Permanyer, Iñaki; Esteve, Albert; García-Román, Joan. «Decomposing patterns of college marital sorting in 118 countries : structural constraints versus assortative mating». Social Science Research, Vol. 83, Núm. 102313 (Setembre 2019). DOI 10.1016/j.ssresearch.2019.06.004

Postprint article:

PERMANYER, Iñaki; ESTEVE, Albert; GARCÍA-ROMÁN, Joan (2019) "Decomposing patterns of college marital sorting in 118 countries: Structural constraints versus assortative mating". ***Social Science Research*** **(online 27 June 2019 / ISSN: 0049-089X).**

Title: Decomposing patterns of college marital sorting in 118 countries: Structural constraints versus assortative mating[1]

Authors:

Iñaki Permanyer (Corresponding Author): Centre d'Estudis Demogràfics (a member of the CERCA Programme / Generalitat de Catalunya). Carrer de Ca n'Altayó, Edifici E-2, Campus de la UAB, 08193 Cerdanyola del Vallès, Spain. E-mail: inaki.permanyer@uab.es. Phone: +34935813060.

Albert Esteve, Centre d'Estudis Demogràfics.

Joan Garcia-Román, Centre d'Estudis Demogràfics.

[1] Acknowledgements: The research leading to these results has received funding from the European Research Council (ERC-2014-StG-637768, EQUALIZE project), from the Spanish Ministry of Economy and Competitiveness "Ramón y Cajal" Research Grant Program (RYC-2013-14196) and the Spanish Ministry of Economy, Industry and Competitiveness National R&D&I Plan GLOBFAM (RTI2018-096730-B-I00).

Abstract: Two broad forces shape the patterns of marital sorting by education: structural constraints and assortative mating. However, we lack specific and comparative quantification of the extent of these two forces. In this paper, we measure the specific contributions of (i) assortative mating, (ii) the level of college education and (iii) the gender gap in education on marital sorting patterns and the corresponding polarization levels between college and non-college educated couples. Unlike previous studies, we adopt a large- cross-national approach including 118 countries and more than 258 observations spanning from 1960 up to 2011. Methodologically, we develop counterfactual modelling techniques to compare observed patterns of marital sorting with expected patterns derived from alternative structural and assortative mating conditions. Our findings indicate that changes in college marital sorting and increases in polarization between college- and non-college-educated populations are overwhelmingly driven by structural constraints, namely the expansion of college education. Instead, educational assortative mating plays a limited role – accounting only for 5% of the observed changes in marriage market polarization.

Keywords: Marital sorting patterns, education assortative mating, education expansion, gender gap in education, college education, polarization.

Decomposing patterns of college marital sorting in 118 countries: Structural constraints versus assortative mating[2]

Iñaki Permanyer
Albert Esteve
Joan Garcia-Román

1. Introduction

The heightened tendency to mate individuals with similar levels of education more frequently that would be expected under random circumstances (i.e. educational assortative mating) has potentially contributed to the increasingly unequal and diverging societies that are at the center of recent scholarship concerns (McLanahan 2004, Reardon, 2011). Yet, while assortative mating captures most of the sociological attention (Mare 1991, Schwartz and Mare 2005), mating patterns are predominantly shaped by the structural constraints of the marriage market, that is, the composition of the marriageable population by educational attainment. Yet we lack specific and comparative cross-national quantification of the extent to which structural constraints, on one side, and assortative mating, on the other, influence the observed patterns of marital sorting across societies and contribute to marriage market polarization between high- and low-educated individuals. This paper documents macro-level patterns and trends in educational marital sorting in 118 countries and measures the influence of structural constraints and assortative mating. We focus on the divide between college and non-college education, the most salient educational boundary of the 21st century.

[2] Acknowledgements: The research leading to these results has received funding from the European Research Council (ERC-2014-StG-637768, EQUALIZE project), from the Spanish Ministry of Economy and Competitiveness “Ramón y Cajal” Research Grant Program (RYC-2013-14196) and the Spanish Ministry of Economy, Industry and Competitiveness National R&D&I Plan GLOBFAM (RTI2018-096730-B-I00).

The paper adopts a macro and global perspective including countries representing more than 98% of today's world population, with multiple observations over time. The data spans from 1960 to 2011. Most data for this analysis come from the international version of Integrated Public Use of Microdata Series (IPUMS-i) database of census microdata samples, with a complementary use of various household surveys. This allows capturing macro structures of marital sorting between college and non-college across societies with markedly different levels of college-educated populations. To assess the degree to which the marriage market is partitioned between college and non-college educated couples, we propose an index of 'marriage market polarization' (or 'polarization' for short). Such index allows ranking societies based on the extent to which married populations split in the arguably two most socially distant groups: college-educated couples and the non-college educated ones. In addition, this index permits easy decompositions and manageability to different counterfactual scenarios, which are employed to quantify the relative contribution of structural constraints and assortative mating to the countries' observed levels of polarization.

Our findings indicate that the increasing polarization between college- and non-college-educated populations is overwhelmingly driven by structural constraints, namely the expansion of college education, and that educational assortative mating plays a very limited role.

2. Background

The importance of being college-educated in the marriage market

As the world embraces the tenets of an increasingly globalized and competitive knowledge-based economy, college education will become the most salient educational boundary of the 21st century as literacy was during the 19th century and a significant part of the 20th century. The importance of being college educated is noticeable in many dimensions of people's lives. College-educated individuals tend to have higher levels of employment, better paying jobs, better health outcomes and greater access to cultural resources than do non-college-educated ones (Hout 2012, Harmon et al. 2001). In addition, these outcomes have implications on the wellbeing of children (Atkinson et al. 2011, Guryan et al. 2008); a circumstance that contributes to the intergenerational transmission of (dis)advantage as children of college-educated parents have, on average, better school and health outcomes (Lamerz et al. 2005, Davis-Kean 2005).

Education is also one of the most important stratification variables of demographic behavior (Lutz, Butz and KC 2014). High-educated people tend to form unions and have children later in time. They also show different propensities to opt for cohabitation instead of marriage and to dissolve their unions. Despite the strength and the direction of the relationship between education and these transitions vary widely across countries and over time, being college-educated matters for most of these transitions. And partner choice is not an exception.

Education influences the age at which individuals enter the marriage market, structures the opportunities of the marriage market and shapes expectations towards marriage and potential mates (Kalmijn 1998, Mare 1991, Blossfeld 2009, Blossfeld and Timm 2003). If not the most important, educational attainment is a major stratifying dimension of modern marriage markets. By level of education, the college educated tend to show the highest levels of homogamy (Smits, Ultee and Lammers 2000, Smits and Park 2009, Scwhartz and Mare 2005, Schwartz 2013). As a result, college education has become the main dividing line in modern marriage markets across the globe (Smits 2003), including in the US (Scwhartz and Mare 2005), Europe (Domański and Przybysz 2007), East Asia (Smits and Park 2009) and Latin America (Esteve and McCaa 2007). Universities are efficient marriage markets. They bring together men and women of similar ages at the age in which they typically engage in their first long-term dating relationships. Universities contribute to homophily in social networks (McPherson et al. 2001), which is later reproduced in the working and leisure environments and in digital social networks as well (Potarca 2017).

Factors that shape who marries with whom

Social scientists have long been interested in *who marries whom* because of its informative power regarding social stratification and its implications for the intergenerational transmission of (dis)advantage. As stated in Kalmijn (1998), mating patterns are driven by structural opportunities, third party influences and individual preferences. Measuring these clearly defined dimensions has always been a challenge because of data limitations and interactions between factors.

Most research on assortative mating relies on cross-sectional observations of mating patterns derived from censuses, register data and surveys. The analyses have consisted in comparing the observed distribution of couples to the random distribution of couples based on the same structural constraints or alternative model specifications. The gap between the observed and the expected distribution has been used as an indirect measure of the force of *assortative mating* (Schwartz 2013, Schwartz and Mare 2005, Smits et al. 1998). Whereas this approach has been criticized for neither including the population at risk nor providing a clear measurement of individual preferences or third party influences, it is widely accepted as an indicator of the extent to which individuals mate within or across groups beyond what it would be expected under random circumstances. By and large, log-linear models are the most popular techniques in this type of analyses. They offer the possibility of examining the interaction between two or more variables beyond marginal constraints. When applied to cross-tabulations of spousal characteristics, in our case educational attainment, evidence consistently shows strong support for assortative mating.

Beyond assortative mating, mating patterns also depend on the structure of the marriage market, i.e., the distribution of the marriageable population by educational attainment. We distinguish between two structural factors: the level of college of education and the gender gap in education. The expansion of *college education*[3] will mechanically contribute to the growth of college-educated couples but will also broaden opportunities for non-college-educated individuals to find a college-educated partner and form a 'mixed' (i.e., heterogamous) couple. An important feature of the expansion of college education and its influence on the marriage market is the *gender gap in education*. The expansion of college education has not been gender neutral (Dorius and Firebaugh 2010; Dorius 2013; Grant and Behrman 2010). Initially favoring males, the gender gap has closed rapidly in recent years and, in many countries, has even reversed in favor of women (Esteve, Garcia and Permanyer 2012), a trend that is expected to continue over the next few decades[4] (KC et al. 2010; Lutz and KC, 2011). Therefore, the 'excess' of college-educated women may reduce the number of college-educated couples and

[3] While by 1970 6.4% of the world's population aged 25-29 had obtained a college degree, three decades later, this proportion had increased to 13%, and the expected figure for 2050 is 29.4% (KC et al. 2010).

[4] In 1970, men represented 63.6% of the total college-educated population. This percentage decreased to 52.6% in 2000, and it is likely to reach 44% in 2050, with most high-income countries reaching lower levels (KC et al. 2010).

increase the number of heterogamous couples, more specifically those in which women have more education than their male partners (Esteve et al. 2012; 2016).

Marriage market polarization

In recent times, some scholars have expressed their concerns about the potentially negative consequences of increasing assortative mating (Schwartz and Mare 2005, Esping-Andersen 2009, Blossfeld 2009, Schwartz 2013, Kalmijn 2013). If resourceful individuals form couples together, and people without resources partner each other too, differences across households are expected to be higher compared to a situation where partnerships are formed across those groups. In association with other family dynamics (like increasing single parenthood among low social status households, maternal employment bias in favor of the higher educated, or the deteriorating position of low-skilled males), assortative mating is expected to further increase the distance between social strata and lead to increasingly unequal and polarized societies (Esping-Andersen 2016) – that is: societies separated in opposite and antagonistic poles (e.g., 'the rich' vs 'the poor', or 'the highly educated' vs 'the low educated'). Scholars have expressed similar concerns when inspecting recent polarizing dynamics in the distribution of income (e.g. Esteban and Ray 2011), or in the distribution of population shares across ethno-linguistic or religious groups (Montalvo and Reynal-Querol 2005, Esteban et al 2012). Put together, these studies provide empirical evidence that increasing polarization along socio-economic lines could have potentially negative consequences for the corresponding societies.

