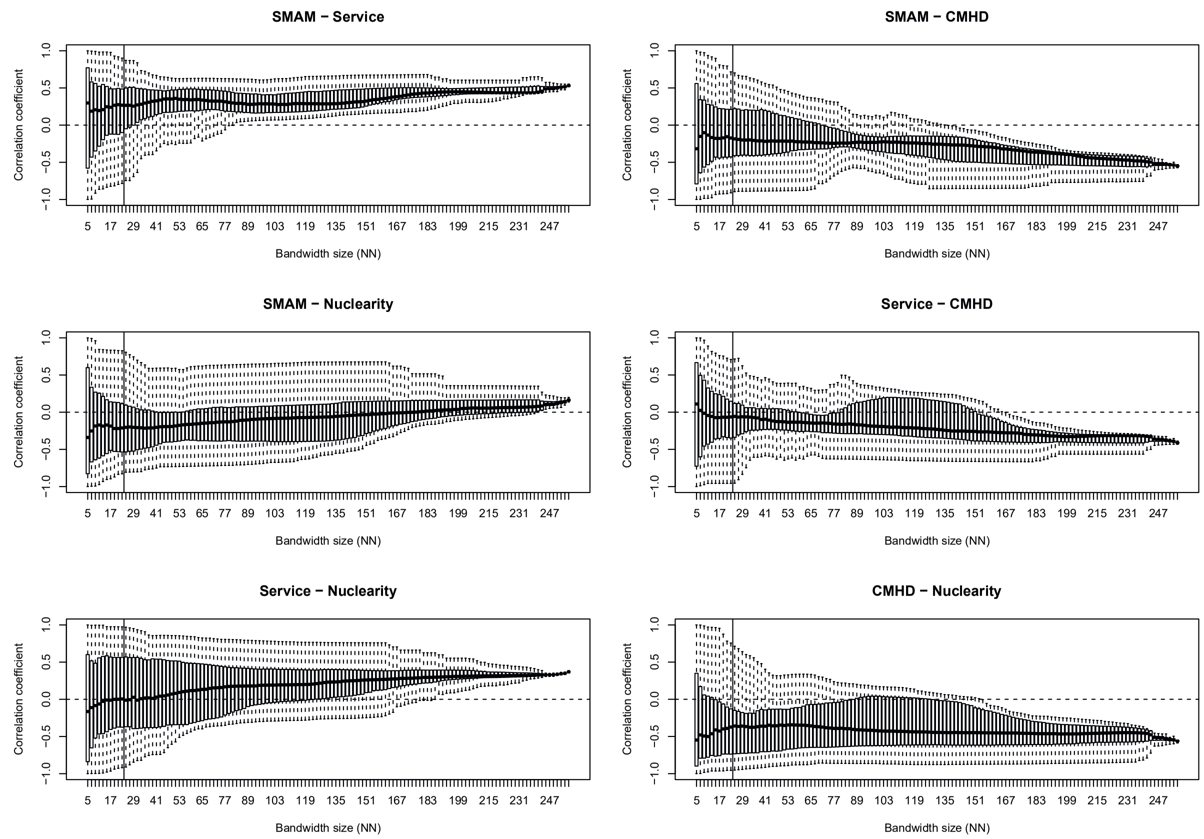
Most frequent (regions ≥ 10) combinations of the relationships between the four household formation markers (local N=24; tricube kernel)

Table 3

Differences between alternative and the baseline local (GW) correlations

Relationships	Female variables, baseline dataset			Male variables, without GB		
	Abs. median difference in GW *p*	No. of correlations with signs unchanged (%)	No. of significant correlations (%)	Abs. median difference in GW *p*	No. of correlations with signs uchanged (%)	No. of significant correlations (%)
SMAM - Service	0.237	198 (77.3)	26 (10.2)	0.050	247 (96.5)	12 (7.1)
Service - CMHD	0.119	218 (85.2)	22 (8.6)	0.075	233 (91.0)	12 (7.1)
Service - Nuclearity	0.130	223 (87.1)	27 (10.5)	0.047	244 (95.3)	16 (9.5)
SMAM - CMHD	0.152	220 (85.9)	25 (9.8)	0.059	240 (93.8)	12 (7.1)
SMAM - Nuclearity	0.165	211 (82.4)	23 (9.0)	0.046	247 (96.5)	17 (10.1)
CMHD - Nuclearity	0.013	248 (96.9)	23 (9.0)	0.068	254 (99.2)	15 (8.9)

Note: for "Female variables, baseline dataset" - local N=27; *tricube* kernel (256 regions included); for "Male variables, without GB" - local N=24; *tricube* kernel (170 regions included).
GW *p* = geographically weighted Spearman's correlation coefficient.
Differences are measured in relation to the baseline correlations as in the main analysis (256 regions; male variables for SMAM and Service).
Source: Mosaic/NAPP data.



Source: Mosaic/NAPP data (baseline correlations).

Figure 10

Changes in the distribution of correlation coefficients between the four variables as a function of number of nearest neighbours used as threshold for the adaptive bandwidth

APPENDIX

Table A

Combinations of the relationships between the four household formation markers

Relationships						N (regions)
SMAM - Service	SMAM - CMHD	SMAM - Nuclearity	Service - CMHD	Service - Nuclearity	CMHD - Nuclearity	
+	–	+	–	+	–	33
+	+	–	+	–	–	26
+	–	–	–	–	–	21
+	–	–	–	–	+	19
+	–	–	+	–	–	17
–	+	–	+	+	–	15
+	–	–	–	+	–	14
–	+	–	–	+	–	10
–	+	–	+	–	–	10
Altogether						165

Note: Only combinations present in at least ten regions are displayed.
Source: Mosaic/NAPP data (local N=24; *tricube* kernel).

Table B

Comparison of the global and local correlations between household formation markers; alternative GWrho (female marriage and service variables)

Relationships	Global *r*	GW correlation *r*					
		Min.	1st Q	Median	3rd Q	Max.	N sign. (%)
SMAM - Service	0.539	−0.778	−0.241	0.181	0.497	0.896	26 (10.2)
Service - CMHD	−0.545	−0.959	−0.385	−0.122	0.090	0.651	22 (8.6)
Service - Nuclearity	0.407	−0.794	−0.249	0.051	0.626	0.979	27 (10.5)
SMAM - CMHD	−0.593	−0.883	−0.447	−0.124	0.146	0.605	25 (9.8)
SMAM - Nuclearity	0.192	−0.871	−0.612	−0.228	0.028	0.784	23 (9.0)
CMHD - Nuclearity	−0.580	−0.948	−0.730	−0.365	−0.163	0.681	23 (9.0)

Source: Mosaic/NAPP data. For local correlations: local N=27; *tricube* kernel.

Table C

Comparison of the global and local correlations between household formation markers; alternative GWrho (male marriage and service variables; GB removed; N=171 regions)

Relationships	Global *r*	GW correlation *r*					
		Min.	1st	Median	3rd	Max.	No. sign. (%)
SMAM - Service	0.578	−0.788	−0.144	0.113	0.304	0.909	12 (7.1)
Service - CMHD	−0.497	−0.834	−0.268	−0.049	0.165	0.649	12 (7.1)
Service - Nuclearity	0.423	−0.678	−0.219	0.259	0.664	0.931	16 (9.5)
SMAM - CMHD	−0.619	−0.921	−0.460	−0.155	0.130	0.687	12 (7.1)
SMAM - Nuclearity	0.275	−0.766	−0.334	−0.122	0.197	0.718	17 (10.1)
CMHD - Nuclearity	−0.676	−0.929	−0.841	−0.569	−0.348	0.421	15 (8.9)

Source: Mosaic/NAPP data. For local correlations: local N=24; *tricube* kernel.

SUPPLEMENTAL MATERIAL

A1. BUILDING THE RURAL DATASET

The urban and rural distinction between population in our combined NAPP/Mosaic database is not for all regions as easily available as for others. In Mosaic, regions are by default divided into rural and urban ones, so it is quite straightforward to select only the rural ones. In NAPP data the definitions of urban residence vary between countries (see Table A1) and we have used those pertaining to particular dataset without trying to harmonise them. Only Denmark in 1787 and Sweden in 1880 could be used without any further adaptation (the few people with unknown status have been excluded). Iceland in 1703 has no information about urban or rural residence, but since Iceland at that time had no city, we treat the whole island as being rural. The file of Norway 1801 has no information about urban or rural residence either, but we used the variable "TownNO" and coded all people not living in one of the coded towns as being rural. The published data of the 1801 census[1] yields an urban population of 10.0 percent and therefore our solution seems very close to this figure (*Table 1A*). The lower proportion of the urban population in our selection can be attributed to some smaller cities included in rural municipalities (*ladestad*), but not captured in the list of cities in the variable "TownNO" (https://www.nappdata.org/napp-action/variables/TOWNNO#codes_section)[2].

England, Wales, and the Islands in the British Seas in 1851 have a high proportion of the population living in places of unknown urban/rural status and only few or no people in urban places (*see Table 1A*). We have therefore decided to treat these people as urban and excluded them from the rural analysis. In Scotland only the 1881 census has the information about urban/rural status and therefore this census has been used instead of the 1851 census as for the other British regions[3]. We have used the 1851 census for England and Wales because this is the oldest available census for these countries with an early industrialisation despite creating a time lag to Scotland of 30 years.

Table A1

Percentages of the urban and rural population in the NAPP data

NAPP dataset	Rural	Urban	Unknown	Definition of urban
Iceland 1703	100.0	–	–	n.a.
Denmark 1787	79.2	20.6	0.2	Copenhagen and royal boroughs (Købstad)
Norway 1801	90.4	9.6	–	Coded as one of the towns in variable TownNO
Sweden 1880	84.3	15.7	–	Parish as classified as part of a city
England 1851	72.6	8.7	18.8	Parish with a population density of more than 75 persons per acre
Wales 1851	84.4	–	15.6	Parish with a population density of more than 75 persons per acre
Islands 1851	–	–	100.0	Parish with a population density of more than 75 persons per acre
Scotland 1881	89.4	10.6	–	Parish with a population density of more than 75 persons per acre

Source: Minnesota Population Center. *North Atlantic Population Project: Complete Count Microdata. Version* 2.3 [Machine-readable database]. Minneapolis: Minnesota Population Center, 2016.

1 Statistisk Sentralbyrå, ed., Folketeljinga 1801 ny bearbeiding (Norges Offisielle Statistikk B 134). Oslo 1980, pp. 56f.

2 A second possibility would be to use the 3rd digit of the variable "MunicNO" which identifies urban municipalities, but the urban population would then be 11.4 percent and therefore less similar to the published data.

3 There are no unknown cases in this Scottish census.

Because our definition of "rural" is predetermined by the structure of the data and not based on the occupational structure of the resident population, it makes it impossible to ascertain if all individuals in rural places such defined would have been engaged in the primary sector (although in the majority of Mosaic data this is more than likely to be the case). However, except for some few small urban places in Norway which are not distinguished from rural ones in the data, and some urban places in England, Wales and Scotland with particularly low population density, none of the rural regions derived according to the rules above would contain people living in towns or other urban centres.

