Rev Econ Household (2022) 20:161–198
https://doi.org/10.1007/s11150-020-09527-6

On the measurement of non-random mating and of its change over time

Jacques Silber[1] · **Sasiwimon Warunsiri Paweenawat**[2] · **Lusi Liao**[2]

Received: 7 March 2020 / Accepted: 6 November 2020 / Published online: 8 January 2021

Abstract
This paper proposes a new approach to the measurement of assortative mating and of the change over time in assortative mating. Non-assortative mating is viewed as independence between the characteristics of the husband and those of the wife. In our paper the characteristic we focus on is the educational level of the spouses. In measuring the change in assortative mating we use an algorithm that allows one to make a distinction between changes in the distribution of husbands and wives by educational level and a "pure change in assortative mating" that is the consequence of a change in the degree of independence between the educational levels of husbands and wives. We present an illustration of our approach, based on data for Thailand covering the period 1985–2019. It appears that while over the whole period 1985–2019 the increase in the Theil index of non-random mating was uniquely due to a change in the educational composition of the males and females (essentially of the female population), there are several sub-periods where the "pure change in assortative mating" played an important role.

Keywords Assortative mating · Education · Marriage · Thailand

JEL classification J12 · D13 · I24

1 Introduction

The study of assortative mating is in a way quite similar to that of inter- or intra-generational mobility or rather immobility. In the case of inter-generational educational mobility, for example, one often analyzes the features of a matrix whose lines correspond to the educational attainment of the parents and the columns to those of

✉ Jacques Silber
jsilber_2000@yahoo.com

[1] Bar-Ilan University, Ramat Gan, Israel

[2] University of the Thai Chamber of Commerce, Bangkok, Thailand

the children. There are then two possibilities to measure the extent of intergenerational mobility (see, Fields and Ok 1999; Fields 2008). Either one adopts a "movement approach" so that if all the observations are on the diagonal (we assume, for simplicity that the educational categories are the same for the parents and the children), mobility will be nil. Or one takes a "time independence" approach and assumes that a perfectly mobile society is a society in which the probability of having a specific educational level is independent of that of one's parents, so that the rows of this matrix, which reflect conditional probabilities, would be identical.

The same two approaches may be taken to analyze the extent of assortative mating. We can first say that there is perfect assortative mating when each man is matched with a woman with the same socio-economic characteristics, like the same educational level. Assuming the educational categories are the same for husbands and wives, we will conclude in such a case that the further away the observations are from the diagonal of the matrix of the spouses educational levels, the less assortative mating there is. One can then use various measures to measure the extent of such "movement", like the mobility indices proposed by Prais (1955) and Bibby (1975, 1980). But, as indicated previously, there is another possible approach to measuring non assortative mating, one where one compares the actual probability for a man with characteristic i to get married a woman with characteristic j, with what this probability would have been, had there been complete independence between the characteristics of the men and those of the women. If, for instance, the characteristics are the educational levels, the latter probability would be equal to the product of the share among males of those with educational level i times the share among females of those with education level j. It is this product that should be compared with the actual share ($n_{ij}/\Sigma_i\Sigma_j n_{ij}$) in the total number of couples n_{ij} where the husband has educational level i and the wife educational level j. To draw general conclusions for the population concerning the extent of assortative mating, one evidently has to take into account all the levels of education i and j.

But a second issue has to be taken into account. It concerns the impact of changes over time in the educational composition of the male and female populations. An increase or decrease in the percentage of spouses having the same level of education does not necessary imply that there was an increase or decrease in what we could call "pure assortative mating", because such changes may be mainly the consequence of variations in the educational composition of the male and female populations. It should be clear, for example, that if there was over time a significant increase in the share of women with a high level of education, such a change is likely to increase the probability that males with a high level of education will get married to females with a high level of education. We therefore argue that a "pure measure of assortative mating" should neutralize such a variation over time in the educational composition of the male and female populations and we propose a technique allowing such a neutralization.