In an attempt to measure the extent to which the population in union is partitioned in two opposite and antagonistic poles (the college-educated couples and the non-college educated ones), we introduce an index of 'marriage market polarization' (or, simply, ‘polarization’). Our polarization measure is maximized in the absence of mixed couples, as this is expected to be scenario where social closure and distance between strata are at their height. The polarization index proposed in this paper serves a double purpose. First, it allows identifying whether societies are approaching those distributions that have triggered recent scholarly and policy concerns[5]. Second, the measure is amenable to different counterfactual exercises that allow quantifying how important the structural constraints and assortative mating are when explaining changes in the observed levels of polarization.

Conditioned by the nature of the data, the measure of polarization we propose is confined to prevailing couples within specific age groups. Given that at these ages a sizeable share of the population may not be in union and the probability of not being in union may vary by educational attainment, we cannot directly extrapolate the observed levels of polarization within couples to the total population. We will perform sensitivity analysis to assess whether excluding the population not in union will have had a large effect on our measure of polarization. Besides, educational homogamy is not the only source of polarization in the marriage markets. Marriage markets are also heavily stratified by other dimensions such as ethnicity, race, wealth, income or socioeconomic status. Education might partially capture these dimensions to the extent they are inter-correlated. However, we are not examining in this paper the relationship between education and other dimensions of the marriage market and, thus, this limitation conditions the interpretations of our results.

[5] Albeit defined in a different way, several scholars have examined the relationship between economic, religious or ethnic polarization and the levels of social tension, unrest, violence, or even the occurrence of Civil Wars (e.g. Montalvo and Reynal-Querol 2005, Esteban and Ray 2011, Esteban et al 2012).

Previous research based on country-specific case studies has already investigated whether changes in inequality can be accounted for by changes in educational assortative mating (Breen and Salazar 2011, Breen and Andersen 2012). As a possible explanation to the small effect of the latter on the former identified in the US (Breen and Salazar 2011), it has been hypothesized that changes in education assortative mating might not have been large enough to produce the expected effect on inequality (Schwartz 2013). Following this line of thought, in this paper we use counterfactual techniques to test whether extreme levels of assortative mating could have substantially altered the levels of marriage market polarization (more details given in section 4.4).

Aims of the paper

This paper contributes to the literature on marital sorting patterns by education in several ways. We measure the specific contributions of (i) assortative mating, (ii) the level of college education and (iii) the gender gap in education on marital sorting patterns and the corresponding polarization levels in a truly global perspective. By adopting a large- cross-national perspective including 118 countries with repeated observations spanning from 1960 up to 2011, we contribute to previous findings that have examined related issues for Europe (De Hauw et al. 2017; Grow and Van Bavel, 2015) or that have investigated the implications of the gender gap reversal in education on assortative mating in an international perspective (Esteve et al. 2012; 2016).

To achieve our goals, we have developed a set of formal identities to characterize patterns of educational marital sorting and the corresponding polarization levels as a function of the three aforementioned factors. These mathematical identities prove very useful to (a) quantify *precisely* the relationships between the two sets of variables (section 4.1), and (b) develop the counterfactual and benchmarking techniques we have introduced to compare observed patterns of marital sorting with the hypothetical patterns that would be observed if other structural conditions or behavioral traits prevailed (sections 4.3 and 4.4). Inter alia, this allows investigating whether polarization levels would be substantially different if alternative levels of assortative mating had been operating.

3. Data

Our analysis is based on a vast collection of census and survey microdata samples from 118 countries spanning 1960 to 2011, which represent more than 98% of today's world population. We have gathered 258 samples of microdata, obtaining 149 census samples from the Integrated Public Use of Microdata Series International project (Minnesota Population Center, 2014); 63 from Demographic and Health Surveys; 37 from European Labor Force Surveys; 5 from the European Statistics on Income and Living Conditions; and 4 from the Generations and Gender Survey. By decade, there are 5 samples from the 1960s, 25 from the 1970s, 28 from the 1980s, 69 from the 1990s, 94 from the 2000s, and 37 from the 2010s. By continent, 57 samples come from 32 African countries, 53 from 28 Asian countries, 74 from 33 European countries, 68 from 24 Latin American countries, 5 from North America (the United States) and 1 from Australia (Oceania). The final dataset includes only samples in which the education of the spouses can be identified.

The analysis is restricted to the population in heterosexual married or cohabiting unions in which the women are 25-34 years old. In this way, we minimize the biases of union dissolution, educational upgrades and remarriage (Schwartz and Mare 2012) and avoid overlapping cohorts across observations because in most cases, our observations are 10 or more years apart. The final dataset totals more than 14 million weighted individual records.

The analysis is restricted to the in union population. As mentioned before, this has some limitations because it does not consider unmarried individuals (e.g. singles, divorcees, separated and widows). This is a recurring limitation in this type of research. Therefore, caution is required when equating mating preferences and our measure of assortative mating and polarization. The later only captures the deviation of the observed distribution of couples from the expected one under a random allocation of spousal characteristics. Whereas this is not a perfect measurement of assortative mating, we expect that we are able to capture the main differences across countries. To assess whether the main findings of the paper are overly affected by the aforementioned methodological choices, we have done two kinds of sensitivity analyses. On the one hand, we have taken into consideration models incorporating the unmarried population[6]. On the other hand, we have considered alternative age ranges of the married population (e.g. 30-39). The main findings of the paper are not altered when choosing these alternative criteria (results not shown here but available from the authors upon request).

Educational attainment is dichotomized into non-college and college education. Educational systems vary widely across the globe, and thus, their harmonization is

[6] Following the methodological approach suggested in Breen and Salazar (2011:822-833), we have implemented counterfactual modelling techniques that take into consideration both the married *and* the unmarried populations. The advantage of this approach is that it encompasses the entire marriage market. The disadvantage is that its estimation requires numerical approximation based iterative optimization techniques (e.g. the Deming-Stephan algorithm) that do *not* admit analytical (i.e. 'solvable') solutions like the ones developed in this paper when restricting the attention to the population in union (see the set of identities in equation [7]). Since the models focusing on the entire population and the ones focusing on the married population generate very similar outcomes, we have opted for the last ones (i.e. the simplest and most parsimonious).

problematic (Esteve and Sobek 2003). However, there are some educational thresholds that are fairly standard across societies. College education is one of them. Virtually all censuses and surveys used in our research identify completed college / tertiary education without ambiguity (coding is available from the authors). Depending on the country, this category might be broken down into several categories. For instance, there are countries that distinguish between graduates and post-graduate levels (e.g. Master's and PhDs). However, this distinction is not available in all censuses and, when it is, the available sub-categories within college education are not always comparable. For the sake of comparability, we use the IPUMS-i harmonized version of educational attainment, which provides careful metadata on the criteria used to classify each of the items available in the enumeration forms. As a rule, IPUMS-i groups all types of post-secondary education that lead to a university degree under 'college completed'. Post-secondary technical education is not considered as 'college completed'.

A more relevant issue concerning data comparability is the fact that for some countries, especially those with small populations with college education (as is the case in most African countries), having college or not may not be a relevant boundary in their respective marriage markets. However, for the sake of comparability and with the intention of measuring the impact of our three explanatory factors in various contexts, we purposely include these countries. Identifying which educational thresholds constitute the most important barriers in each marriage market is beyond the scope of this paper. In addition, the use of country specific measures of educational attainment would also oblige to differentiate within college educated populations in those countries in which the majority of the population has attained this level. In doing this, we would miss the intention of this paper, which is to measure the extent to which marriage markets are divided between college and non-college educated couples.

4. Analytical strategy

Since we focus on the population living in union and we only consider two educational attainment groups (those with college and those without college education), we use 2×2 contingency tables with 4 possible combinations to describe marital sorting patterns. The first combination corresponds to couples in which neither member has a college education; their share in the coupled population (or population in union) is denoted by '*a*'. Analogously, '*d*' represents the share of couples in which both partners are college educated; '*b*' represents the share of couples in which men have no college education but women have; and '*c*' represents the opposite combination. Technically speaking, the couples counted in '*a*' and '*d*' are homogamous, whereas the couples counted in '*b*' and '*c*' are heterogamous. More specifically, the couples in '*b*' are hypogamous (i.e. couples where women 'marry down'), and the couples in '*c*' are hypergamous (i.e. where women 'marry up'). These shares will be indistinctly referred to as *educational mating distribution*, or *marital sorting distribution*, and will be briefly denoted as (*a, b, c, d*). Because *a, b, c, d* are shares, their sum is 1. It is important to highlight these four numbers simply describe the observed mating distribution among the population in union, which might follow assortative, random or disassortative patterns (an important aspect associated to the distribution that will be defined below).

4.1. Three factors influencing the education mating distribution

In our approach, any observed educational mating distribution (a, b, c, d) results from the interplay of the following three factors:

(*i*) The *level of college education*, measured as the share of the college-educated population within the coupled population, is denoted by E. Other factors kept constant, high values of E increase the share of college-educated couples, decrease the share of non-college-educated ones, and increase the share of mixed couples. Formally, we have

$$E = \frac{b + c + 2d}{2} \qquad [1]$$

In addition, the share of college-educated women among women in unions is

$$E_w = b + d \qquad [2]$$

and the share of college-educated men among men in unions is

$$E_m = c + d \qquad [3]$$

Clearly, we have

$$E = \frac{E_w + E_m}{2} \qquad [4]$$

(*ii*) The level of *assortative mating* is denoted by H and measures the tendency among individuals to marry within the same educational groups, beyond the structural constraints imposed by the marriage market. *Ceteris paribus*, high levels of assortative mating tend to increase the shares of couples in which both members are either college or non-college educated while decreasing the share of mixed couples.

Formally, the level of assortative mating can be measured with the following indicator[7]:

$$H = ad - bc \qquad [5]$$

(*iii*) The *gender gap in college education* within the coupled population, denoted by *G*, measures the imbalances in the education distribution of women and men. When it is very different from zero, the number of homogamous couples that can be formed diminishes, whereas the potential number of mixed couples increases. Formally, the gender gap in education is defined as

$$G = b - c = E_w - E_m \qquad [6]$$

As shown in equations [1], [5] and [6], each educational mating distribution (*a, b, c, d*) has a corresponding level of college education (*E*), assortative mating (*H*) and gender gap (*G*). Interestingly, the opposite is true. For any specific combination of these three factors, there is one *and only one* educational mating distribution (*a, b, c, d*). This is shown in the following equations:

$$\left.\begin{aligned} a &= \varphi_a(E,H,G) = (1-E)^2 + H - \left(\frac{G}{2}\right)^2 \\ b &= \varphi_b(E,H,G) = E(1-E) - H + \frac{G}{2}\left(\frac{G}{2}+1\right) \\ c &= \varphi_c(E,H,G) = E(1-E) - H + \frac{G}{2}\left(\frac{G}{2}-1\right) \\ d &= \varphi_d(E,H,G) = E^2 + H - \left(\frac{G}{2}\right)^2 \end{aligned}\right\} \qquad [7]$$

[7] This way of measuring the level of assortative mating was suggested in Permanyer et al. (2013). It bears some resemblance to the classical odds ratio parameter that is the basis of log-linear models $\Omega=(a/c)/(b/d)=ad/bc$. Details on why assortative mating is defined in this way are given in Appendix C.