A2. SYSTEMATIC BIAS IN SMALLER REGIONAL GROUPINGS

Can 58 regions from our database with smaller number of households (N<1,000) be systematically biased with regards to the values of our focal variables (household formation markers)?

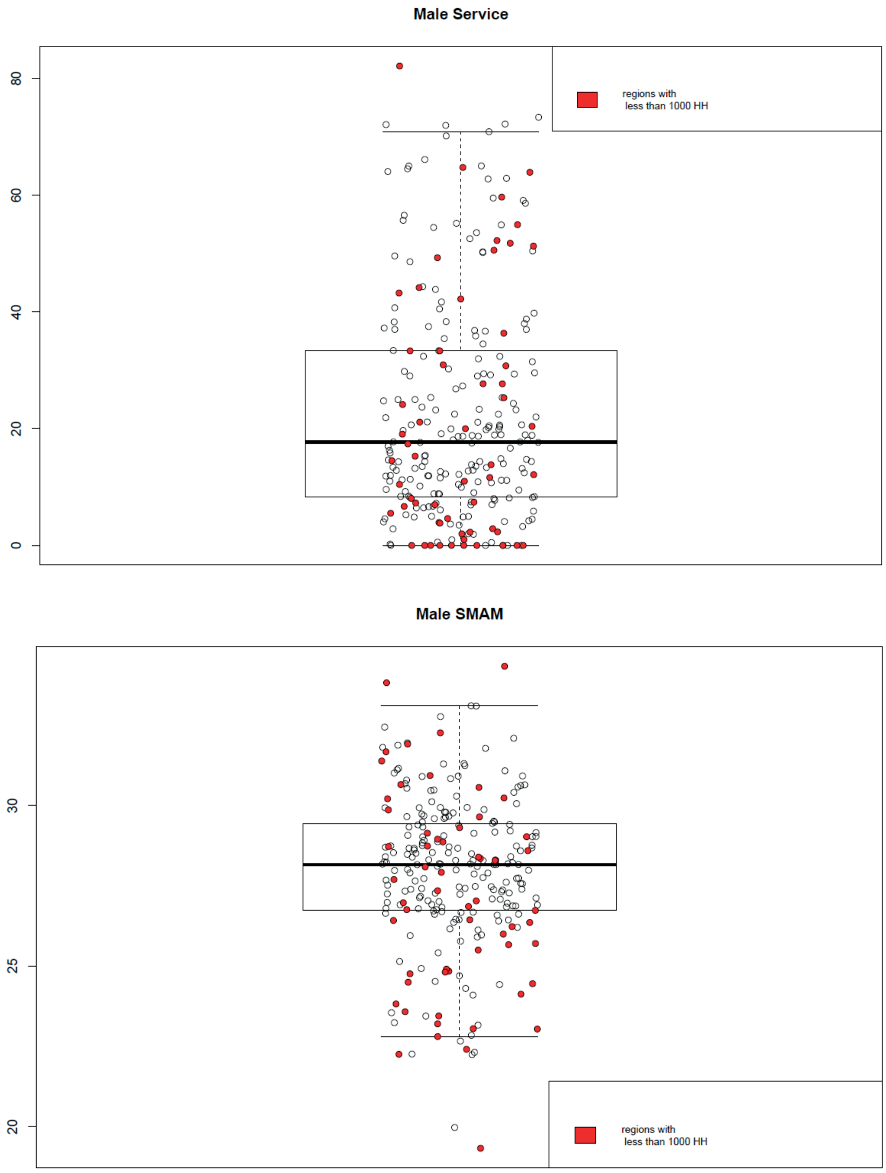
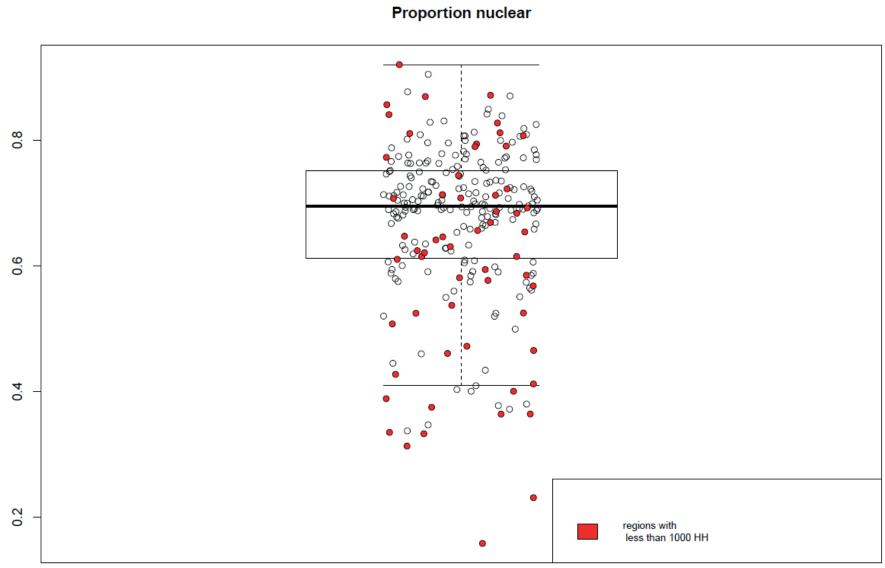


Figure A2_1_4

Dispersion of the values of the four household formation markers employed in the analysis (NAPP and Mosaic data combined)



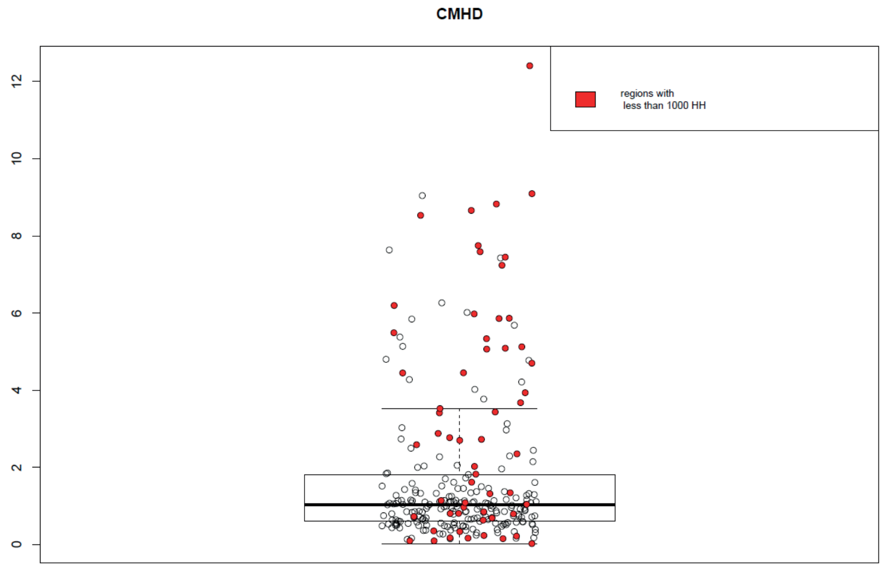


Figure A2_1_4 (continued)

Dispersion of the values of the four household formation markers employed in the analysis (NAPP and Mosaic data combined)

Figures A2_1_4 show that in general our N<1000 hh regions are distributed across all parts of the respective data distribution, and there is no clear evidence that most outlying cases are clustered among these particular regions. The fact that for SMAM and nuclear families some smaller regions are leaning towards minimal values does not seem to be a function of their size, but rather of their actual familial specificity, since these are mostly Eastern and Southeastern European populations with low age at marriage and low proportion of simple family households.

A3. ASSESSING "DEMOGRAPHICALLY TRADITIONAL" FEATURES OF THE REGIONAL POPULATIONS

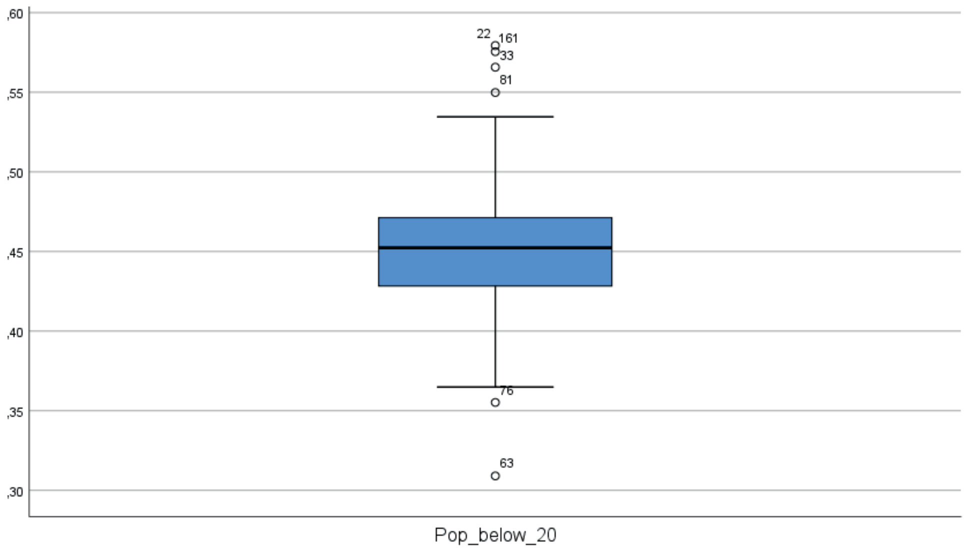


Figure A3_1

Distribution of 256 regional populations by their respective shares of population below 20 years of age

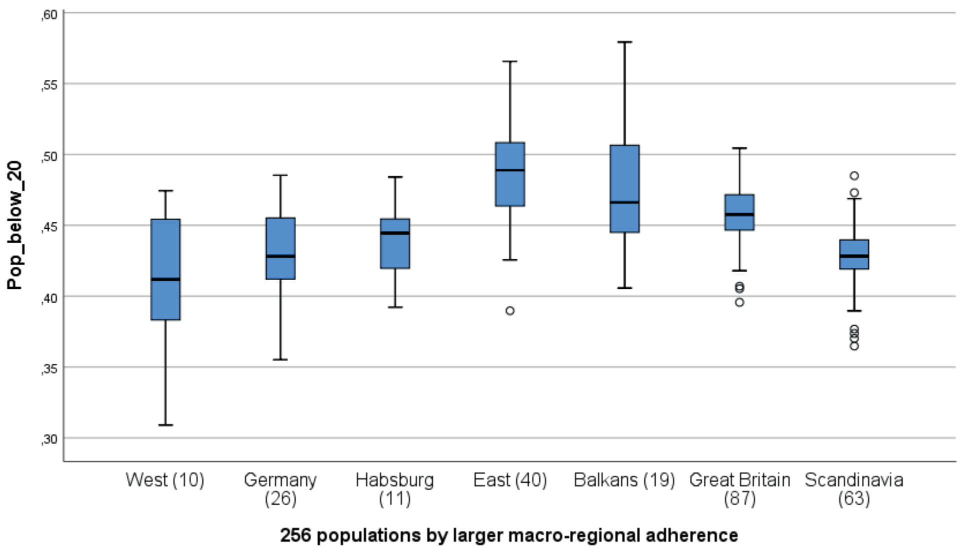


Figure A3_2

Distribution of 256 regional populations by their respective of population below 20 years of age, by broad geographic groupings

Notes: Seven bigger territorial groupings followed major institutional and socioeconomic distinctions across historic Europe. "Great Britain": England, Wales, and Scotland; "Scandinavia": Danish, Swedish, and Norwegian data, as well as Iceland; "Germany": German-dominated areas other than the Habsburg territories; "West": areas west and south-west of Germany; "Habsburg": Austrian, Hungarian, Croatian, as well as Slovakian data; "East": east-central and eastern Europe, including the former Polish-Lithuanian Commonwealth and Russia; "Balkans": areas south and/or east of Croatia and Hungary.