The paper is organized as follows. Section 2 shortly summarizes the literature on assortative mating. Section 3 explains our approach to the measurement of assortative mating, stressing the need to make a distinction between actual and expected assortative mating. Section 4 then describes the database and the marriage market in Thailand, the country for which we give an empirical illustration. Section 5 presents empirical results for Thailand for the period 1985–2019. Section 6 describes the methodology allowing us to decompose the change over time in assortative mating into a component related to variations over time in the educational composition of the male and female populations and a component assumed to represent the impact of a "pure change" in assortative mating. An illustration of this methodology is then given, based again on data for Thailand and comparing the degree of assortative mating in 1985 and in 2019. Concluding comments are given in Section 7.

2 On assortative mating: a short review of literature

The topic of assortative mating has been of interest to social scientists working in various disciplines. There are thus papers written on this topic by sociologists, demographers, economists, anthropologists and social psychologists. This literature has shown that marriages are not random. As stressed by Hitsch et al. (2010), "marriage partners are similar in age, education levels, and physical traits such as looks, height, and weight". There may be various reasons for such a sorting. It may first be due to search frictions rather than preferences. In such a case sorting would simply be due to the fact that individuals are most likely to meet individuals who are similar to them in age, education or even religion. Sorting however can also exist in the absence of search frictions. Hitsch et al. (2010) thus mention the possibility of "horizontal" mate preferences, which means that a man (woman) often might prefer mating with a woman (man) who has traits similar to his (hers). But they also discuss the case of "vertical" preferences, "in the sense that each mate ranks all potential partners in the same way. In the equilibrium of a frictionless market, the ranks of the matched men and women will then be perfectly correlated. If the ranks are monotonically related to the mate's attributes, there will also be sorting along these attributes".

In addition, as stressed by Esteve et al. (2012), "union formation is in many ways a gender-asymmetrical process". In societies in which marriages are arranged, there is certainly similarity in several traits of the spouses, such as ethnicity, religion, caste. But marriage in such societies has also asymmetrical aspects, the custom of dowry being the most typical of these asymmetries. In contemporaneous Western societies there are generally free-choice unions and mate selection tends to be influenced by the personal traits of the individuals, education being often a crucial individual characteristic. A distinction has then to be made between homogamous and heterogamous couples. In the former case the spouses have similar characteristics, such as religion, age, ethnicity or education. In the latter case a distinction should be made between hypergamic and hypogamic unions. Hypergamic unions refer to the case when a "woman marries up", for example when her husband's education is higher than hers, the opposite being true for hypogamic couples.

In this section we present a succinct review of the literature, making a distinction between on one hand papers that are mainly of a theoretical nature, and on the other hand more empirically oriented studies.

2.1 Models of assortative mating

Becker's (1973, 1974) theory of marriage has certainly been the benchmark model of marriage among economists. Becker applied the theory of comparative advantage to the marriage market and concluded that if men are better than women at earning money in the labor market while women are better at taking care of the home and children, there will be an incentive for them to get married so that they can specialize in what they do best. According to Becker individuals before deciding to marry compare the utility they expect from getting married with the utility they presently have as single as well as with the utility they would have if looking for another mate. This kind of comparison holds also when an individual considers getting separated from his spouse/companion. There is hence in Becker's model a market for marriage and the kind of exchange described by Becker is assumed to be voluntary. Actually the development of the internet certainly confirms the existence of a marriage market since there are sites where individuals can specify the main features of their demand for a spouse, at the same time as they describe the characteristics of what they can supply.

Becker (1973, 1974) made however additional assumptions. He assumed that families pool their resources and optimize a single objective function. This assumption of pooling was criticized, among others, by Lundberg et al. (1997) who found that when, in the United Kingdom in the late 1970s, child allowances were transferred to wives, the household increased its expenditures on women's and children's clothing compared to husbands' clothing.

Becker also ignored the fact that the utilities of spouses are generally different, so that bargaining is likely to take place, even if not explicitly, and this bargaining will de facto determine the equilibrium outcome, as far as the allocation of resources (material resources and time) within the household is concerned. Manser and Brown (1980) and McElroy and Horney (1981) were probably the first to introduce household bargaining in this literature. A nice survey of the issues to be dealt with when adopting a household behavior model that includes bargaining may be found in Bourguignon and Chiappori (1992).