These mathematical identities (whose involved derivation is explained in Appendix A) show how the three factors, *E*, *H*, and *G*, are related to the educational mating distribution. Inspecting them, we corroborate that education expansion favors the increase in college-educated couples and the decrease in non-college-educated ones. In its initial stages (when E<1/2), it favors the increase in mixed couples, but in its later stages (when E>1/2), the opposite happens. Assortative mating (*H*) and the gender gap in education (*G*) have opposing effects: *H* favors the increase in equally educated couples, and *G* fosters the increased in mixed ones. While these relationships are straightforward and go in the direction one should a priori expect, the usefulness of the formal identities shown in [7] comes from the fact that they allow quantifying and measuring *exactly* not only the direction but also the *magnitude* of such relationships – which, otherwise, would be loosely labeled as 'positive' or 'negative'. In addition, these equations are the cornerstone upon which our novel counterfactual-based trend decomposition techniques are based (see section 4.3).

Even if derived through an entirely different procedure, the identities shown in [7] are reminiscent of a saturated 2×2 log-linear model[8] . In that setting, the specification of the different modelling parameters allows predicting *exactly* the observed number of cell counts. In our case, the specification of three variables (*E*, *H* and *G*) also "predicts" *exactly* the shares of each type of couple (i.e. a, b, c, d). Yet, the similarities between both approaches stop here. While the log-linear approach aims at predicting cell counts

[8] In those saturated models, the log of the expected number of cell counts in a 2 × 2 bivariate table is commonly written as $\log(\mu_{ij}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$, where the parameters capture different effects associated to the variables we are working with (see Knoke and Burke 1980).

based on parsimonious models and relies on several methodological assumptions[9], the method suggested in this paper has been specifically designed for the 2×2 setting and aims at quantifying the relative importance of E, H and G to explain changes in the marital sorting distribution (i.e. $\Delta a, \Delta b, \Delta c$ and Δd) and the corresponding polarization levels (see below).

4.2. Measurement of polarization

"Marriage market polarization" (or, simply, "polarization") measures the degree to which married populations split in two groups: college-educated couples and non-college educated ones. In the economics literature, polarization is defined as the grouping of the population into significantly sized clusters such that each cluster has members with similar attributes (e.g. 'the rich' and 'the poor'; see Esteban and Ray 1994). Adapted to our setting, an index of polarization aims to assess how far a given distribution is from a hypothetical scenario in which the population is divided into two equally sized and antagonistic groups (i.e., those with a college education vs. those without a college education). Formally, we work with the following polarization index:

$$P^{\alpha}(a,b,c,d) = 1 - 2^{\alpha-1}\left(\left|\frac{1}{2} - a\right|^{\alpha} + \left|\frac{1}{2} - (a+b+c)\right|^{\alpha}\right) \qquad [8]$$

where α is a non-negative number that can be interpreted as a polarization sensitivity parameter[10]. We use the intermediate value of α = 2, but other values have been investigated with analogous results (which are available from the authors). This index is

[9] For instance, 'Observed frequencies are normally distributed around expected frequencies over repeated samples', or 'The logarithm of the expected value of the response variable is a linear combination of the explanatory variables'.

[10] When $\alpha \rightarrow 0$, the relative contribution of the median category to polarization levels increases, whereas for increasing values of α, the contribution of the median category decreases.

an *ad hoc* adaptation of the ordinal polarization index suggested by Apouey (2007: 885) for the case in which one works with 3 ordered categories. P^{α} is a standard index of polarization that measures the distance between a given distribution (*a, b, c, d*) and the bipolar case (*1/2, 0, 0, 1/2*), where the population is split in two equal-sized groups concentrated at the opposite extremes of the education distribution. P^{α} satisfies the following classical properties expected from a polarization index: (i) P^{α} is bounded between 0 and 1 (i.e. the values associated to minimal and maximal polarization, respectively); (ii) P^{α}(*1/2,0,0,1/2*)=1 (i.e. polarization is maximized in the bipolar case, where half of the couples are college educated and the other half are non-college educated); and (iii) P^{α}(1,0,0,0)= P^{α}(0,1,0,0)= P^{α}(0,0,1,0)= P^{α}(0,0,0,1)=0 (i.e. polarization is minimized when the entire population in union is concentrated in one of the four possible cells and there is no variability).

Polarization levels can thus be either expressed as a function of educational mating distributions (see equation [8]) or as a function of the level of college education, assortative mating and the gender gap (combine equation [7] in equation [8]). Both representations will be used to estimate the impact that these factors have had on polarization levels and its changes over time (see sections 4.3 and 4.4).

4.3. Counterfactual modeling

To assess the impact of E, H and G on educational mating distributions and polarization, we ask what would have happened to the shares a, *b, c, d* and to P^{α} if we held constant two of the three quantities that appear in [7] (E, H and G) at their value in an earlier period of time (t_1) and allowed the third to take a value observed later in time (t_2). In this way, we generate a counterfactual educational mating distribution for the later period of time (t_2). By comparing the observed and the counterfactual patterns, we can assess how important the aforementioned factors are in explaining *changes* in the marital sorting distribution and the corresponding polarization levels[11] (details shown in Appendix B). For that purpose, let us denote the educational mating distributions at times t_1 and t_2 as (a_1, b_1, c_1, d_1) and (a_2, b_2, c_2, d_2), respectively. Likewise, P_1^α and P_2^α denote the observed levels of polarization in times t_1 and t_2. We can now show how changes over time of our key variables can be neatly decomposed in the following three blocs:

$$\Delta a := a_2 - a_1 = \Delta_E a + \Delta_H a + \Delta_G a \qquad [9]$$

$$\Delta b := b_2 - b_1 = \Delta_E b + \Delta_H b + \Delta_G b \qquad [10]$$

$$\Delta c := c_2 - c_1 = \Delta_E c + \Delta_H c + \Delta_G c \qquad [11]$$

$$\Delta d := d_2 - d_1 = \Delta_E d + \Delta_H d + \Delta_G d \qquad [12]$$

$$\Delta P^\alpha := P_2^\alpha - P_1^\alpha \cong \Delta_E P^\alpha + \Delta_H P^\alpha + \Delta_G P^\alpha \qquad [13]$$

where the different $\Delta_E x, \Delta_H x, \Delta_G x$ respectively measure the influence of college education, assortative mating and the gender gap in education when explaining changes in the marital sorting distribution and the corresponding polarization levels (the detailed definitions and formulae are presented in Appendix B). The additive formulas shown in [9] – [13] are the ones we use to measure the contribution of E, H and G when explaining marital sorting and polarization trends over time.

---

[11] Under somewhat stringent conditions, some factors (E, H or G) can be written as a perfect function of the other factors (for instance, when b=c=0, then H=E·(1-E)), so it is formally not possible to assess the effect of one factor 'while holding others constant'. Yet, these special conditions turn out to be so restrictive that they have not been observed in our empirical sample and are unlikely to be observed in practice (e.g. they would require that absolutely *all* couples in a given country should be educationally homogamous) – so they are not affecting the decomposition methods presented in the paper. We are grateful to an anonymous referee for highlighting this point.

It is relevant to emphasize that the methods presented in this paper are based on mathematical decomposition techniques, but they are *not* predictive models as generally understood. Unlike classical predictive modelling techniques (e.g. Ordinary Least Squares or Log-linear models) we are *not* aiming at predicting the estimated value of a dependent variable on the basis of a set of independent variables, and we are neither assuming there is an error term capturing unexplained factors and measurement errors. Likewise, we do not use classical methods like the Oaxaca-Blinder decomposition – which is derived from the OLS framework – because we are not relying on model approximations of the variables of interest. On the contrary, the variables of interest (i.e. educational mating distributions and the corresponding polarization levels) are expressed as an *exact* function of other factors (*E*, *H*, *G*) and, hence, are able to unravel the effects that these factors have had on the changes over time in the 'dependent variable'[12]. The approach followed in this paper is among the commonly used methods in sensitivity analysis – which can be defined as 'the study of how the variation in the critical outcomes of a given system can be categorized and assigned, qualitatively or quantitatively, to different sources of variation in the system' (Saltelli et al., 2008). Our decompositions are purely formal exercises which prove to be extremely useful to test how educational marital sorting and the corresponding polarization levels react to *hypothetical* changes in *E*, *H* and *G*. Indeed, similar ideas have been recently used, albeit in somewhat different contexts, by Breen and Salazar (2010:147, 2011:824), Breen and Andersen (2012:876) and Permanyer et al (2013: 2214).

4.4. Benchmarking exercises

[12] In a way, our approach is like working with an accounting identity (that is: an identity that, by construction, is necessarily true – e.g. 'population change' = 'births' – 'deaths' + 'immigration' – 'outmigration') and assessing the contribution of its different components.

A simple way of assessing the influence of assortative mating is to investigate the extent to which couples' education distribution and the corresponding polarization levels would have been different under alternative levels of assortative mating. We derive the educational mating distribution and the polarization levels that would be observed in the two extreme and hypothetical scenarios of 'absent' (a_0, b_0, c_0, d_0) and 'maximal' $(a_{max}, b_{max}, c_{max}, d_{max})$ assortative mating (see Appendix C). The observed educational mating distribution (a, b, c, d) and the corresponding polarization levels $P^{\alpha}(a, b, c, d)$ are bounded from below and from above by the aforementioned hypothetical education distributions with 'absent' and 'maximal' assortative mating, as observed in the following inequalities (see Appendix C for details):

$$\left.\begin{array}{c} a_0 \leq a \leq a_{max} \\ b_{max} \leq b \leq b_0 \\ c_{max} \leq c \leq c_0 \\ d_0 \leq d \leq d_{max} \\ P_0^{\alpha} \coloneqq P^{\alpha}(a_0, b_0, c_0, d_0) \leq P^{\alpha}(a, b, c, d) \leq P^{\alpha}(a_{max}, b_{max}, c_{max}, d_{max}) \coloneqq P_{max}^{\alpha} \end{array}\right\} \quad [14]$$

To benchmark these variables between the corresponding bounds, we normalize them to a [0,1] scale via the standard transformations

$$\left.\begin{array}{l} a^* = \dfrac{a - a_0}{a_{max} - a_0} \\ b^* = \dfrac{b_0 - b}{b_0 - b_{max}} \\ c^* = \dfrac{c_0 - c}{c_0 - c_{max}} \\ d^* = \dfrac{d - d_0}{d_{max} - d_0} \\ P^* = \dfrac{P^{\alpha} - P_0^{\alpha}}{P_{max}^{\alpha} - P_0^{\alpha}} \end{array}\right\} \quad [15]$$

These indicators are used to assess the potential influence of assortative mating on couples' education distribution and the corresponding polarization levels.

5. Empirical Results

5.1. Descriptive findings

Figure 1 shows the temporal and regional variation of the following key variables: college education, assortative mating, gender gap in education, and polarization levels. Every panel follows the same structure: the variable of interest is represented on the vertical axis; the horizontal axis represents time; each dot represents country observations at a specific point in time; and lines connect dots to show country trends over time. Color lines indicate the geographical region, based on the United Nations continental classification of countries, except for Mexico that was included among the Latin American countries.