In order to assess whether particular populations as captured in the censuses stored in our database precede the onset of a monotonic fertility decline, our regional data were matched (by means of a spatial joint) to the corresponding province-level estimates of the onset of the fertility decline derived from the European Princeton Fertility Project's capstone volume (Coale and Watkins 1986; for territories not covered by the Princeton data - such as Turkey, Albania, or Siberia - we used indicators derived from other existing literature, e.g. Falkingham and Gjonca 2001 or Coale et al. 1979, on Asiatic Russia). A dummy variable for each of the regional populations in our dataset was then created indicating whether given population at its specific census time belonged to a province which already experienced/or not the onset of monotonic fertility decline (following Coale' definition of the fall of the fertility as measured by his Ig below 90 percent of a previous plateau).

We were able to approximate that 240 out of 256 rural regions used in the analysis could be safely considered as including populations which did not yet experience fertility declines. Except for three out of four regions of France, which has been the forerunner of the European fertility decline, the 16 regions not conforming with this characteristic include populations widely scattered amongst the dominant "pre-transitional" populations[4], and as such cannot change the overall picture here suggested.

A4. DEFINING THE SERVICE HOUSEHOLD FORMATION MARKER

Although Hajnal's household formation rules (1982, 452) were presented as applying to both sexes, presenting the results of the geographically weighted analyses including service variables for both men and women would be burdensome. Since we are dealing with six pairs of bivariate correlations (each assessed locally), redoing the analysis with the service variable for both men and women would double the number of local correlations to be computed (from 1536 to 3072). Although technically this is not a problem, presenting results obtained in this way would be very difficult, as the number of maps would also double (from six pairs to twelve pairs).

Furthermore, it would also be counterproductive, since our measures of male and female premarital female service are strongly (and significantly) correlated with each other (Spearman's rho= 0.846; Fig. A4_1 below). The regional incidence of premarital male service predicts 70 percent of the variation in the female variable, and both variables nearly equally predict the overall percentage of the servant population in a region (R2=0.70). Given this covariation patterns it might be argued that male and female variables could be used interchangeably.

4 Southwest France 1831, Sweden/Gotland 1880, France/Northeast 1846, Upper Austria 1910, France/Northwest 1846, Austria/Tyrol 1910, Sweden/Uppsala region 1880, Catalonia rural 1880-1890, Sweden/Västmanland 1880, Sweden/Stockholm region 1880, Hungary/Southern Transdanubia 1869, Austria/Styria 1910, Bulgaria/Rhodope region 1877-1947, Scotland/Roxburghshire 1881, Scotland/Berwickshire 1881, Scotland/Selkirk 1881.

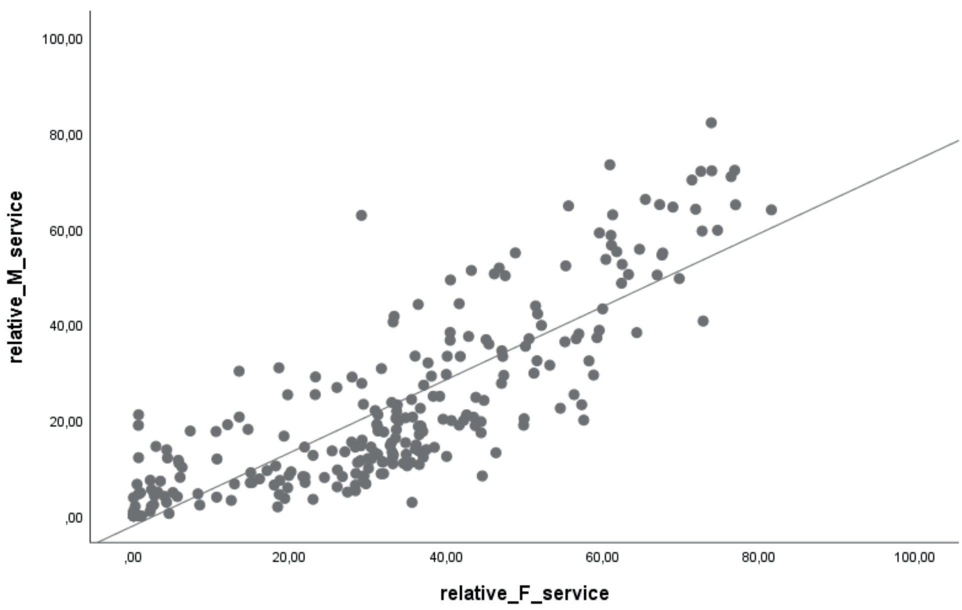


Figure A4_1

Correlation between male and female premarital service variables across combined NAPP/Mosaic data (256 rural regions)

However, since Hajnal was interested in agricultural related farm service in rural areas (1982, 470, 473), male population seems to better reflect this phenomenon than the female one represented in our database mostly by domestic labourers. To focus on female service would be questionable since Hajnal defined his servants as those individuals who "were in the main not servants in the now customary meaning of the term, that is, people ministering to the personal comforts of the more prosperous section of the population", but who "participated in the productive tasks-mainly in farming or craft activities-of the households in which they lived. (...) Most servants were not primarily engaged in domestic tasks, but were part of the work force of their master's farm or craft enterprise" (1982, 470, 473).

Meanwhile, in the overwhelming majority of the Mosaic listings, a strict distinction between farm servants and domestic servants is impossible, as most men and women are generally called "servants". The situation with the majority of NAPP censuses is similar: in the censuses of Iceland (1703), Denmark (1787) and Norway (1801) the absolute majority of men and women categorized as "servants" by the relationship to household head code provided by NAPP, either have no occupation title provided, or were just called "servants". The situation in the British censuses from the NAPP used in this paper was somewhat different, as in addition to standard coding of servants based on the relationship to head of household variable provided by the NAPP[5], it was possible to partly rely on the OCCHISCO occupational coding scheme, and hence to provide a more nuanced description of (some) servants. Had we done this, we would find that for England and Wales in 1851 roughly 90 percent of female servants were categorized as "house servants" (an equivalent of the "domestic service"), and 91 percent of male servants were farm servants without specialization (the situation in the Scottish census of 1881 was generally similar). In other words, these female servants would barely fit Hajnal's 1982 definition of service.

5 See: https://www.nappdata.org/napp-action/variables/SERVANTS#description_section, and https://www.nappdata.org/napp-action/variables/RELATE#codes_section

Thus, in all likelihood, the male service variable serves better in this regard. In occupational terms - available, however, for only a part of our datafiles, the absolute majority of these men were just called "servants", or were denoted as farm workers and agricultural labourers (*see Table A4_1* below), so this variable is by all means closer to what Hajnal had in mind.

Table A4_1

Occupational coding of servants in the NAPP data (relate codes 1211 thru 1218/England and Wales codes 5200 and 5210): only rural areas, men

	N	Servants	House servants	Peronal servants	Military	Farm workers	No occupation/ unknown
Iceland 1703	4,209	93.0	0.0	0.1	0.0	0.0	0.2
Denmark 1787	67,224	0.0	4.4	0.1	10.7	1.8	75.8
Norway 1801	37,659	5.3	0.0	0.0	11.8	0.1	80.7
Sweden 1880	102,939	0.5	0.0	0.0	0.0	99.5	0.0
England 1851	25,435	0.3	15.7	0.1	0.1	43.2	4.6
Wales and Islands 1851	35,792	0.3	10.9	0.1	0.0	59.5	5.5
Scotland 1881	45,488	0.5	7.2	0.0	0.0	62.9	0.9

Still, the male service variable may be problematic for other reasons. According to the general wisdom (esp. Kussmaul 1981), male service was already in decline, if not extinct, in Britain of 1851 (Kussmaul's argument), so by focusing on males only we may still be far away from properly capturing the institution that Hajnal had in mind[6]. The extent of the demise of agricultural service during the nineteenth-century remain, however, a matter of dispute. Kussmaul's contention about the near extinction of the male farm-service before 1851 has been challenged by a more recent literature which suggested the continued centrality of service to the rural economy in various parts of England, Wales and Scotland well into the late nineteenth-century (see review in Verdon 2002, 78-83). This literature is, however, still inconclusive.

Also, while the superiority of the male service variable is clear, possibilities of using the female service variable cannot be entirely rejected. Even if the majority of female servants were "domestic servants", it could be argued that since we are dealing with the rural datafile, they were unlikely to represent primarily the people "ministering to the personal comforts of the more prosperous section of the population", but rather people participating in various "productivity" or "consumption" tasks that Hajnal used to define servants. This view is supported by increasing recognition that differences between female domestic and agricultural servants were blurred and in many instances may be difficult if not impossible to establish, as some individuals probably came into both categories, with many female servants in husbandry also undertaking domestic work in the house, and domestic servants also performing "productive" work on the farm and in the dairy (Higgs 1995, 707; Holland 2017, 190 ff.; Gritt 2002, 87; esp. Verdon 2002, 81-83; also Snell 1985, 283) .[7]

6 Unlike the male service, female service was much less responsive to changes associated with the transition period to capitalist/industrial economy, and the Demand for female domestic servants remained high throughout the nineteenth-century (Cooper 2004, 287).

7 According to Holland (2017, 190), "there was a considerable overlap in the work undertaken by farm servants and domestic servants, especially on small farms, which mean that farm-work was not always distinguished from domestic duties" (190 ff). According to Wall (2004, 25), "there was a nebulous border between "productive" work and domestic tasks". Verdon (2002, 83) argued that "an indiscriminate mixing of tasks in the house and on the farm was the norm for servants on many farms into the twentieth century [England]". John Rickman, an English governmental statistician, expressed this problem succinctly when he wrote in 1831: 'In some instances a doubt is expressed, whether a Female hired by an Occupier of land, and resident in his family is to be deemed a Household Servant, or an Agricultural Servant. Certainly she is also a Household Servant, and to be reckoned as such" (q. in Gritt 2002, 87).

A5. ILLUSTRATION OF THE WEIGHTING MATRICES USED IN THE ANALYSIS

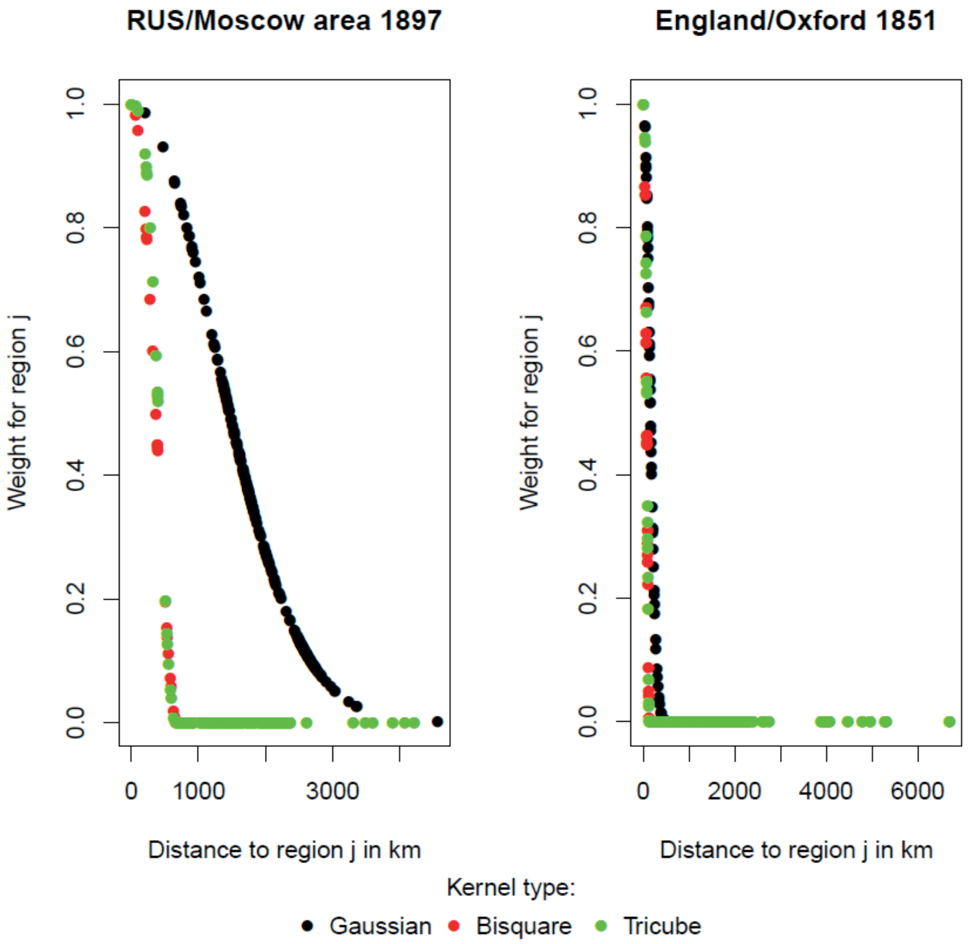


Figure A5_1

The distribution of weights as a function of distance in local GWrho models for Moscow area and Oxford county by type of the kernel density functions

Literature cited in sections A1–A5

Coale, A. J. and Watkins, S.C. (Eds.) (1986). *The Decline of Fertility in Europe*. Princeton: Princeton University Press.

Coale, A. J., Anderson, B. A., and Härm, E. (1979). *Human Fertility in Russia since the Nineteenth Century*. Princeton, NJ: Princeton University.

Cooper S. M. (2004). From Family Member to Employee: Aspects of Continuity and Discontinuity in English Domestic Service, 1600–2000. In Fauve-Chamoux, A. (ed.): *Domestic service and the formation of European identity. Understanding the globalization of domestic work, 16th-21st centuries*. Bern: Peter Lang, . 277–296.

Falkingham, J. and Gjonça, A. (2001). Fertility transition in Communist Albania, 1950-90. *Population Studies*, 55(3), 309–318.

Gritt, A. J. (2002). The "Survival" of Service in the English Agricultural Labour Force: Lessons from Lancashire, c. 1650–1851, *Agricultural History Review* 50(1), 25–50.

Hajnal, J. (1982). Two kinds of preindustrial household formation system, *Population and Development Review* 8, 449–494.

Higgs, E. (1995). Occupational censuses and the agricultural workforce in Victorian England and Wales1. *The Economic History Review*, 48, 700–716.

Holland, S. (2017). Farm Service and Hiring Practices in Mid-Nineteenth-Century England: The Doncaster Region in the West Riding of Yorkshire. In Whittle, J. (ed.): *Servants in Rural Europe: 1400–1900.* Boydell & Brewer, 183–202.

Kussmaul, A. (1981). *Servants in Husbandry in Early Modern England*. Cambridge: Cambridge University Press.

Snell, K. D. M. (1985). *Annals of the Labouring* Poor*: Social Change and Agrarian England, 1660–1900*. Cambridge.

Verdon, N. (2002). *Rural Women Workers in Nineteenth-Century England: Gender, Work and Wages*. Rochester, N. Y.: The Boydell Press.

A6. REGIONAL DATASETS BY SELECTED FEATURES, INCLUDING CORRELATION COEFFICIENTS OF ALL LOCAL CORRELATIONS

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
AL/Berat 1918	Mosaic	after 1850	Balkans	7424	1275	−0.19	0.09	−0.18	0.25	−0.12	−0.94
AL/Gora 1918	Mosaic	after 1850	Balkans	11298	2284	−0.08	0.17	−0.21	0.20	−0.04	−0.93
AL/Kruja 1918	Mosaic	after 1850	Balkans	4276	591	−0.17	0.06	−0.15	0.22	−0.06	−0.94
AL/Puka 1918	Mosaic	after 1850	Balkans	5008	786	−0.13	0.08	−0.15	0.21	−0.03	−0.93
AL/Shkodra 1918	Mosaic	after 1850	Balkans	12340	1968	−0.16	0.01	−0.11	0.18	0.02	−0.93
AL/Tirana North 1918	Mosaic	after 1850	Balkans	14529	2703	−0.17	0.07	−0.17	0.23	−0.08	−0.94
AL/Tirana South 1918	Mosaic	after 1850	Balkans	12206	2507	−0.16	0.13	−0.21	0.24	−0.11	−0.94
AL/Zhuri 1918	Mosaic	after 1850	Balkans	15565	2820	−0.10	0.14	−0.19	0.21	−0.05	−0.93
AT/Styria 1910	Mosaic	after 1850	Habsburg	6693	1256	0.25	−0.90	−0.43	−0.27	0.00	0.44
AT/Tyrol 1910	Mosaic	after 1850	Habsburg	6514	1266	0.35	−0.63	−0.75	−0.42	−0.43	0.23
AT/Upper Austria 1910	Mosaic	after 1850	Habsburg	1675	379	0.20	−0.86	−0.59	−0.27	0.02	0.43
BE/Western Flanders 1814	Mosaic	1800–1850	West	13666	2662	0.44	−0.03	0.25	−0.37	0.36	−0.14
BG/Rhodope region 1877–1947	Mosaic	after 1850	Balkans	6590	1332	−0.11	0.42	−0.42	0.00	0.03	−0.92
CH/Zürich North rural 1671–1685	Mosaic	before 1800	West	6400	1346	0.24	−0.40	−0.67	−0.11	−0.34	−0.10
CH/Zürich South rural 1678–1762	Mosaic	before 1800	West	4588	848	0.23	−0.44	−0.71	−0.18	−0.32	−0.03
DE/Arnsberg 1846	Mosaic	1800–1850	Germany	2111	379	−0.02	0.17	0.10	0.11	−0.54	−0.55
DE/Braunschweig 1846	Mosaic	1800–1850	Germany	2014	539	−0.05	0.05	−0.26	−0.60	−0.08	−0.34
DE/Danzig and Posen 1858	Mosaic	after 1850	Germany	3468	730	0.21	−0.38	−0.25	−0.01	0.56	−0.27

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
DE/Düsseldorf 1846	Mosaic	1800–1850	Germany	2185	370	−0.11	0.27	−0.07	0.05	−0.45	−0.44
DE/Koblenz 1846	Mosaic	1800–1850	Germany	2092	467	0.05	0.10	−0.17	0.21	−0.58	−0.48
DE/Liegnitz 1846	Mosaic	1800–1850	Germany	2607	717	0.19	−0.51	−0.40	−0.07	0.53	−0.04
DE/Mecklenburg–Schwerin Northeast 1819	Mosaic	1800–1850	Germany	9602	2018	−0.28	0.17	−0.67	0.01	0.58	0.11
DE/Mecklenburg–Schwerin Northeast 1867	Mosaic	after 1850	Germany	14107	2518	−0.28	0.17	−0.67	0.01	0.58	0.12
DE/Mecklenburg–Schwerin Northwest 1819	Mosaic	1800–1850	Germany	2061	427	−0.35	0.09	−0.70	−0.11	0.61	0.13
DE/Mecklenburg–Schwerin Northwest 1867	Mosaic	after 1850	Germany	3914	692	−0.35	0.09	−0.71	−0.12	0.61	0.13
DE/Mecklenburg–Schwerin Southeast 1867	Mosaic	after 1850	Germany	3232	524	−0.33	0.04	−0.70	−0.11	0.53	0.08
DE/Mecklenburg–Schwerin Southwest 1819	Mosaic	1800–1850	Germany	10209	2023	−0.36	0.00	−0.71	−0.21	0.55	0.12
DE/Mecklenburg–Schwerin Southwest 1867	Mosaic	after 1850	Germany	16025	2707	−0.36	−0.01	−0.71	−0.22	0.55	0.12
DE/Meppen 1749	Mosaic	before 1800	Germany	12360	2350	0.00	0.24	0.17	−0.56	−0.29	−0.13
DE/Meppen and Cloppenburg 1700	Mosaic	before 1800	Germany	8607	1629	0.01	0.22	0.17	−0.58	−0.26	−0.16
DE/Merseburg 1846	Mosaic	1800–1850	Germany	1939	412	0.20	−0.16	−0.30	−0.57	−0.18	−0.31
DE/Münster 1846	Mosaic	1800–1850	Germany	1278	212	−0.07	0.23	0.03	−0.09	−0.42	−0.42
DE/Rheine–Bevergern 1700	Mosaic	before 1800	Germany	3506	618	−0.04	0.20	0.10	−0.25	−0.37	−0.37
DE/Rheine–Bevergern 1749	Mosaic	before 1800	Germany	5558	1069	−0.04	0.21	0.11	−0.30	−0.35	−0.33