FIGURE 1

The top left panel portrays the percentage of people with a college education within the coupled population (*E*) on a 0 to 1 scale. College education has expanded dramatically worldwide over the last five decades, but cross-national differences are large. The percentage of college-educated populations ranges from values below 1% in countries such as Benin, Burkina Faso, Chad, Ethiopia, Guinea, Rwanda and Sierra Leone to values above 40% in Australia, Belgium, Denmark, Portugal, Switzerland and the United Sates in recent years. Time trends show very little growth among African countries. European countries and the United States (North America) show the largest increases in college education (see Table 1).

The level of assortative mating (H), i.e., the tendency to marry within the same educational group, has increased worldwide, with the exception of the African countries, where there is technically no assortative mating (see top right panel in Figure 1). European countries and the United States have the highest levels of assortative mating. Average values of H within Europe moved from 0.07 to 0.1 between 2000 and 2010 (see Table 1), but there is a lot of heterogeneity across countries (see Figure 1 and Table A1 in the online appendix). The patterns and trends of assortative mating are fairly similar to those of the percentages of college-educated people (E). The correlation coefficient between the two variables is 0.91. Assortative mating is highest where the largest percentages of college-educated populations exist and vice versa.

The gap in college education (G) between men and women is shown in the bottom left panel of Figure 1. Negative values indicate that the number of college-educated men is larger than the number of college-educated women. Positive values indicate the opposite. The general trend indicates that the gap between men and women is closing and even reversing in a growing number of countries. By 2010, the gap was completely reversed in all European countries except in Austria, where there was no gap in either direction. By 2000, most Latin American countries had completely closed the gap, with only a few of them lagging behind (e.g., Honduras and Bolivia). By the same year, the gender gap in college education was in favor of men in 12 out of 20 Asian countries for which we have data. In the other 8 countries, the gap was close to 0 or slightly in favor of women (e.g., Jordan, Palestinian Territory). In Africa, the gap is systematically in favor of men (see Table 1 for the regional average levels and Table A1 in the online appendix for country-specific results).

Lastly, on the bottom right panel of Figure 1, we report the level of marriage-market polarization between college-educated and non-college-educated populations, P^{α}, which is bounded between 0 and 1. P^{α} increases in all regions of the world but at markedly different levels and speeds, with Europe at the upper end, Africa at the lower end and Latin America and Asia and Oceania somewhere in between. P^{α} values range from almost 0 in countries such as Malawi, Niger and Tanzania to values above 0.9 in Luxembourg in 2011 (0.94), Ukraine in 2007 (0.92), the United Kingdom in 2011 (0.91) and France in 2011 (0.91). In the early 2000s, polarization levels in Latin American, except Puerto Rico, were between 0.11 (Jamaica 2001) and 0.37 (Cuba 2002). On average, marriage-market polarization in Asia is higher than it is in Latin America, but internal diversity is higher, with values ranging from 0.05 in Vietnam to 0.85 in the Russian Federation (see Table 1 and Table A1 in the online appendix).

Overall, the four panels in Figure 1 are remarkably similar. The correlation coefficients between the level of polarization in the marriage market and the percentage of college-educated people, assortative mating, and the gender gap are 0.97, 0.97 and 0.67, respectively, which implies that the highest levels of polarization are found in countries with the largest shares of college-educated populations and highest assortative mating. However, the question is which of these factors accounts for the most variation in marital sorting patterns and polarization observed across countries and over time. To answer this question, we use the decomposition methods presented in section 4.3.

TABLE 1 ABOUT HERE

5.2. Decomposing changes over time using counterfactual techniques

To investigate the dynamics of change, we focus on change over time within countries. The analysis is restricted to countries with at least two observations in time, but only the most recent period of observation is taken into consideration. On average, these periods are 10.8 years long. This leaves us with 79 different countries covering all regions of the world.

Using the decomposition formulas [9]-[13], we infer the absolute and relative contribution of education expansion, assortative mating and the gender gap in education to the changes over time in marital sorting distributions (a, b, c, d) and the corresponding polarization levels P^{α}. The results are presented in Figures 2 and 3, respectively. Figure 2 is a collection of four ternary plots, each indicating the contribution of E, H and G to changes in the corresponding 'marital sorting shares' (i.e. a, b, c and d). The '+' and '–' signs indicate increases and decreases over time in the corresponding shares. The closer a '+/–' sign is to a given vertex, the more important the corresponding component is in explaining the changes in the corresponding share. For the case of a (shares of non-college educated couples), we can see that (i) they tend to decrease over time, and (ii) such decrease is overwhelmingly explained by the expansion of college education (assortative mating and, particularly, the gender gap in education, play a very limited role; see upper left panel). For the cases of b and c (shares of mixed couples) we observe more increases than decreases over time. In addition, such changes are mostly explained by increasing college education, followed by the gender gap in education (see upper right and lower left panels). Lastly, the generalized increases in college-educated couples can be more attributable to assortative mating than the expansion of college education (with the gender gap in education playing a negligible role; see lower right panel).

FIGURE 2 ABOUT HERE

In Figure 3, the absolute difference between polarization at time t and polarization at time $t + 1$ is represented on the vertical axis. Negative values indicate that polarization has decreased over the observed period, and positive values signal an increase. On the horizontal axis, we represent the 79 observed periods ordered from the smallest to the largest amount of change. Change over time is decomposed into the contribution of E, H and G (see equation [13]). College education is by far the most important factor contributing to the generalized increase in polarization over time, followed distantly by assortative mating. The gender gap in education tends to decrease polarization, but its overall impact is of negligible size. In relative terms (results not shown), the expansion of college education accounts for 94.3% of the change over time in P^{α}, assortative mating accounts for 5%, and the gender gap accounts for less than 1%. The vast majority of countries conform to this pattern.

Despite the concerns that higher assortative mating might have contributed to generate increasingly unequal and diverging societies, we actually observe that the tendency to form homogamous couples has played a very modest role when accounting for the generalized raises in marriage market polarization.

FIGURE 3 ABOUT HERE

5.3. Could extreme changes in assortative mating affect marriage market polarization?

A plausible hypothesis to explain the limited impact of assortative mating on changes in marriage market polarization is that the former might have been too small to have an important effect on the latter. To test this hypothesis, we investigate how polarization levels would look if alternative levels of assortative mating prevailed. We compare the observed educational mating distributions (a, b, c, d) with the ones that would be observed in the *absence* of assortative mating (a_0, b_0, c_0, d_0) and in the case of *maximum* assortative mating $(a_{max}, b_{max}, c_{max}, d_{max})$, and we compute the corresponding P_0^α and P_{max}^α levels. The results are shown in Figure 4. On the vertical axis, it displays the value of polarization. On the horizontal axis, it represents the 258 data observations ordered from the lowest to the highest observed values of P^α. For every country-year observation, we represent the observed level of polarization, P^α, the level of polarization assuming absence of assortative mating, P_0^α, and the level assuming maximum assortative mating, P_{max}^α.

FIGURE 4 ABOUT HERE

By definition, the observed level of polarization is always higher than P_0^α and smaller than P_{max}^α (see Appendix C). $P^\alpha - P_0^\alpha$ measures *in absolute terms* the 'amount of polarization' that is attributable to assortative mating. If we divide $P^\alpha - P_0^\alpha$ by P^α, we obtain a measure in relative terms of the amount of polarization that can be attributable to assortative mating. On average, P^α values would have decreased only 6.1 percentage points had couples been formed at random (see the lower whiskers in Figure 4). This figure varies across regions but not substantially. In Europe and North America, an average of 12 percent of P^α is attributable to assortative mating. In Africa, such contribution is negligible and among Latin American and Asian countries it is below 5% (see Table 1 for the regional average levels and Table A1 in the online appendix for country-specific results).

Finally, we can divide $P^\alpha - P_0^\alpha$ by $P_{max}^\alpha - P_0^\alpha$ to obtain a relative measure of the distance between the observed polarization attributable to assortative mating and the maximum it could have potentially achieved (see equation [15]). As observed in Figure 4, the observed levels of polarization are systematically closer to P^{α}_{max} than to P^{α}_{0}, indicating that assortative mating tends to maximize its potential. Furthermore, it turns out that there are no major variations across regions: all of them approach the global average of 71% (i.e., polarization levels would not have been much higher even if assortative mating had maximized its full potential).

These findings confirm that assortative mating is relatively unimportant when explaining changes in marriage market polarization – which are mostly accounted for by the expansion of college education.

6. Discussion and Concluding Remarks

In this article, we have developed a methodology to disentangle and understand the intertwined effects of three factors that affect the educational mating distributions and the corresponding polarization levels. These factors are the expansion of college education, assortative mating, and the gender gap in college education. We have applied this methodology to data from 118 countries and more than 258 observations to investigate worldwide trends in polarization between low- and high-educated couples.

We have shown that at high levels of college education, marriage markets become more polarized between college-educated and non-college-educated populations, as demonstrated by the growing absolute and relative numbers of college-educated couples in the marriage market. Together with the expansion of college education, the tendency to marry within the same educational group—i.e., assortative mating—is positively related to the size of college-educated people in the marriage market. The changes in marriage market polarization have been mainly driven by the process of education expansion rather than by assortative mating. The gender gap in education has played a negligible role. Therefore, the polarizing trends we observe in the world and its regions are barely influenced by the strengthening of assortative mating but are influenced by the expansion of college education, which is likely to continue expanding in the coming years.

A series of benchmarking and counterfactual exercises have allowed us to investigate the extent to which polarization levels would have been different under different levels of assortative mating. We have found that if couples were formed purely at random, polarization levels would have decreased by an average of only 6.1% of their observed values. This finding shows a rather modest contribution of assortative mating to polarization, despite being at 71% of its maximum potential contribution.

In this context, one might wonder whether the global polarizing trends in terms of education shown in this paper are likely to have an effect on economic inequality around the world. Indeed, several researchers have expressed concerns as regards the current and future implications of increasing assortative mating on patterns of inequality (McLanahan and Percheski 2008, McCall and Percheski 2010, Schwartz 2013). While this is an extremely relevant issue that deserves thorough investigation, we can speculate on the findings that such research might reveal. To begin with, since inequality *between* countries depends on country-level average economic distances and, furthermore, assortative mating is not expected to have a major effect on those averages, we expect a null relationship between the education polarizing trends and inter-country inequality. However, it might have been the case that education-polarizing trends have been partially responsible for the observed increases in income inequality *within* countries. In this regard, there are a few studies explicitly addressing this issue (see Breen and Salazar 2010, 2011, Breen and Andersen 2012, Boertien and Permanyer 2017) but the evidence is so far scarce and inconclusive[13]. Since (i) global expansion of college education is expected to continue in the next decades, and (ii) given its close relationship with education polarization, it is not unlikely that the effects of the latter on within-country inequality will rise over time.

Are the results presented in this paper "unavoidable" given the bounded nature of our variables and the parsimony of our models? Obviously, when no one is educated and when everyone is educated, there is no variability, so polarization is zero. In the process of education expansion, when some population groups receive extra education, there is increasing variability and therefore polarization. However, far less obvious are the following conclusions: (i) assortative mating plays a secondary role in driving the levels of polarization—a result that seems in line with the findings of Breen and Salazar (2011) in the US context, and (ii) the global reversal of the gender gap in higher education has had almost no effect on polarization. From a methodological perspective, we have developed an analytical strategy that neatly analyzes the contribution of the key social forces that drive changes in polarization and provides a broad overview of the macro-level trends that are taking place at the global level.