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
DE/Sachsen–Coburg 1846	Mosaic	1800–1850	Germany	2096	442	−0.18	0.08	−0.15	0.12	−0.33	−0.78
DE/Sachsen–Gotha 1846	Mosaic	1800–1850	Germany	2284	497	−0.16	0.17	0.00	−0.03	−0.34	−0.73
DE/Sigmaringen 1861	Mosaic	after 1850	Germany	6541	1345	0.26	−0.39	−0.62	0.01	−0.38	−0.20
DE/Stromberg 1749	Mosaic	before 1800	Germany	16251	2454	−0.02	0.14	0.14	−0.07	−0.45	−0.50
DE/Trier 1846	Mosaic	1800–1850	Germany	1712	374	0.05	0.14	−0.31	0.30	−0.60	−0.45
DE/Vechta 1700	Mosaic	before 1800	Germany	10897	2130	0.00	0.17	0.13	−0.57	−0.19	−0.28
Denmark/Aalborg 1787	NAPP	before 1800	Scandinavia	42674	8879	0.61	0.03	0.43	−0.09	0.88	−0.29
Denmark/Aarhus 1787	NAPP	before 1800	Scandinavia	23703	4654	0.10	0.32	−0.16	0.03	0.74	−0.39
Denmark/Bornholm 1787	NAPP	before 1800	Scandinavia	17674	3856	−0.15	0.41	−0.35	0.19	0.65	−0.15
Denmark/Frederiksborg 1787	NAPP	before 1800	Scandinavia	41842	9037	−0.34	0.52	−0.44	−0.04	0.90	−0.21
Denmark/Hjorring 1787	NAPP	before 1800	Scandinavia	40791	8563	0.67	−0.06	0.43	−0.13	0.89	−0.23
Denmark/Holbak 1787	NAPP	before 1800	Scandinavia	44388	9431	−0.26	0.59	−0.40	0.11	0.91	−0.13
Denmark/Kobenhavn 1787	NAPP	before 1800	Scandinavia	111301	23396	−0.29	0.50	−0.39	0.01	0.93	−0.15
Denmark/Maribo 1787	NAPP	before 1800	Scandinavia	47937	10154	−0.33	0.40	−0.62	0.07	0.68	−0.03
Denmark/Odense 1787	NAPP	before 1800	Scandinavia	59432	11474	−0.13	0.48	−0.29	−0.08	0.79	−0.54
Denmark/Prasto 1787	NAPP	before 1800	Scandinavia	49364	10261	−0.34	0.57	−0.50	0.23	0.85	−0.02
Denmark/Randers 1787	NAPP	before 1800	Scandinavia	45400	9234	0.13	0.34	−0.18	0.06	0.72	−0.33
Denmark/Ribe 1787	NAPP	before 1800	Scandinavia	31248	6212	0.08	0.22	−0.04	−0.32	0.81	−0.70
Denmark/Ringkobing 1787	NAPP	before 1800	Scandinavia	34964	7262	0.57	−0.10	0.35	−0.27	0.74	−0.65
Denmark/Roskilde 1787	NAPP	before 1800	Scandinavia	21195	4326	−0.33	0.56	−0.44	0.07	0.92	−0.13
Denmark/Skanderborg 1787	NAPP	before 1800	Scandinavia	25094	5217	0.03	0.27	−0.15	−0.08	0.76	−0.52

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
Denmark/Soro 1787	NAPP	before 1800	Scandinavia	39403	8020	−0.33	0.58	−0.49	0.24	0.85	−0.02
Denmark/Svendborg 1787	NAPP	before 1800	Scandinavia	51049	9933	−0.19	0.56	−0.39	0.11	0.81	−0.28
Denmark/Thisted 1787	NAPP	before 1800	Scandinavia	27438	5701	0.89	0.03	0.70	0.02	0.80	−0.28
Denmark/Tonder 1787	NAPP	before 1800	Scandinavia	8887	1866	0.30	0.05	0.17	−0.40	0.80	−0.72
Denmark/Vejle 1787	NAPP	before 1800	Scandinavia	35611	6886	0.02	0.28	−0.12	−0.23	0.78	−0.65
Denmark/Viborg 1787	NAPP	before 1800	Scandinavia	39228	7881	0.37	0.06	0.13	−0.17	0.71	−0.57
England/Bedfordshire 1851	NAPP	after 1850	Great Britain	10821	2231	0.27	−0.33	−0.21	−0.36	−0.19	−0.28
England/Berkshire 1851	NAPP	after 1850	Great Britain	16754	3597	0.13	−0.38	−0.15	0.53	−0.01	−0.08
England/Buckinghamshire 1851	NAPP	after 1850	Great Britain	13349	3288	0.22	−0.30	−0.29	−0.12	−0.17	−0.22
England/Cambridgeshire 1851	NAPP	after 1850	Great Britain	14213	2990	0.39	−0.48	−0.11	−0.33	−0.14	−0.25
England/Cheshire 1851	NAPP	after 1850	Great Britain	38451	7626	0.59	−0.45	−0.44	0.13	−0.79	−0.19
England/Cornwall 1851	NAPP	after 1850	Great Britain	33702	6893	0.41	−0.20	−0.28	−0.26	−0.47	0.46
England/Cumberland 1851	NAPP	after 1850	Great Britain	14682	3036	0.75	0.04	−0.70	0.47	−0.58	0.14
England/Derbyshire 1851	NAPP	after 1850	Great Britain	15151	3153	0.42	−0.32	−0.30	−0.28	−0.48	−0.34
England/Devonshire 1851	NAPP	after 1850	Great Britain	48977	10025	0.48	−0.24	−0.29	−0.27	−0.35	0.55
England/Dorset 1851	NAPP	after 1850	Great Britain	15449	3418	0.37	−0.11	−0.06	0.11	0.02	0.23
England/Durham 1851	NAPP	after 1850	Great Britain	35062	7541	0.41	−0.17	−0.82	0.50	−0.44	0.18
England/Essex 1851	NAPP	after 1850	Great Britain	27750	5750	0.39	−0.36	−0.28	−0.14	−0.19	−0.30
England/Gloucestershire 1851	NAPP	after 1850	Great Britain	27169	5747	0.56	−0.56	−0.40	−0.11	−0.18	0.51
England/Hampshire 1851	NAPP	after 1850	Great Britain	25832	5451	0.16	−0.17	−0.01	0.56	0.04	−0.31
England/Herefordshire 1851	NAPP	after 1850	Great Britain	8651	1911	0.78	−0.58	−0.77	−0.48	−0.73	0.75

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
England/Hertfordshire 1851	NAPP	after 1850	Great Britain	16136	3297	0.36	−0.31	−0.26	−0.17	−0.24	−0.31
England/Huntingdonshire 1851	NAPP	after 1850	Great Britain	5123	1110	0.35	−0.44	−0.11	−0.36	−0.13	−0.27
England/Kent (extra London) 1851	NAPP	after 1850	Great Britain	41894	8553	0.44	−0.28	−0.20	0.01	−0.16	−0.36
England/Lancashire 1851	NAPP	after 1850	Great Britain	133234	26052	0.52	−0.41	−0.59	0.25	−0.61	−0.19
England/Leicestershire 1851	NAPP	after 1850	Great Britain	12548	2782	0.43	−0.32	−0.29	−0.47	−0.20	−0.18
England/Lincolnshire 1851	NAPP	after 1850	Great Britain	31732	6638	0.65	−0.27	−0.37	−0.23	−0.42	−0.36
England/London (Parts Of Middlesex	NAPP	after 1850	Great Britain	99115	21365	0.44	−0.28	−0.20	0.01	−0.16	−0.36
England/Middlesex (extra London) 1851	NAPP	after 1850	Great Britain	12443	2582	0.33	−0.31	−0.26	−0.03	−0.17	−0.32
England/Norfolk 1851	NAPP	after 1850	Great Britain	32145	6913	0.49	−0.49	−0.20	−0.24	−0.18	−0.25
England/Northamptonshire 1851	NAPP	after 1850	Great Britain	16988	3737	0.29	−0.41	−0.15	−0.47	−0.09	−0.22
England/Northumberland 1851	NAPP	after 1850	Great Britain	24486	5280	0.60	0.13	−0.75	0.71	−0.57	−0.10
England/Nottinghamshire 1851	NAPP	after 1850	Great Britain	20599	4317	0.65	−0.26	−0.39	−0.39	−0.44	−0.32
England/Oxford 1851	NAPP	after 1850	Great Britain	12837	2850	0.16	−0.49	−0.28	−0.03	0.00	0.03
England/Rutland 1851	NAPP	after 1850	Great Britain	2153	433	0.47	−0.42	−0.11	−0.43	−0.18	−0.22
England/Shropshire 1851	NAPP	after 1850	Great Britain	19902	4162	0.70	−0.52	−0.72	−0.14	−0.91	0.29
England/Somerset 1851	NAPP	after 1850	Great Britain	41187	8638	0.57	−0.32	−0.30	−0.21	−0.16	0.61
England/Staffordshire 1851	NAPP	after 1850	Great Britain	43720	8580	0.50	−0.40	−0.60	−0.25	−0.69	−0.01
England/Suffolk 1851	NAPP	after 1850	Great Britain	27989	5960	0.41	−0.44	−0.26	−0.23	−0.13	−0.24
England/Surrey (extra London) 1851	NAPP	after 1850	Great Britain	16468	3394	0.31	−0.32	−0.17	0.20	−0.09	−0.35
England/Sussex 1851	NAPP	after 1850	Great Britain	27912	5658	0.31	−0.27	−0.17	0.23	−0.05	−0.41
England/Warwick 1851	NAPP	after 1850	Great Britain	24241	5292	0.39	−0.49	−0.48	−0.61	−0.15	0.14

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
England/Westmorland 1851	NAPP	after 1850	Great Britain	3888	778	0.58	–0.23	–0.76	0.35	–0.53	0.25
England/Wiltshire 1851	NAPP	after 1850	Great Britain	19457	4212	0.33	–0.32	–0.13	0.31	–0.06	0.19
England/Worcestershire 1851	NAPP	after 1850	Great Britain	19276	3971	0.70	–0.65	–0.75	–0.51	–0.56	0.63
England/Yorkshire East Riding 1851	NAPP	after 1850	Great Britain	14747	3155	0.45	–0.17	–0.71	–0.11	–0.43	–0.32
England/Yorkshire North Riding 1851	NAPP	after 1850	Great Britain	18227	3995	0.34	–0.29	–0.82	0.25	–0.44	0.06
England/Yorkshire West Riding 1851	NAPP	after 1850	Great Britain	106069	21630	0.36	–0.34	–0.67	0.02	–0.55	–0.21
ES/Catalonia rural 1880–1890	Mosaic	after 1850	West	7115	1601	0.38	–0.24	–0.22	–0.01	–0.37	–0.49
F/Northeast 1846	Mosaic	1800–1850	West	4444	1183	–0.04	–0.18	–0.26	0.22	–0.37	–0.28
F/Northwest 1846	Mosaic	1800–1850	West	5914	1389	0.75	0.66	0.15	0.66	0.11	–0.49
F/South 1846	Mosaic	1800–1850	West	6609	1605	0.32	–0.17	–0.21	0.04	–0.38	–0.55
F/Southwest 1831–1901	Mosaic	1800–1850	West	5109	1146	0.51	0.09	–0.18	0.00	–0.24	–0.76
H/Great Plain 1869	Mosaic	after 1850	Habsburg	3781	811	0.67	–0.79	0.27	–0.30	0.15	–0.20
H/North–East 1869	Mosaic	after 1850	Habsburg	2072	470	0.61	–0.55	0.42	–0.13	0.03	–0.71
H/Northern Transdanubia 1869	Mosaic	after 1850	Habsburg	4067	836	0.57	–0.88	–0.10	–0.36	0.08	0.18
H/Southern Transdanubia 1869	Mosaic	after 1850	Habsburg	3804	810	0.64	–0.88	0.22	–0.45	0.35	–0.11
HR/Dubrovnik area 1674	Mosaic	before 1800	Habsburg	1880	358	–0.10	–0.34	0.12	–0.12	0.36	–0.82
Iceland 1703	NAPP	before 1800	Scandinavia	51003	7572	0.81	–0.85	0.82	–0.96	0.97	–0.95
LT/Kovno 1847	Mosaic	1800–1850	East	9954	1530	0.40	–0.47	0.25	–0.03	0.13	–0.89
LT/Vilna 1847	Mosaic	1800–1850	East	9963	1398	0.48	–0.47	0.24	–0.13	0.21	–0.86
LV/Courland Goldingen+Pilten 1797	Mosaic	before 1800	East	13377	994	–0.01	–0.51	0.35	0.21	–0.05	–0.94
LV/Courland Mitau 1797	Mosaic	before 1800	East	10490	626	0.25	–0.48	0.28	0.07	0.05	–0.93

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
LV/Courland Selburg 1797	Mosaic	before 1800	East	7200	554	0.42	−0.45	0.20	−0.01	0.08	−0.90
LV/Courland Tuckum 1797	Mosaic	before 1800	East	4740	301	0.12	−0.51	0.34	0.13	0.00	−0.94
NL/North Brabant 1810	Mosaic	1800–1850	West	6488	1181	−0.06	0.27	−0.15	0.12	−0.45	−0.42
NL/Zeeland 1811	Mosaic	1800–1850	West	14746	2900	0.39	0.13	−0.12	−0.46	0.13	0.00
Norway/Aggershuus 1801	NAPP	1800–1850	Scandinavia	56761	10490	−0.18	−0.07	−0.28	−0.48	0.85	−0.75
Norway/Bratsberg 1801	NAPP	1800–1850	Scandinavia	42523	8161	0.20	−0.16	−0.14	−0.16	0.77	−0.42
Norway/Buskerud 1801	NAPP	1800–1850	Scandinavia	51373	9607	−0.02	−0.13	−0.23	−0.34	0.82	−0.62
Norway/Christians 1801	NAPP	1800–1850	Scandinavia	66661	11983	0.18	−0.18	−0.25	−0.70	0.67	−0.81
Norway/Finmarken 1801	NAPP	1800–1850	Scandinavia	26378	4837	0.62	−0.57	0.28	−0.86	0.87	−0.79
Norway/Hedemarken 1801	NAPP	1800–1850	Scandinavia	61025	11093	−0.18	−0.07	−0.28	−0.48	0.85	−0.75
Norway/Jarlsberg O 1801	NAPP	1800–1850	Scandinavia	34269	6606	−0.01	−0.14	−0.23	−0.33	0.82	−0.61
Norway/Lister Og M 1801	NAPP	1800–1850	Scandinavia	34722	6969	0.71	−0.36	0.29	−0.04	0.73	0.03
Norway/Nedenas 1801	NAPP	1800–1850	Scandinavia	28002	5441	0.71	−0.36	0.29	−0.04	0.73	0.03
Norway/Nordland 1801	NAPP	1800–1850	Scandinavia	51455	8942	0.31	−0.13	0.00	−0.77	0.85	−0.78
Norway/Nordre Berg 1801	NAPP	1800–1850	Scandinavia	52601	8773	0.68	−0.10	−0.17	−0.41	0.37	−0.71
Norway/Nordre Tron 1801	NAPP	1800–1850	Scandinavia	43443	8033	0.32	−0.28	−0.21	−0.50	0.57	−0.61
Norway/Romsdal 1801	NAPP	1800–1850	Scandinavia	55283	9748	0.42	−0.08	−0.39	−0.38	0.28	−0.66
Norway/Smaalehnene 1801	NAPP	1800–1850	Scandinavia	40231	7371	−0.09	−0.17	−0.28	−0.35	0.83	−0.62
Norway/Sondre Berg 1801	NAPP	1800–1850	Scandinavia	59101	10983	0.71	0.09	−0.21	−0.12	0.33	−0.63
Norway/Sondre Tron 1801	NAPP	1800–1850	Scandinavia	51272	9592	0.32	−0.28	−0.21	−0.50	0.57	−0.61
Norway/Stavanger 1801	NAPP	1800–1850	Scandinavia	38542	7074	0.78	−0.05	0.25	0.24	0.52	−0.22

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
PL/Central Belarus 1768–1804	Mosaic	before 1800	East	19176	3381	0.42	−0.52	0.04	−0.10	0.08	−0.73
PL/Chelmska Land 1791–1792	Mosaic	before 1800	East	25193	4520	0.73	−0.60	0.68	−0.47	0.63	−0.64
PL/Greater Poland 1666–1809	Mosaic	before 1800	East	5763	846	0.20	−0.37	−0.30	0.00	0.55	−0.26
PL/Kujavia 1766–1792	Mosaic	before 1800	East	13320	2065	0.24	−0.39	−0.09	−0.01	0.55	−0.33
PL/Lesser Poland 1789–1792	Mosaic	before 1800	East	14371	2511	0.53	−0.57	0.21	−0.29	0.46	−0.36
PL/Ostrzeszow County 1790–1791	Mosaic	before 1800	East	8358	1317	0.38	−0.47	0.05	−0.17	0.65	−0.35
PL/Podolia 1785–1819	Mosaic	1800–1850	East	5526	1131	0.32	0.06	0.32	0.19	−0.24	−0.73
PL/Polesia 1795	Mosaic	before 1800	East	25332	3884	0.10	−0.53	−0.05	0.14	−0.13	−0.58
PL/Silesia 1747–1805	Mosaic	before 1800	Germany	12265	2445	0.27	−0.55	−0.32	−0.12	0.57	−0.04
PL/Warmia 1695–1772	Mosaic	before 1800	East	2543	545	0.00	−0.40	0.07	0.14	0.42	−0.51
PL/Wielunskie County 1790–1792	Mosaic	before 1800	East	9945	1909	0.47	−0.52	0.17	−0.24	0.59	−0.38
PL/Zhytomyr County 1791	Mosaic	before 1800	East	14026	2100	−0.06	−0.17	0.41	0.14	−0.27	−0.81
RO/Eastern Wallachia 1838	Mosaic	1800–1850	Balkans	5089	1124	0.15	0.57	−0.18	−0.04	0.03	−0.73
RO/Moldavia Catholics 1781–1787	Mosaic	before 1800	Balkans	1992	374	0.08	0.52	−0.12	0.07	−0.22	−0.74
RO/Moldavia Catholics 1866–1879	Mosaic	after 1850	Balkans	3299	776	0.10	0.52	−0.12	0.08	−0.23	−0.74
RO/Northern Wallachia 1838	Mosaic	1800–1850	Balkans	5806	1508	0.19	0.59	−0.17	−0.04	0.10	−0.74
RO/Partium 1869	Mosaic	after 1850	Balkans	3471	742	0.56	0.30	0.04	0.30	−0.25	−0.72
RO/Southern Wallachia 1838	Mosaic	1800–1850	Balkans	5411	1370	0.09	0.59	−0.28	−0.12	0.17	−0.78
RO/Southwestern Wallachia 1838	Mosaic	1800–1850	Balkans	5240	1261	0.15	0.48	−0.30	−0.06	0.16	−0.84
RO/Transylvania 1869	Mosaic	after 1850	Balkans	5801	1298	0.35	0.53	−0.14	0.17	−0.18	−0.71
RUS/Braclav Governorate 1795	Mosaic	before 1800	East	8050	1537	−0.09	0.15	0.23	0.11	−0.29	−0.81

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
RUS/Gagarin villages 1814	Mosaic	1800–1850	East	2955	347	0.40	−0.45	−0.06	0.07	−0.13	−0.62
RUS/Moscow area 1897	Mosaic	after 1850	East	11559	1891	0.48	−0.64	−0.15	0.01	−0.11	−0.37
RUS/Ural Centre 1710	Mosaic	before 1800	East	4213	519	0.09	−0.52	0.39	−0.01	0.05	−0.89
RUS/Ural East 1710	Mosaic	before 1800	East	2723	355	0.10	−0.58	0.39	−0.03	0.05	−0.87
RUS/Ural North 1710	Mosaic	before 1800	East	5253	788	0.16	−0.51	0.40	−0.06	0.06	−0.90
RUS/Ural North Centre 1710	Mosaic	before 1800	East	3886	568	0.13	−0.50	0.39	−0.04	0.05	−0.90
RUS/Ural South 1710	Mosaic	before 1800	East	4713	660	0.08	−0.54	0.38	−0.01	0.06	−0.88
RUS/Ural West 1710	Mosaic	before 1800	East	3566	433	0.08	−0.45	0.40	0.00	0.06	−0.91
Scotland/Aberdeenshire 1881	NAPP	after 1850	Great Britain	217496	44481	0.18	0.56	−0.73	−0.16	−0.41	−0.49
Scotland/Angus and Forfarshire 1881	NAPP	after 1850	Great Britain	266383	60623	0.47	0.28	−0.56	0.19	−0.62	0.06
Scotland/Argyll 1881	NAPP	after 1850	Great Britain	78491	16725	0.55	0.69	−0.66	0.46	−0.82	−0.31
Scotland/Ayr 1881	NAPP	after 1850	Great Britain	217561	44551	0.53	0.65	−0.65	0.45	−0.81	−0.29
Scotland/Banff 1881	NAPP	after 1850	Great Britain	63003	13319	−0.15	0.56	−0.76	−0.41	−0.05	−0.60
Scotland/Berwickshire 1881	NAPP	after 1850	Great Britain	35401	7551	0.49	0.16	−0.61	0.57	−0.72	−0.11
Scotland/Bute 1881	NAPP	after 1850	Great Britain	17642	4035	0.53	0.70	−0.66	0.43	−0.84	−0.32
Scotland/Caithness 1881	NAPP	after 1850	Great Britain	38834	8601	−0.37	0.28	−0.66	−0.64	0.53	−0.73
Scotland/Clackmannanshire 1881	NAPP	after 1850	Great Britain	25674	5440	0.53	0.60	−0.64	0.40	−0.77	−0.20
Scotland/Dumfriesshire 1881	NAPP	after 1850	Great Britain	64268	13857	0.68	0.48	−0.59	0.58	−0.76	−0.24
Scotland/Dunbartonshire 1881	NAPP	after 1850	Great Britain	75389	14654	0.54	0.63	−0.65	0.42	−0.77	−0.23
Scotland/Fife 1881	NAPP	after 1850	Great Britain	170708	38658	0.53	0.59	−0.64	0.41	−0.77	−0.21
Scotland/Inverness-shire 1881	NAPP	after 1850	Great Britain	89194	18682	−0.19	0.57	−0.76	−0.52	0.01	−0.62