In conclusion, polarization in the marriage markets inevitably increases because increasing shares of the population have access to and complete college education but

[13] Breen and Salazar (2010, 2011) report a little to null effect of education assortative mating on income inequality in the UK and the US respectively, and Boertien and Permanyer (2017) reach similar conclusions in several high-income countries. On the other hand, the paper by Breen and Andersen (2012) finds that changes in assortative mating actually increased income inequality in Denmark. Currently, there are no analogous studies published for low- and middle-income settings, so the overall relationship between education assortative mating and income inequality remains unclear.

not because of the strengthening of assortative mating. If college-educated people were to become a majority, polarization levels would mechanically go down. A similar effect may have already occurred with basic literacy skills (Permanyer et al. 2013). The literate population was initially very scarce, and then, it began to grow; now, it is almost universal in many places.

These results underscore the importance of structural constraints and understate that of assortative mating. This finding is, of course, not new, as previous research in the US has shown that mating patterns are predominantly driven by the marginal constraints. Since marginal distributions had little sociological interest compared to assortative mating, scholarship focused on the latter. Yet, our findings based on a global perspective comparing societies with markedly different levels of college education suggest that the strength of assortative mating might have been magnified.

References

Atkinson, A., Piketty, T. and Saez, E. 2011. “Top incomes in the long run of history”, *Journal of Economic Literature* 49(1):3-71.

Apouey, B. (2007). Measuring health polarization with self-assessed health data. *Health Economics*, *16*(9), 875-894.

Barro, R. and Lee,J.-W.2013.“A New Data Set of Educational Attainment in the World, 1950-2010.”,*Journal of Development Economics*104:184-198.

Benavot, A. and Riddle, P. 1988. “The Expansion of Primary Education, 1870–1940: Trends and Issues.” *Sociology of Education* 61(3):191-210.

Benavot, A., Cha, Y.-K., Kamens, D. Meyer, J. and Wong, S.-Y. 1991. “Knowledge for the Masses: World Models and National Curricula, 1920-1986.” *American Sociological Review* 56(1):85-100.

Blossfeld, H.-P. 2009. “Educational assortative marriage in comparative perspective.” *Annual Review of Sociology* 35:513-530.

Blossfeld, H.-P.and Timm, A. (Eds.). 2003. *Who marries whom? Educational systems as marriage markets in modern societies*. Dordrecht: Kluwer Academic.

Boertien, D. and Permanyer, I. 2017. “Educational assortative mating as a determinant of changing household income inequality: A 22-country study”, Luxembourg Income Studies Working Paper No. 719.

Breen, R. and Andersen, S. 2012. “Educational assortative mating and income inequality in Denmark.” *Demography* 49:867-887.

Breen, R. and Salazar, L. 2010. “Has increased women’s educational attainment led to greater earning inequality in the United Kingdom? A multivariate decomposition analysis.” *European Sociological Review* 26(2):143-157.

Breen, R. and Salazar, L. 2011. “Educational assortative mating and earnings inequality in the United States.” *American Journal of Sociology* 117(3):808-843.

Cohen, D. and Soto, M. 2007. “Growth and human capital: good data, good results.” *Journal of Economic Growth* 12:51-76.

Crafts, N. 2002. “The Human Development Index, 1870-1999: Some revised estimates.” *European Review of Economic History* 6(3):395-405.

Davis-Kean, P. E. 2005. The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of family psychology*, 19(2), 294.

De Hauw, Y., Grow, A. and Van Bavel, J., 2017. The reversed gender gap in education and assortative mating in Europe. *European Journal of Population*, 33(4), pp.445-474.

DomaŒski, H., & Przybysz, D. 2007. Educational homogamy in 22 European countries. *European Societies*, 9(4), 495-526.

Dorius, S. and Firebaugh, G. 2010. “Trends in global gender inequality.” *Social Forces* 88(5):1941-1968.

Dorius, S. 2013. “The rise and fall of worldwide education inequality from 1870 to 2010: Measurement and trends.” *Sociology of Education* 86(2):158-173.

Esping-Andersen, G. 2009. The incomplete revolution: Adapting welfare states to women's new roles. Polity.

Esping-Andersen, G. 2016. Families in the 21st Century. SNS Förlag, Stockholm.

Esteban, J. and Ray, D. 1994. “On the Measurement of Polarization”, *Econometrica* 62:819–851.

Esteban, J. and Ray, D. 2011. “Linking Conflict to Inequality and Polarization”, *American Economic Review*, 101:1345-1374.

Esteban, J., Mayoral, L. and Ray, D. 2012. “Ethnicity and Conflict: Theory and Facts”, *Science,* 336 (6083): 858-865..

Esteve, A., and Sobek, M. 2003. “Challenges and methods of international census harmonization.”*Historical Methods: A Journal of Quantitative and Interdisciplinary History* 36(2):66-79.

Esteve, A., Garcia, J. and Permanyer, I. 2012. “The gender-gap reversal in education and its effect in union formation: The end of hypergamy?” *Population and Development Review* 38(3):535-546.

Esteve, A., & McCaa, R. 2007. Homogamia educacional en México y Brasil, 1970-2000: pautas y tendencias. *Latin American Research Review*, 56-85.

Esteve, A., Schwartz, C.R., Van Bavel, J., Permanyer, I., Klesment, M. and Garcia, J., 2016. The end of hypergamy: Global trends and implications. *Population and development review*, 42(4), p.615.

Grant, M. and Behrman, J. 2010. “Gender gaps in educational attainment in less developed countries.” *Population and Development Review* 36(1):71-89.

Grow, A. and Van Bavel, J., 2015. Assortative mating and the reversal of gender inequality in education in Europe: An agent-based model. *PloS one*, 10(6), p.e0127806.

Guryan, J., Hurst, E., & Kearney, M. 2008. Parental education and parental time with children. *Journal of Economic perspectives*, 22(3), 23-46.

Harmon, C., Walker, I., and Westergærd-Nielsen, N. C. (Eds.). 2001. Education and earnings in Europe: a cross-country analysis of the returns to education. Edward Elgar Publishing.

Hout, M. 2012. Social and economic returns to college education in the United States. *Annual Review of Sociology*, 38, 379-400.

Kalmijn, M. 1998. “Intermarriage and homogamy: Causes, patterns, trends.” *Annual review of sociology*, 395-421.

Kalmijn, M. 2013. “The educational gradient in Marriage: A comparison of 25 European countries”, *Demography* 50(4):1499-1520.

KC, S. et al. 2010. “Projection of population by level of educational attainment, age, and sex for 120 countries for 2005–2050.”*Demographic Research* 22:383-472.

Knoke, D. and Burke, P. 1980. *Log-linear Models*. Sage University Paper 20. Series: Quantitative Applications in the Social Sciences.

Lamerz, A., Kuepper-Nybelen, J., Wehle, C., Bruning, N., Trost-Brinkhues, G., Brenner, H., ... & Herpertz-Dahlmann, B. 2005. Social class, parental education, and obesity prevalence in a study of six-year-old children in Germany. *International journal of obesity*, 29(4), 373.

Lutz, W. and KC, S. 2011. “Global human capital: Integrating education and population.” *Science* 333:587-592.

Lutz, W., Butz, W. and KC, S. (Eds.). 2014. *World Population and Human Capital in the Twenty-First Century*. Oxford: Oxford University Press.

McCall, L. and Percheski, C. 2010. “Income Inequality: New trends and research directions”, *Annual Review of Sociology* 36:329-347.

McLanahan, S. and Percheski, C. 2008. “Family structure and the reproduction of inequalities”, *Annual Review of Sociology* 34:257-276.

McLanahan, S. 2004. Diverging destinies: How children are faring under the second demographic transition. Demography, 41(4), 607-627.

McPherson, M., Smith-Lovin, L., & Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1), 415-444.

Mare, R. 1991. "Five decades of educational assortative mating." *American Sociological Review* 56:15-32.

Meyer, J., Ramirez, F., Rubinson, R. and Boli-Bennett, J. 1977. "The world education revolution, 1950-1970." *Sociology of Education* 4:242-258.

Meyer, J., Ramirez, F. and Soysal, Y. 1992. "World Expansion of Mass Education, 1870-1980." *Sociology of Education* 65(2):128-149.

Minnesota Population Center. 2014. Integrated Public Use Microdata Series, International: Version 6.3 [Machine-readable database]. Minneapolis: University of Minnesota.

Montalvo, J. and Reynal-Querol, M. 2005. "Ethnic Polarization, Potential Conflict, and Civil Wars", *American Economic Review*, 95(3):796-816.

Morrison, C. and Murtin, F. 2009."The century of education", *Journal of Human Capital* 3(1):1-42.

Permanyer, I., García, J., and Esteve, A. 2013. "The impact of educational homogamy on isolated illiteracy levels" *Demography* 50 (6):2209-2225.

Potarca, G. 2017. Does the internet affect assortative mating? Evidence from the US and Germany. *Social science research*, 61, 278-297.

Ramirez, F. and Meyer, J. 1980. "The Social Construction of the Modern World System", *Annual Review of Sociology* 6:369-99.

Reardon, S. F. 2011. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. Whither opportunity, 91-116.

Schwartz, C. and Mare, R. 2012. “The Proximate Determinants of Educational Homogamy: The Effects of First Marriage, Marital Dissolution, Remarriage, and Educational Upgrading.” *Demography* 49:629-650.

Schwartz, C. 2013. “Trends and variations in assortative mating: causes and consequences.” *Annual Review of Sociology* 39:451-470.

Schwartz, C. and Mare, R. 2005.“Trends in educational assortative mating from 1940 to 2003.”*Demography* 49:629-650.

Smits, J. 2003. Social closure among the higher educated: Trends in educational homogamy in 55 countries. *Social Science Research*, 32(2), 251-277.

Smits, J., Ultee, W., and Lammers, J. (1998). Educational homogamy in 65 countries: An explanation of differences in openness using country-level explanatory variables. *American Sociological Review*, 264-285.

Smits, J., Ultee, W., and Lammers, J. 2000.“More or less educational homogamy?A test of different versions of modernization theory using cross-temporal evidence for 60 countries”, *American Sociological Review* 65:781–788.

Smits, J., and Park, H. (2009). Five decades of educational assortative mating in 10 East Asian societies. *Social Forces*, *88*(1), 227-255.

Appendix A.

The derivation of [7] is long and involved; it is explained in the following steps.