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
Scotland/Kincardinshire 1881	NAPP	after 1850	Great Britain	34471	7446	0.27	0.52	−0.71	−0.07	−0.48	−0.44
Scotland/Kinross–shire 1881	NAPP	after 1850	Great Britain	7466	1888	0.49	0.52	−0.61	0.33	−0.72	−0.13
Scotland/Kirkudbrightshire 1881	NAPP	after 1850	Great Britain	42153	9106	0.68	0.48	−0.59	0.58	−0.76	−0.24
Scotland/Lanarkshire 1881	NAPP	after 1850	Great Britain	701180	142369	0.53	0.65	−0.65	0.45	−0.81	−0.29
Scotland/Mid Lothian and Edinburghshire 1881	NAPP	after 1850	Great Britain	351237	72286	0.49	0.45	−0.60	0.38	−0.74	−0.10
Scotland/Morayshire and Elginshire 1881	NAPP	after 1850	Great Britain	44020	9676	−0.15	0.56	−0.76	−0.41	−0.05	−0.60
Scotland/Nairnshire 1881	NAPP	after 1850	Great Britain	9552	2100	−0.21	0.55	−0.76	−0.47	0.03	−0.62
Scotland/Orkney 1881	NAPP	after 1850	Great Britain	32043	7129	−0.29	0.20	−0.59	−0.66	0.56	−0.75
Scotland/Peeblesshire 1881	NAPP	after 1850	Great Britain	14265	2769	0.50	0.42	−0.61	0.51	−0.77	−0.17
Scotland/Perthshire 1881	NAPP	after 1850	Great Britain	128350	29450	0.47	0.30	−0.54	0.20	−0.62	0.07
Scotland/Renfrewshire 1881	NAPP	after 1850	Great Britain	200491	40377	0.53	0.65	−0.65	0.45	−0.81	−0.29
Scotland/Ross and Cromarty 1881	NAPP	after 1850	Great Britain	78745	17251	−0.19	0.57	−0.76	−0.52	0.01	−0.62
Scotland/Roxburghshire 1881	NAPP	after 1850	Great Britain	53450	11113	0.49	0.16	−0.61	0.57	−0.72	−0.11
Scotland/Selkirk 1881	NAPP	after 1850	Great Britain	25552	5295	0.48	0.16	−0.61	0.55	−0.72	−0.11
Scotland/Shetland 1881	NAPP	after 1850	Great Britain	29735	6179	0.00	−0.05	−0.07	−0.69	0.75	−0.64
Scotland/Stirlingshire 1881	NAPP	after 1850	Great Britain	112640	23171	0.53	0.59	−0.64	0.41	−0.77	−0.21
Scotland/Sutherland 1881	NAPP	after 1850	Great Britain	23401	5149	−0.30	0.47	−0.74	−0.52	0.18	−0.64
Scotland/West Lothian and Linlithgowshire 1881	NAPP	after 1850	Great Britain	42175	8495	0.52	0.55	−0.63	0.44	−0.78	−0.21
Scotland/Wigtown 1881	NAPP	after 1850	Great Britain	35083	7363	0.47	0.68	−0.61	0.32	−0.87	−0.25

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
SE/Jasenica 1863	Mosaic	after 1850	Balkans	7128	1099	0.18	0.06	0.01	0.05	0.23	−0.88
SE/Jasenica 1884	Mosaic	after 1850	Balkans	9434	1474	0.18	0.06	0.01	0.05	0.23	−0.88
SK/Central 1869	Mosaic	after 1850	Habsburg	1779	371	0.50	−0.77	−0.09	−0.29	0.19	0.05
SK/East 1869	Mosaic	after 1850	Habsburg	3601	682	0.55	−0.53	0.38	−0.24	0.20	−0.61
SK/West 1869	Mosaic	after 1850	Habsburg	3030	564	0.50	−0.86	−0.24	−0.32	0.06	0.26
Sweden/Älvsborg 1880	NAPP	after 1850	Scandinavia	268237	65791	0.22	−0.34	−0.10	−0.24	0.75	−0.49
Sweden/Blekinge 1880	NAPP	after 1850	Scandinavia	103566	27662	−0.20	0.34	−0.28	0.19	0.55	−0.28
Sweden/Gävleborg 1880	NAPP	after 1850	Scandinavia	147988	36210	−0.19	−0.30	0.07	−0.67	0.75	−0.77
Sweden/Göteborg och Bohus 1880	NAPP	after 1850	Scandinavia	184739	43342	0.22	−0.34	−0.10	−0.24	0.75	−0.49
Sweden/Gotland 1880	NAPP	after 1850	Scandinavia	48351	11845	−0.76	−0.56	0.32	0.57	−0.32	−0.90
Sweden/Halland 1880	NAPP	after 1850	Scandinavia	130184	30703	−0.22	0.62	−0.48	−0.03	0.69	−0.37
Sweden/Jämtland 1880	NAPP	after 1850	Scandinavia	85788	21622	0.31	−0.48	0.20	−0.58	0.84	−0.60
Sweden/Jönköping 1880	NAPP	after 1850	Scandinavia	167208	40812	−0.30	0.23	−0.27	0.13	0.39	−0.53
Sweden/Kalmar 1880	NAPP	after 1850	Scandinavia	235246	61746	−0.48	0.01	−0.17	0.20	0.19	−0.76
Sweden/Kopparberg 1880	NAPP	after 1850	Scandinavia	190009	43018	−0.25	−0.27	−0.05	−0.67	0.78	−0.81
Sweden/Kristianstad 1880	NAPP	after 1850	Scandinavia	212890	58319	−0.29	0.50	−0.39	0.01	0.93	−0.15
Sweden/Kronoberg 1880	NAPP	after 1850	Scandinavia	156754	39850	−0.30	0.23	−0.27	0.13	0.39	−0.53
Sweden/Malmöhus 1880	NAPP	after 1850	Scandinavia	278631	72782	−0.29	0.50	−0.39	0.01	0.93	−0.15
Sweden/Norrbotten 1880	NAPP	after 1850	Scandinavia	85587	18265	0.44	−0.45	0.21	−0.77	0.89	−0.74
Sweden/Örebro 1880	NAPP	after 1850	Scandinavia	168058	40113	−0.27	−0.32	0.13	−0.67	0.66	−0.86
Sweden/Östergötland 1880	NAPP	after 1850	Scandinavia	226062	57508	−0.53	−0.31	0.11	−0.30	0.36	−0.87

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
Sweden/Skaraborg 1880	NAPP	after 1850	Scandinavia	239319	56194	−0.56	−0.18	−0.26	−0.47	0.70	−0.78
Sweden/Södermanland 1880	NAPP	after 1850	Scandinavia	122941	30557	−0.43	−0.34	0.16	−0.39	0.40	−0.88
Sweden/Stockholm 1880	NAPP	after 1850	Scandinavia	126046	32735	−0.28	−0.33	0.16	−0.56	0.52	−0.89
Sweden/Uppsala 1880	NAPP	after 1850	Scandinavia	118412	30120	−0.27	−0.33	0.16	−0.57	0.53	−0.88
Sweden/Värmland 1880	NAPP	after 1850	Scandinavia	255831	61066	−0.33	−0.31	−0.10	−0.59	0.77	−0.78
Sweden/Västerbotten 1880	NAPP	after 1850	Scandinavia	89373	19783	0.35	−0.38	0.18	−0.74	0.89	−0.74
Sweden/Västernorrland 1880	NAPP	after 1850	Scandinavia	150353	39977	−0.01	−0.36	0.16	−0.55	0.80	−0.65
Sweden/Västmanland 1880	NAPP	after 1850	Scandinavia	106001	26511	−0.22	−0.31	0.15	−0.66	0.62	−0.87
UKR/Berdychiv region North rural 1897	Mosaic	after 1850	East	3646	642	−0.06	−0.07	0.38	0.11	−0.26	−0.81
UKR/Berdychiv region South rural 1897	Mosaic	after 1850	East	2267	429	−0.12	−0.05	0.37	0.10	−0.28	−0.81
UKR/Hetmanate North 1765	Mosaic	before 1800	East	4046	546	0.23	−0.38	−0.05	0.17	−0.20	−0.66
UKR/Hetmanate Southeast 1765	Mosaic	before 1800	East	5479	724	−0.11	−0.27	0.38	0.16	−0.29	−0.80
UKR/Hetmanate Southwest 1765	Mosaic	before 1800	East	4564	698	−0.22	−0.22	0.41	0.18	−0.35	−0.81
UKR/Hetmanate West 1765	Mosaic	before 1800	East	4302	674	−0.09	−0.29	0.21	0.23	−0.31	−0.77
USSR/Central 1926/27	Mosaic	after 1850	East	3928	834	0.33	−0.88	0.57	−0.33	0.69	−0.75
USSR/East 1926/27	Mosaic	after 1850	East	2293	431	0.32	−0.86	0.56	−0.32	0.73	−0.73
USSR/North 1926/27	Mosaic	after 1850	East	3849	795	0.39	−0.89	0.55	−0.49	0.81	−0.76
USSR/Northeast 1926/27	Mosaic	after 1850	East	2201	441	0.34	−0.87	0.53	−0.40	0.80	−0.72
USSR/Northwest 1926/27	Mosaic	after 1850	East	1523	312	0.73	−0.71	0.41	−0.92	0.89	−0.82
USSR/South 1926/27	Mosaic	after 1850	East	3233	664	0.28	−0.87	0.61	−0.18	0.53	−0.75
USSR/Southeast 1926/27	Mosaic	after 1850	East	1238	284	0.26	−0.85	0.60	−0.16	0.59	−0.72

REGION	Data source	Period	Macro-region	Sample size	No. of houshs.	Spearman's correlation coefficients of the local correlations					
						SMAM-SERV	SMAM-CMHD	SMAM-NUCL	SERV-CMHD	SERV-NUCL	CMHD-NUCL
USSR/West 1926/27	Mosaic	after 1850	East	4997	1121	0.38	−0.86	0.55	−0.35	0.47	−0.81
Wales/Anglesey 1851	NAPP	after 1850	Great Britain	42165	9266	0.62	−0.32	−0.52	0.20	−0.66	−0.06
Wales/Brecknockshire 1851	NAPP	after 1850	Great Britain	43971	8980	0.74	−0.45	−0.71	−0.49	−0.88	0.66
Wales/Cardiganshire 1851	NAPP	after 1850	Great Britain	92728	20524	0.53	−0.35	−0.55	−0.30	−0.88	0.47
Wales/Carmarthenshire 1851	NAPP	after 1850	Great Britain	72439	15571	0.53	−0.37	−0.54	−0.51	−0.90	0.61
Wales/Carnarvonshire 1851	NAPP	after 1850	Great Britain	92576	19795	0.62	−0.37	−0.53	0.07	−0.74	0.06
Wales/Denbighshire 1851	NAPP	after 1850	Great Britain	76397	15885	0.69	−0.45	−0.53	0.06	−0.74	0.00
Wales/Flintshire 1851	NAPP	after 1850	Great Britain	27780	5848	0.69	−0.46	−0.52	0.05	−0.71	−0.04
Wales/Glamorganshire 1851	NAPP	after 1850	Great Britain	204962	39978	0.61	−0.38	−0.54	−0.45	−0.68	0.72
Wales/Merionethshire 1851	NAPP	after 1850	Great Britain	46797	10058	0.65	−0.43	−0.58	−0.02	−0.79	0.16
Wales/Monmouthshire 1851	NAPP	after 1850	Great Britain	122926	24403	0.73	−0.44	−0.57	−0.34	−0.49	0.71
Wales/Montgomeryshire 1851	NAPP	after 1850	Great Britain	64069	13073	0.67	−0.46	−0.65	−0.12	−0.85	0.30
Wales/Pembrokeshire 1851	NAPP	after 1850	Great Britain	68700	14577	0.44	−0.25	−0.47	−0.40	−0.88	0.50
Wales/Radnorshire 1851	NAPP	after 1850	Great Britain	23352	4773	0.76	−0.48	−0.77	−0.37	−0.92	0.52

A7. DATA REFERENCES

Mosaic data:

Karl Kaser, Siegfried Gruber, Gentiana Kera, Enriketa Pandelejmoni. *1918 Census of Albania, Version 0.1* [SPSS file]. Graz, 2011.