Step 1. Write a, b, c and d in terms of E_m, E_w and H (see equations [2], [3] and [5] for definitions). This involves solving the following equations system:

$$\left.\begin{array}{c} a + b = 1 - E_m \\ c + d = E_m \\ a + c = 1 - E_w \\ b + d = E_w \\ ad - bc = H \end{array}\right\} \qquad [A1]$$

Solving [A1], we obtain

$$\left.\begin{array}{c} a = (1 - E_w)(1 - E_m) + H \\ b = E_w(1 - E_m) - H \\ c = E_m(1 - E_w) - H \\ d = E_w E_m + H \end{array}\right\} \qquad [A2]$$

Therefore, *any* education distribution $(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d})$ with gender equality (i.e., $E_w = E_m = E$) can be written as

$$\left.\begin{array}{c} \tilde{a} = (1 - E)^2 + H \\ \tilde{b} = E(1 - E) - H \\ \tilde{c} = E(1 - E) - H \\ \tilde{d} = E^2 + H \end{array}\right\} \qquad [A3]$$

Step 2. Starting from the education distribution (a, b, c, d), we derive another distribution (a_g, b_g, c_g, d_g) with the same marginals and the same level of assortative mating as the original one but with no gender gap in education. For that purpose, we need to solve the following equations system:

$$\left.\begin{array}{c} a_g + b_g = a + b \\ c_g + d_g = c + d \\ a_g + c_g = a + c \\ b_g + d_g = b + d \\ a_g d_g - b_g c_g = ad - bc \\ b_g = c_g \end{array}\right\} \qquad [A4]$$

Solving [A4], we obtain

$$\left.\begin{aligned} a_g &= a + \left(\frac{G}{2}\right)^2 \\ b_g &= \frac{b+c}{2} - \left(\frac{G}{2}\right)^2 \\ c_g &= \frac{b+c}{2} - \left(\frac{G}{2}\right)^2 \\ d_g &= d + \left(\frac{G}{2}\right)^2 \end{aligned}\right\} \qquad [A5]$$

Step 3. Because [A5] is obtained after imposing gender equality, (a_g, b_g, c_g, d_g) can be considered a particular case of $(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d})$. From [A3] and [A5], we can deduce that

$$\left.\begin{aligned} (1-E)^2 + H &= a + \left(\frac{G}{2}\right)^2 \\ E(1-E) - H &= \frac{b+c}{2} - \left(\frac{G}{2}\right)^2 \\ E^2 + H &= d + \left(\frac{G}{2}\right)^2 \end{aligned}\right\} \qquad [A6]$$

Lastly, the identities in [7] are obtained after basic algebraic manipulations of [A6].

Appendix B

In this appendix, we show how the decomposition formulas [9], [10], [11], [12] and [13] are arrived at. For that purpose, we introduce the following notations. Let the college education levels, the level of assortative mating and the gender gap in college education observed at times t_1 and t_2 be denoted as E_1, H_1, G_1 and $E_2, H_2, G_2,$ respectively. We define

$$\left.\begin{aligned} a_2^E &= \varphi_a(E_2, H_1, G_1) \\ a_2^H &= \varphi_a(E_1, H_2, G_1) \\ a_2^G &= \varphi_a(E_1, H_1, G_2) \end{aligned}\right\} \qquad [B1]$$

as the counterfactual shares of non-college-educated couples that would be observed in t_2 if we only changed over time *E*, *H* or *G*, while keeping the other two factors constant (see equation [7] for the definition of $\varphi_a(.,.,.)$). Analogously, we can define the counterfactual shares of college-educated couples d_2^E, d_2^H, d_2^G and heterogamous couples $b_2^E, b_2^H, b_2^G, c_2^E, c_2^H, c_2^G$.

$$\left.\begin{aligned} d_2^E &= \varphi_d(E_2, H_1, G_1) \\ d_2^H &= \varphi_d(E_1, H_2, G_1) \\ d_2^G &= \varphi_d(E_1, H_1, G_2) \end{aligned}\right\} \qquad [B2]$$

$$\left.\begin{aligned} b_2^E &= \varphi_b(E_2, H_1, G_1) \\ b_2^H &= \varphi_b(E_1, H_2, G_1) \\ b_2^G &= \varphi_b(E_1, H_1, G_2) \end{aligned}\right\} \qquad [B3]$$

$$\left.\begin{aligned} c_2^E &= \varphi_c(E_2, H_1, G_1) \\ c_2^H &= \varphi_c(E_1, H_2, G_1) \\ c_2^G &= \varphi_c(E_1, H_1, G_2) \end{aligned}\right\} \qquad [B4]$$

Now, we can define the following quantities:

$$\left.\begin{aligned} \Delta_E a &= a_2^E - a_1;\ \Delta_E b = b_2^E - b_1;\ \Delta_E c = c_2^E - c_1;\ \Delta_E d = d_2^E - d_1 \\ \Delta_H a &= a_2^H - a_1;\ \Delta_H b = b_2^H - b_1;\ \Delta_H c = c_2^H - c_1;\ \Delta_H d = d_2^H - d_1 \\ \Delta_G a &= a_2^G - a_1;\ \Delta_G b = b_2^G - b_1;\ \Delta_G c = c_2^G - c_1;\ \Delta_G d = d_2^G - d_1 \end{aligned}\right\} \qquad [B5]$$

Thus, $\Delta_E a$ measures the difference between the share of non-college educated couples that would be observed in t_2 if only college education had changed over time and the share of non-college educated couples observed in t_1. In this way, $\Delta_E a$ can be interpreted as the influence of education expansion to the changes in the share of non-college educated couples. The remaining quantities defined in equation [B5] are defined in an analogous way. With these definitions and using the identities shown in [7], after simple algebraic manipulations one has that

$$\Delta_E a + \Delta_H a + \Delta_G a = (a_2^E - a_1) + (a_2^H - a_1) + (a_2^G - a_1) =$$

$$= \big(\varphi_a(E_2, H_1, G_1) - \varphi_a(E_1, H_1, G_1)\big) + \big(\varphi_a(E_1, H_2, G_1) - \varphi_a(E_1, H_1, G_1)\big)$$
$$+ \big(\varphi_a(E_1, H_1, G_2) - \varphi_a(E_1, H_1, G_1)\big) =$$

$$= \left(\left((1 - E_2)^2 + H_1 - \left(\frac{G_1}{2}\right)^2\right) - \left((1 - E_1)^2 + H_1 - \left(\frac{G_1}{2}\right)^2\right)\right)$$
$$+ \left(\left((1 - E_1)^2 + H_2 - \left(\frac{G_1}{2}\right)^2\right) - \left((1 - E_1)^2 + H_1 - \left(\frac{G_1}{2}\right)^2\right)\right)$$
$$+ \left(\left((1 - E_1)^2 + H_1 - \left(\frac{G_2}{2}\right)^2\right) - \left((1 - E_1)^2 + H_1 - \left(\frac{G_1}{2}\right)^2\right)\right)$$

$$= ((1 - E_2)^2 - (1 - E_1)^2) + (H_2 - H_1) + \left(\left(\frac{G_1}{2}\right)^2 - \left(\frac{G_2}{2}\right)^2\right) =$$

$$\left((1 - E_2)^2 + H_2 - \left(\frac{G_2}{2}\right)^2\right) - \left((1 - E_1)^2 + H_1 - \left(\frac{G_1}{2}\right)^2\right) = a_2 - a_1$$

This shows how the exact decomposition formula shown in [9] is arrived at. Following exactly the same techniques, it is straightforward to derive the decomposition formulas [10], [11] and [12]. For the decomposition formula [13], we introduce the following notation:

$$\left.\begin{aligned} \Delta_E P^\alpha &= P^\alpha(a_2^E, b_2^E, c_2^E, d_2^E) - P^\alpha(a_1, b_1, c_1, d_1) \\ \Delta_H P^\alpha &= P^\alpha(a_2^H, b_2^H, c_2^H, d_2^H) - P^\alpha(a_1, b_1, c_1, d_1) \\ \Delta_G P^\alpha &= P^\alpha(a_2^G, b_2^G, c_2^G, d_2^G) - P^\alpha(a_1, b_1, c_1, d_1) \end{aligned}\right\} \qquad [B6]$$

Thus, $\Delta_E P^\alpha$ measures the change in polarization levels that would be observed if only education had changed over time (i.e. leaving assortative mating and the gender gap in education at their t_1 levels). The quantities $\Delta_H P^\alpha$ and $\Delta_G P^\alpha$ are defined analogously. It turns out that the decomposition formula

$$\Delta P^\alpha := P_2^\alpha - P_1^\alpha \cong \Delta_E P^\alpha + \Delta_H P^\alpha + \Delta_G P^\alpha \qquad [B7]$$

is "quasi-exact": the changes in polarization over time (ΔP^α) do not coincide exactly with the sum of the contributions of E, H and G to changes in polarization, but they are extremely similar. To prove this assertion, the scatterplot shown below compares for the samples included in our paper the values of $P_2^\alpha - P_1^\alpha$ on the horizontal axis versus the values of $\Delta_E P^\alpha + \Delta_H P^\alpha + \Delta_G P^\alpha$ in the vertical one. As can be seen, the relation is almost perfectly linear (the correlation coefficient between both variables equals 0.998), and the approximation error is extremely small. Therefore, it makes perfect sense to say that changes in polarization can be decomposed in the contributions of E, H and G, respectively, using equation [13].

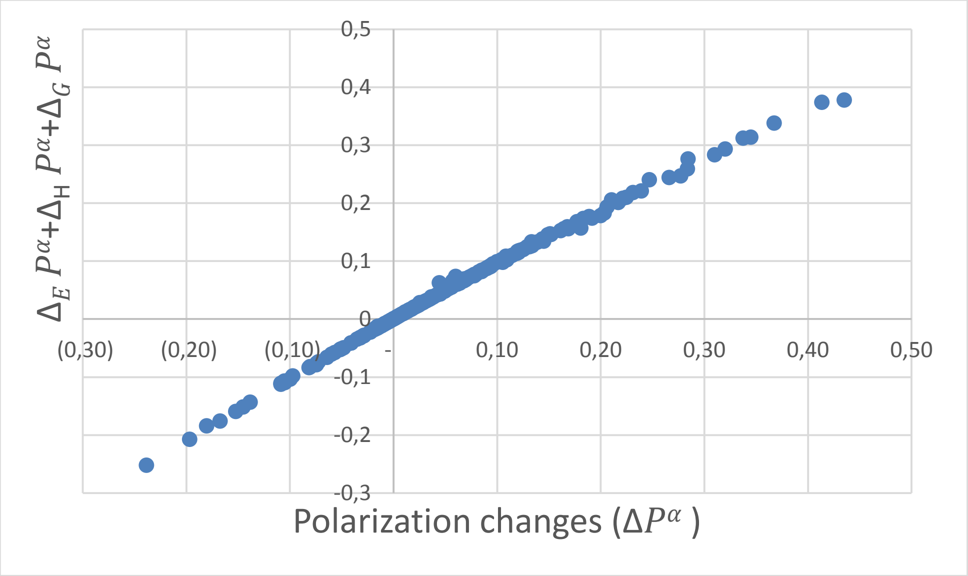


Figure B1. Scatterplot comparing the values of $P_2^\alpha - P_1^\alpha$ (horizontal axis) and the values of $\Delta_E P^\alpha + \Delta_H P^\alpha + \Delta_G P^\alpha$ (vertical axis) for the set of 79 countries considered in this paper for which we have at least two observations over time. Authors' calculations based on IPUMS, DHS, EU-LFS, GGS and EU-SILC data.

Appendix C.

In this appendix, we show how to obtain the hypothetical education distributions (a_0, b_0, c_0, d_0) and $(a_{max}, b_{max}, c_{max}, d_{max})$ that would be observed under extreme assortative mating assumptions.

A simple way of measuring assortative mating is to compare the observed educational mating distribution (a, b, c, d) with the hypothetical distribution (a_0, b_0, c_0, d_0) that would be observed if individuals did not care about their partners' education (i.e., if couples were formed purely at random) while keeping the marginal education distribution of women and men unchanged. It is well known that under such an independence assumption, one has

$$a_0 = (a + b)(a + c); b_0 = (a + b)(b + d); c_0 = (c + d)(a + c); d_0 = (c + d)(b + d) \quad [C1]$$

Because these are the expected frequencies that would be observed if partners' education played no role in the process of union formation, the difference between observed and expected values could be interpreted as measuring assortative mating. These differences will be labeled as

$$a_p = a - a_0; b_p = b - b_0; c_p = c - c_0;\ d_p = d - d_0 \qquad [C2]$$

As shown in Permanyer et al. (2013), one has

$$a_p = d_p = ad - bc \qquad [C3]$$

$$b_p = c_p = bc - ad \qquad [C4]$$

Therefore, if one defines $H = ad - bc$, then any educational mating distribution (a, b, c, d) can be rewritten as

$$\left.\begin{aligned} a &= a_0 + H \\ b &= b_0 - H \\ c &= c_0 - H \\ d &= d_0 + H \end{aligned}\right\} \qquad [C5]$$

Equation [C5] shows a decomposition of observed cell frequencies as a sum of frequencies that would be observed if education status were irrelevant for couples' formation plus a term H that could be interpreted as the level of assortative mating. Positive values of H indicate that in the population under study, there is a tendency toward assortative mating (indeed, this is the case for all observations in our sample.)

Thus far, we have compared the education distribution shares with a hypothetical education distribution that results from assuming the absence of a relationship between education status and couples' formation. A conceptually related but somewhat different way of approaching the same problem is to attempt to answer the following question: to what extent would the education distribution shares be different if maximal assortative mating patterns prevailed? It is straightforward to verify that *when the marginal education distributions of women and men are fixed*, the distribution that maximizes assortative mating is the one that concentrates the maximum number of couples in the main diagonal of the couples' education distribution table:

	Non-college Woman	College Woman	Total
Non-college Man	$a + \min\{b, c\}$	$b - \min\{b, c\}$	$a + b$

College Man	$c - \min\{b, c\}$	$d + \min\{b, c\}$	$c + d$
Total	$a + c$	$b + d$	1

Therefore, we define $(a_{max}, b_{max}, c_{max}, d_{max})$ as

$$\left.\begin{aligned} a_{max} &= a + \min\{b, c\} \\ b_{max} &= b - \min\{b, c\} \\ c_{max} &= c - \min\{b, c\} \\ d_{max} &= d + \min\{b, c\} \end{aligned}\right\} \qquad [C6]$$

Under the assumption that $H \geq 0$ (a restriction that holds true in all our empirical observations), using equations [C5] and [C6] it is straightforward to prove the validity of the several inequalities shown in [14], that is:

$$a_0 \leq a \leq a_{max};\ b_{max} \leq b \leq b_0;\ c_{max} \leq c \leq c_0;\ d_0 \leq d \leq d_{max}; P_0^{\alpha} \leq P^{\alpha}(a, b, c, d) \leq P_{max}^{\alpha}$$

With these definitions, it is straightforward to check that the first four identities in [15] are indeed the same:

$$\frac{a - a_0}{a_{max} - a_0} = \frac{b_0 - b}{b_0 - b_{max}} = \frac{c_0 - c}{c_0 - c_{max}} = \frac{d - d_0}{d_{max} - d_0} = \frac{H}{H + \min\{b, c\}} = m^* \qquad [C7]$$

Figures

Figure 1. Percentage of college-educated people (E), assortative mating (H), gender gap in college education (G) and observed polarization P^{α} across 258 observations spanning 118 countries since the 1960s. Authors' calculations based on IPUMS, DHS, EU-LFS, GGS and EU-SILC data.

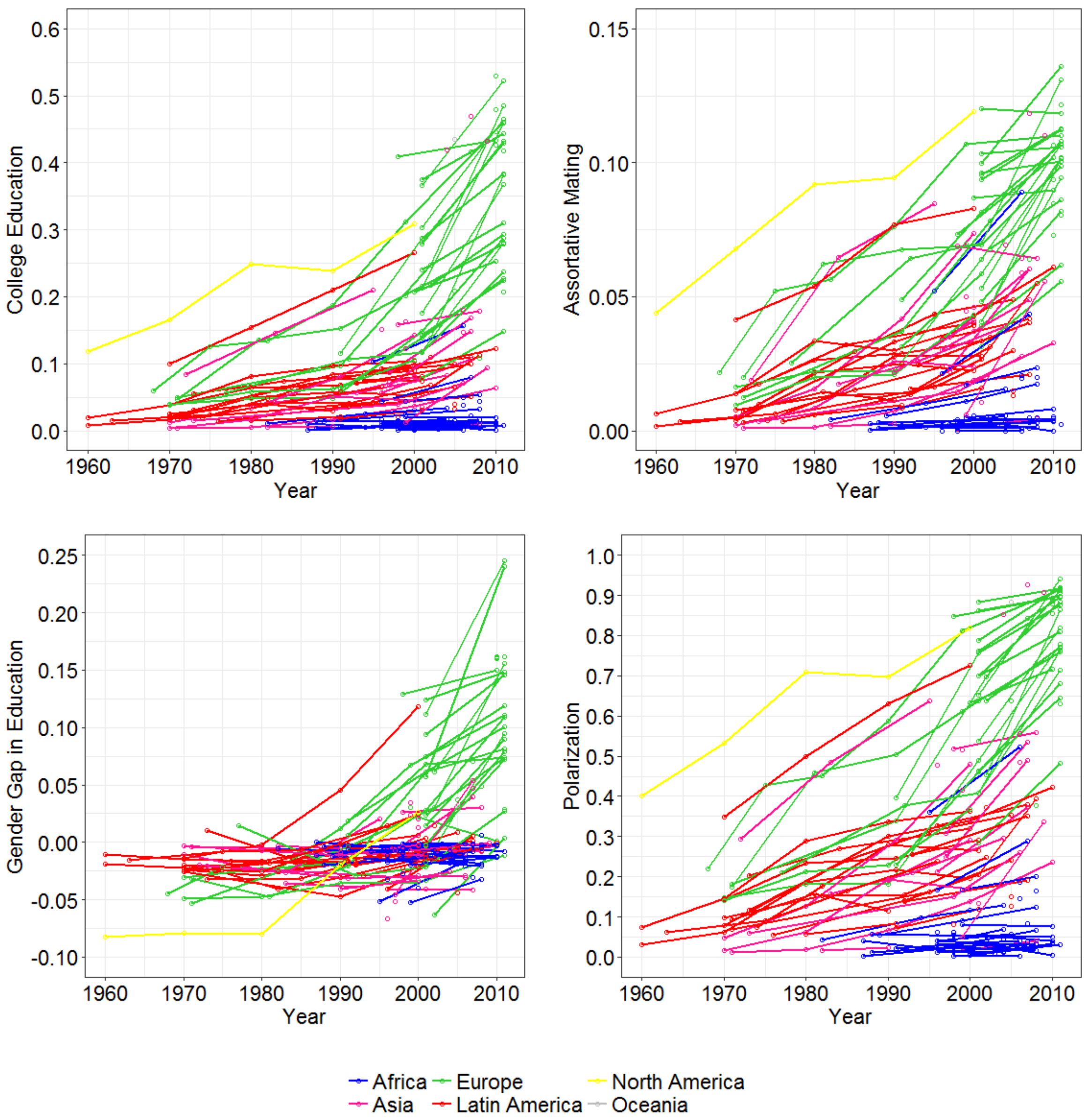


Figure 2. Contribution of the college education expansion (*E*), assortative mating (*H*) and the gender gap in education (*G*) to changes over time in marital sorting distributions, in 79 countries ('+' and '–' signs indicate increases and decreases in the corresponding couple education combination, respectively). The three numbers within each triangle indicate the percentage of observations where *E*, *H* or *G* are the most important contributing factors, respectively. The blue dot indicates the average across observations. Authors' calculations based on IPUMS, DHS, EU-LFS, GGS and EU-SILC data.

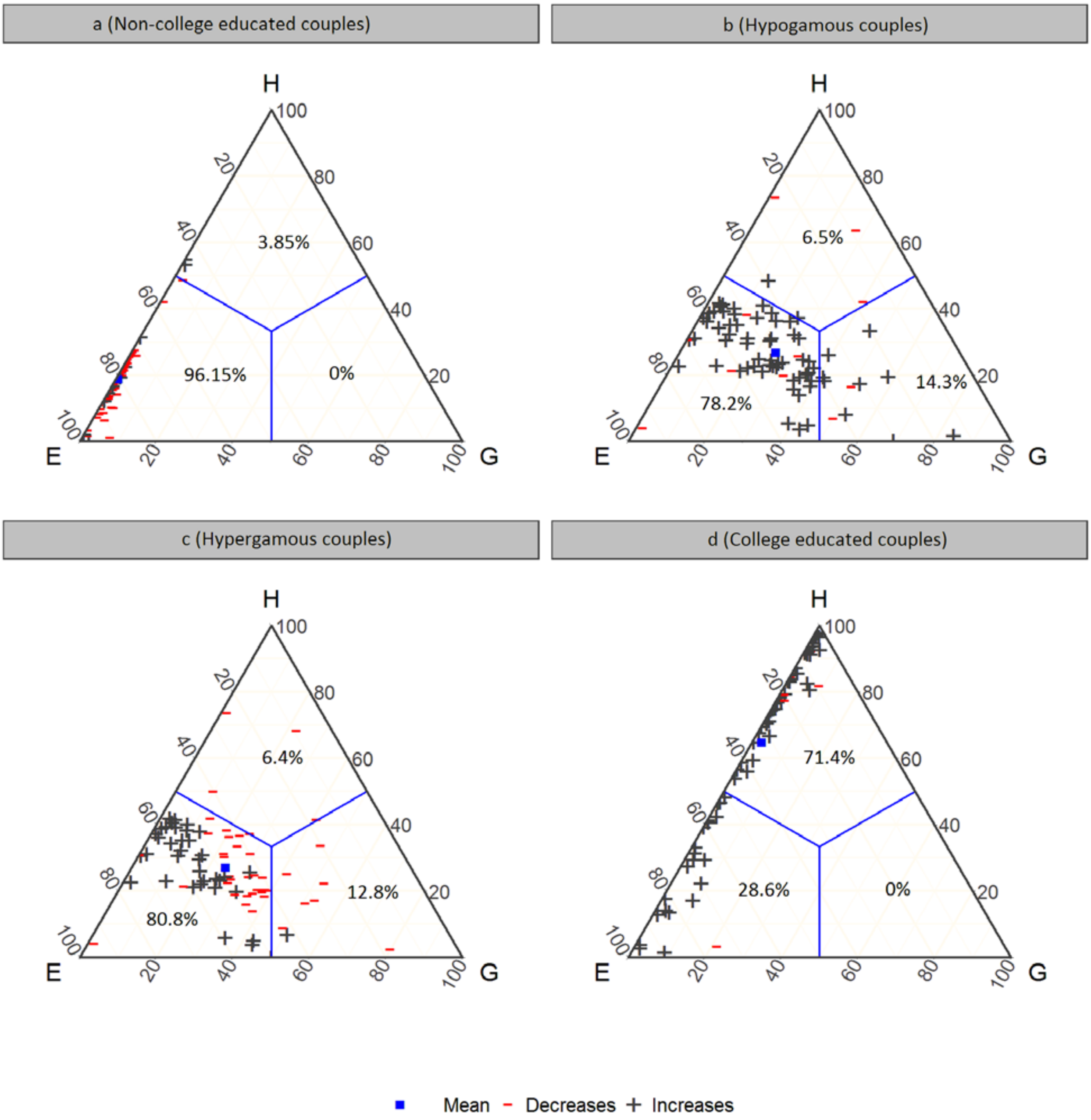


Figure 3. Contribution of the college education expansion (E), assortative mating (H) and the gender gap in education (G) to changes over time in the marriage market polarization between college and non-college educated populations, $P_{t+1}^{\alpha} - P_{t}^{\alpha}$, in 79 countries. Authors' calculations based on IPUMS, DHS, EU-LFS, GGS and EU-SILC data.