Laboratory of Historical Demography (MPIDR). *1869 Census of Hungary, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1910 Census of Austria, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Familiekunde Vlaanderen and Laboratory of Historical Demography (MPIDR). *1814 Census of Western Flanders, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Ulf Brunnbauer. *Household registers of Rhodope region, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). Status Animarum for Lisac and Pridvorje, *Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.

Danish Data Archive. *1803 Census of Schleswig and Holstein, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

Laboratory of Historical Demography (MPIDR). *1846 Census of France, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

University of Bordeaux. *1831 Census of Sallespisse, Version 1.2* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2013.

University of Bordeaux. *1836 Census of Boulazac, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

University of Bordeaux. *1841 Census of St. Jean de Luz, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

University of Bordeaux. *1841 Census of Targon, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

University of Bordeaux. *1876 Census of Boulazac, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

University of Bordeaux. *1901 Census of Sauternes, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.

University of Bordeaux. 1846 Census of Saint-Émilion, Version 1.2 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

University of Bordeaux. 1856 Census of Saint-Émilion, Version 1.2 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1846 German Customs Union Census, Version 2.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1846 Census of Höhscheid, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1858 German Customs Union Census, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

Laboratory of Historical Demography (MPIDR). *1861 Census of Haigerloch, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.

State Main Archive Schwerin, Laboratory of Historical Demography (MPIDR), and Department of Multimedia and Data Processing, University of Rostock. *1819 Census of Mecklenburg-Schwerin, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2016.

State Main Archive Schwerin, Laboratory of Historical Demography (MPIDR), and Department of Multimedia and Data Processing, University of Rostock. 1819 Census of

Rostock, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.
State Main Archive Schwerin, Laboratory of Historical Demography (MPIDR), and Department of Multimedia and Data Processing, University of Rostock. 1867 Census of Mecklenburg-Schwerin, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2016.
State Main Archive Schwerin, Laboratory of Historical Demography (MPIDR), and Department of Multimedia and Data Processing, University of Rostock. 1867 Census of Rostock, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.
State Main Archive Schwerin, Laboratory of Historical Demography (MPIDR), and Department of Multimedia and Data Processing, University of Rostock. *1900 Census of Rostock, Version 1.0* [Mosaic Historical Microdata File]. Rostock, Germany: www.censusmosaic.org, 2013.
Laboratory of Historical Demography (MPIDR). *1749 Status Animarum of Münster, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.
Laboratory of Historical Demography (MPIDR). *1690-1713 Status Animarum of Oldenburger Münsterland, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.
Laboratory of Historical Demography (MPIDR). 1811 Census of Zeeland, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.
Laboratory of Historical Demography (MPIDR). 1810 Census of North Brabant, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.
Mikołaj Szołtysek (2012) CEURFAMFORM database, Version 23 [SPSS file]. Rostock.
Laboratory of Historical Demography (MPIDR). 1781-1879 Status Animarum in Moldavia, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.
Laboratory of Historical Demography (MPIDR). *1838 Census of Wallachia, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.
Laboratory of Historical Demography (MPIDR). *1710 Russian enumeration of Ural region, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.
Laboratory of Historical Demography (MPIDR). *1765 Rumyantsev census of Hetmanate region, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.
Laboratory of Historical Demography (MPIDR). *1795 Braclav Region Revision Lists, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.
Laboratory of Historical Demography (MPIDR). *1797 Revision lists of Courland, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.
Laboratory of Historical Demography (MPIDR). *1814 Russian list of inhabitants, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2014.
Laboratory of Historical Demography (MPIDR). 1847 Lithuanian Estate Household Listings, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.
Laboratory of Historical Demography (MPIDR). *1897 Russian census, Moscow region, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org,2014.
Tomasz Jankowski. *1897 Russian Census, Berdychiv region, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.
Joel M. Halpern and Siegfried Gruber. *1863 Census of Jasenički srez, Serbia, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.
Joel M. Halpern and Siegfried Gruber. *1884 Census of Jasenički srez, Serbia, Version 1.1* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2012.
Laboratory of Historical Demography (MPIDR). 1880-1890 Local Censuses in Catalonia, Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2015.
Ulrich Pfister. *1634-1764 Soul listings of canton Zürich, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.
Vienna Database on European Family History. *1671-1685 Church listings of canton Zürich,*

Version 1.0 [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.

Vienna Database on European Family History. *1870 Census of Zürich, Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2017.

Anderson, D., Gruber, S., Szołtysek, M. (2018). *The Polar Census 1926-27 Mosaic sample*. *Version 1.0* [Mosaic Historical Microdata File]. www.censusmosaic.org, 2018.

NAPP data:

Minnesota Population Center. *North Atlantic Population Project: Complete Count Microdata. Version 2.3* [Machine-readable database]. Minneapolis: Minnesota Population Center, 2016.

- England and Wales 1881: K. Schürer and M. Woollard, National Sample from the 1851 Census of Great Britain [computer file], Colchester, Essex: History Data Service, UK Data Archive [distributor], 2003
- Scotland 1881: K. Schürer and M. Woollard, National Sample from the 1881 Census of Great Britain [computer file], Colchester, Essex: History Data Service, UK Data Archive [distributor], 2003.
- Denmark 1787: Nanna Floor Clausen, Danish National Archives. 1787 Census of Denmark, Version 1.0
- Iceland 1703: Ólöf Garðarsdóttir (University of Iceland) and National Archives of Iceland (NAI). 1703 Census of Iceland, Version 1.0.
- Norway 1801: The Digital Archive (The National Archive), University of Bergen, and the Minnesota Population Center. Census of Norway 1801, Version 1.0. Bergen, Norway: University of Bergen, 2011.
- Sweden 1880: The Swedish National Archives, Umeå University, and the Minnesota Population Center. National Sample of the 1880 Census of Sweden, Version 1.0. Minneapolis: Minnesota Population Center [distributor], 2014.

LIST OF WORKING PAPERS

1. László Hablicsek, Pál Péter Tóth: The Role of International Migration in Maintaining the Population Size of Hungary between 2000–2050
2. Maritetta Pongrácz: Birth out of Wedlock
3. Attila Melegh: East/West Exclusions and Discourses on Population in the 20^{th} Century
4. Zsolt Spéder: Fertility and Structural Change in Hungary
5. Sándor Illés: Foreigners in Hungary: Migration from the European Union
6. Magdalena Muszyńska: Family Models in Europe in the Context of Women's Status
7. Attila Melegh, Elena Kondratieva, Perttu Salmenhaare, Annika Forsander, László Hablicsek, Adrienn Hegyesi: Globalisation, Ethnicity and Migration. The Comparison of Finland, Hungary and Russia
8. Zsolt Spéder, Balázs Kapitány: Poverty and Deprivation: Assessing Demographic and Social Structural Factors
9. Etelka Daróczi: Ageing and Health in the Transition Countries of Europe – the Case of Hungary
10. Péter Őri: Demographic Patterns and Transitions in $18–20^{th}$ Century Hungary. County Pest-Pilis-Solt-Kiskun in the Late 18^{th} and Early 20^{th} Centuries
11. Zsolt Spéder, Balázs Kaptány: Ideational Factors and Parenthood. A Gender- and Parity Specific Analysis in a Post-Communist Society
12. Irén Gödri: The Role of Ethnicity and Social Capital in Immigration to Hungary
13. Attila Melegh, Arland Thornton, Dimiter Philipov, Linda Young-DeMarco: Mapping Societal Developmental Hierarchies in Europe: a Bulgarian Perspective
14. Balázs Kapitány, Zsolt Spéder: Factors Affecting the Realisation of Child-Bearing Intentions in Four European Countries
15. Zsolt Spéder, Balázs Kapitány: Realising Birth Intention in European Comparison – Understanding the Post-Communist Fertility Transition
16. Tamás Faragó: Historical Demography in Hungary: a History of Research
17. Attila Melegh: Net Migration and Historical Development in Southeastern Europe since 1950
18. Róbert I. Gál, Endre Szabó, Lili Vargha: The age profile of invisible transfers: the true size of asymmetry in inter-age reallocations
19. Irén Gödri, Béla Soltész, Boróka Bodacz-Nagy: Immigration or emigration country? Migration trends and their socio-economic background in Hungary: A longer-term historical perspective
20. Péter Őri, Levente Pakot: Residence patterns in nineteenth century Hungary: Evidence from the Hungarian MOSAIC sample.
21. Zsuzsa Blaskó: Studying emigration by extending a large-scale household survey. Methodology, evaluation and descriptive findings
22. Lili Vargha, Róbert Iván Gál, Michelle O. Crosby-Nagy: Household production and consumption over the lifecycle: the National Time Transfer Accounts in 14 European countries
23. Kálmán Joubert, Gyula Gyenis: The Hungarian Longitudinal Growth Study: From birth to the age of 18 years
24. Róbert Iván Gál, Árpád Törzsök: The Savings Gap in Hungary
25. Lajos Bálint: Suicide in the Hungarian Kingdom
26. Róbert Iván Gál, Pieter Vanhuysse, Lili Vargha: Pro-elderly welfare states within pro-child societies: Incorporating family cash and time into intergenerational transfers analysis
27. Zsolt Spéder, Lívia Murinkó, Lívia Sz. Oláh: Sailing close to the wind?
The effects of third birth policies in post-communist Hungary
28. Róbert Iván Gál, Márton Medgyesi: Financing the lifecycle or mitigating poverty: Redistribution in the Hungarian welfare system by age and income
29. Irén Gödri, Gábor Attila Feleky: Selection of migrants and realization of migration intentions – lessons from a panel study
30. Zsuzsanna Veroszta (ed.): Technical Report. Growing Up in Hungary - Cohort '18 Hungarian Birth Cohort Study. Prenatal Research, Preparational phase.
31. Róbert Iván Gál, Márta Radó: Participation and Postponed Retirement in Central and Eastern Europe
32. Zsuzsanna Veroszta (ed.): Conceptual framework. Growing Up in Hungary – Cohort '18 Hungarian birth cohort study

The above Working Papers can be ordered at the following e-mail addresses:
workingpapers@demografia.hu

The Working Papers are also available online at:
http://www.demografia.hu/workingpapers