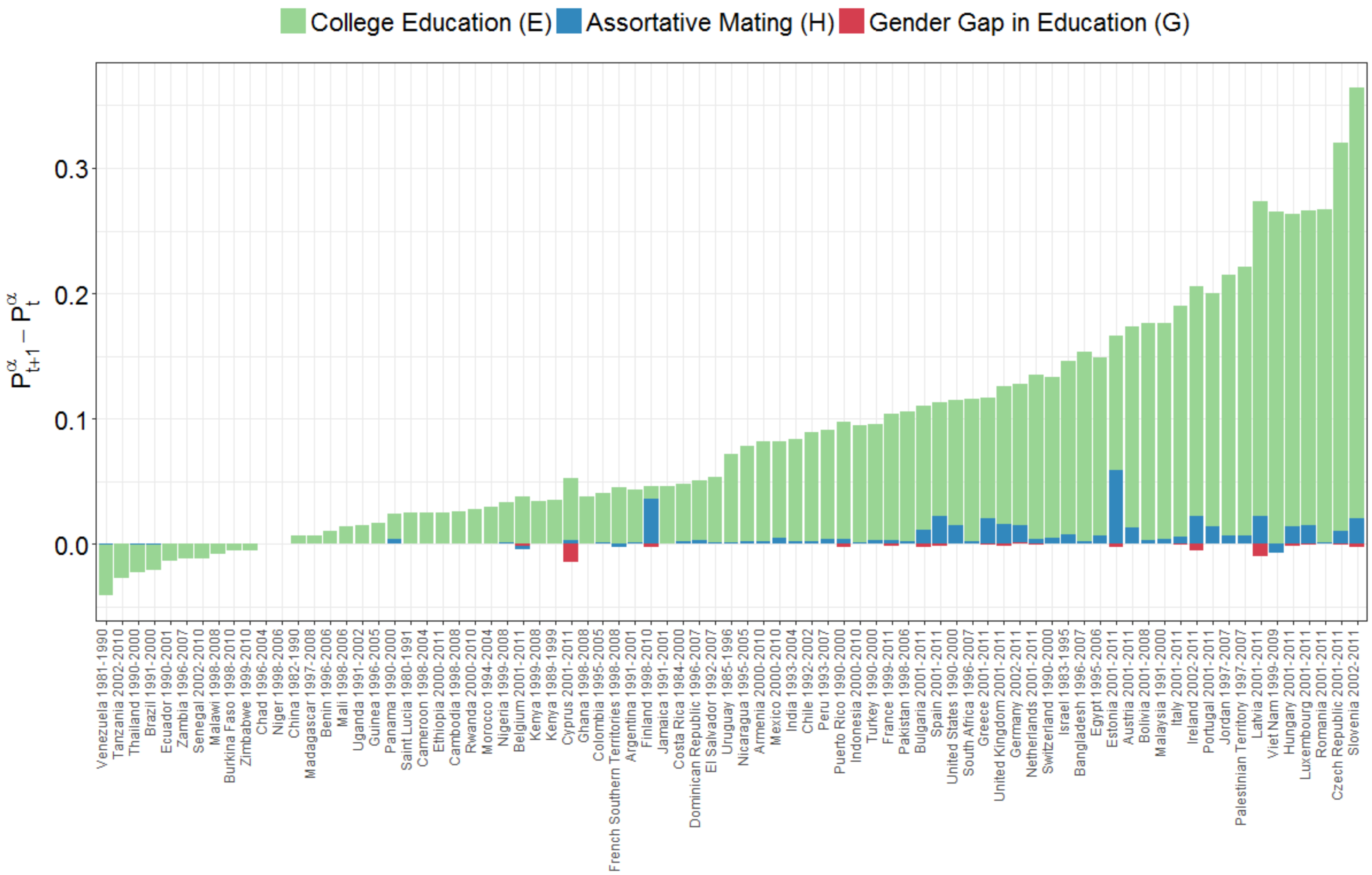
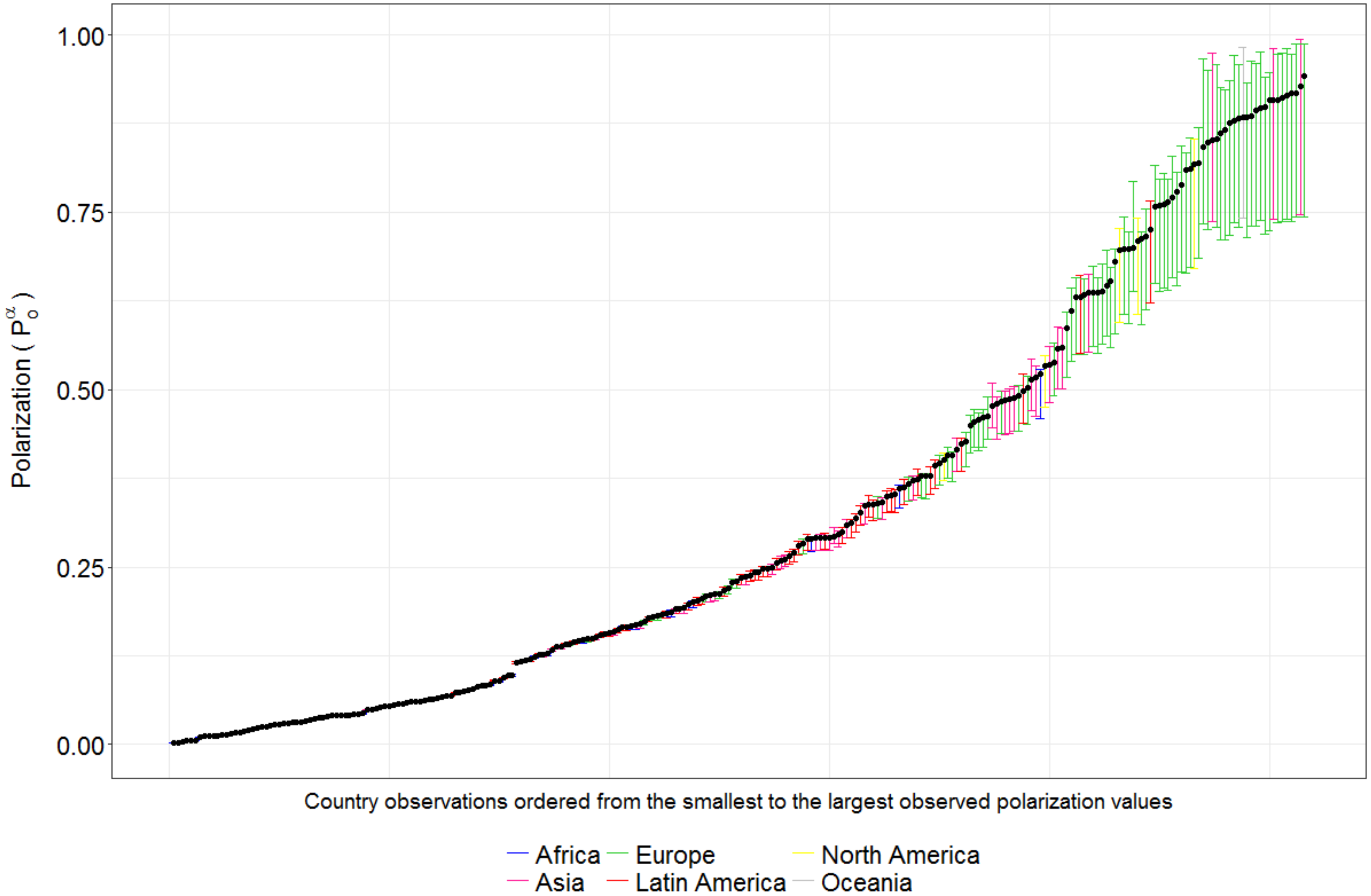


Figure 4. Levels of polarization in the marriage market between college and non-college educated populations according to observed values, P^{α}, and two counterfactual scenarios—absence of assortative mating, P_0^{α}, and maximum assortative mating, P_{max}^{α}—across 258 observations spanning 118 countries. Authors' calculations based on IPUMS, DHS, EU-LFS, GGS and EU-SILC data.



Tables

Table 1. College education, assortative mating, gender gap in education and polarization levels by regions over time.

Region	Period	Countries	Samples	E	H	G	P	Pmax	PzeroH
Africa	Before 1995	8	9	0,010	0,003	-0,008	0,039	0,039	0,038
	1995-2004	25	27	0,017	0,006	-0,014	0,063	0,063	0,061
	2005 and later	22	22	0,027	0,013	-0,011	0,101	0,101	0,096
	Total	32	**58**	**0,020**	**0,008**	**-0,012**	**0,073**	**0,074**	**0,070**
Asia & Oceania	Before 1995	8	17	0,037	0,016	-0,017	0,134	0,136	0,128
	1995-2004	23	23	0,095	0,034	-0,013	0,300	0,312	0,276
	2005 and later	13	13	0,180	0,057	0,007	0,467	0,493	0,408
	Total	28	**53**	**0,097**	**0,034**	**-0,010**	**0,288**	**0,300**	**0,261**
Europe & US	Before 1995	9	23	0,104	0,043	-0,030	0,347	0,356	0,317
	1995-2004	25	25	0,225	0,074	0,036	0,627	0,663	0,545
	2005 and later	32	32	0,351	0,095	0,100	0,792	0,843	0,660
	Total	35	**80**	**0,241**	**0,074**	**0,043**	**0,613**	**0,647**	**0,525**
LAC	Before 1995	17	40	0,050	0,018	-0,017	0,179	0,183	0,170
	1995-2004	17	17	0,079	0,031	0,001	0,272	0,279	0,254
	2005 and later	10	10	0,075	0,033	-0,006	0,266	0,271	0,249
	Total	23	**67**	**0,061**	**0,024**	**-0,011**	**0,215**	**0,220**	**0,203**
World	Before 1995	42	89	0,057	0.023	-0.019	0.200	0.204	0.187
	1995-2004	90	92	0,104	0.036	0.003	0.314	0.328	0.282
	2005 and later	77	77	0,194	0.057	0.039	0.471	0.498	0.403
	Total	118	**258**	**0,115**	**0.038**	**0.006**	**0.321**	**0.336**	**0.285**

Authors' calculations based on IPUMS, DHS, EU-LFS, GGS and EU-SILC data.