Becker (1973, 1974) discussed also assortative mating in marriage markets. His views on this topic appear also in his quite famous book (Becker 1981). There he argues (page 66) that "an efficient market usually has positive assortative mating, where high-quality men are matched with high-quality women and low-quality men with low-quality women, although negative assortative mating is sometimes important". The implications of Becker's model have been tested (see, for example, Grossbard-Shechtman 1993). Becker's model has been labeled "a transferable utility model of the marriage market" (Choo and Siow 2006) but such a model has seldom been estimated. As stressed by Choo and Siow (2006) two issues need to be solved before such a transferable utility model of the marriage market can be estimated. They first stress that equilibrium transfers in modern marriages are seldom observed. Second, individuals are not identical as they are likely to have different characteristics (such as age, education ethnicity, etc…). In Choo and Siow (2006) words "different types of individuals may not agree on the rankings of individuals of the opposite gender as spouses. Thus an empirical model of the marriage market should not impose too much a priori structure on the nature of preferences for marriage partners. However, without a priori structure, it is unclear what can be identified from the data". This is why Choo and Siow (2006) developed a structural empirical marriage matching model in which the individual preferences of a potential partner depend only on observable traits, such as the educational level. Assuming a distribution of male and female preferences over all possible spouse types, they derived market demand functions for matching partners which, when combined with an equilibrium market clearing condition, allowed them to obtain a marriage matching function.

Such marriage matching functions had actually been quite popular among mathematical demographers (see, for example, McFarland 1972). But, as stressed by Mourifié (2019), "one of the main advantages of the marriage matching functions proposed by economists over the demographers' approach is that they have clear microeconomic foundations and allow for spillover effects to be captured". While Mourifié and Siow (2014, 2017) assumed a peer effect specification, where the utilities of individuals are affected only by individuals of the same type, Mourifié (2019) took a more general approach where individuals may be affected by decisions made by other types. Galichon and Salanié (2015) have also investigated a model of one-to-one matching with transferable utility when some of the characteristics of the players are not observable. They show that the stable matching maximizes a social gain function that trades off complementarities in observable characteristics and matching on unobserved characteristics. Chiappori et al. (2017) construct a model of household decision-making which they then embed into a transferable utility matching framework with random preferences, as done in Choo and Siow (2006). The authors then show that as returns to human capital increase, couples at the top of the income distribution should spend more time with their children. In their model this should reinforce assortative matching. Noting that in Becker's (1973) model, the equilibrium matching is assumed to be unique and assortative, Cao et al. (2019) show that when only a subset of relevant characteristics is observed, the unique assortative matching does not uniquely determine a distribution of observed characteristics. Jaffe and Weber (2019) extended the matching model of Choo and Siow (2006) to allow for the possibility that the rate at which potential partners meet affects their probability of matching.

A different approach was taken by Saint-Paul (2015). He built an economic model of marriage that is based on biological differences between men and women. According to Saint-Paul the first important difference is that in nature women know for sure whom their children are, while men do not. The second key difference is that men can potentially have children with many women while women cannot. One of the main implications of Saint-Paul's model is that marriage markets will tend to be hypergamous. Saint-Paul's model allows only two possible equilibria. Either each individual get married to someone with the same rank in the distribution of incomes, what he calls the "Victorian" type of marriage, or women get married to men with a higher rank (human capital) than theirs, what St-Paul calls the "Sex and the City" type, a situation where you end up with many unmarried men at the bottom of the distribution of human capital and many single women at the top of the distribution.

Almås et al. (2020) discuss the theoretical arguments of why hypergamy may exist even when the distribution of earnings potential is the same for men and women. They show that hypergamy may be related to biological reasons (men are fertile for a longer period than women) but also to asymmetric valuation of partner attributes.

2.2 Empirical studies of assortative mating

Mare (1991), looking at Census and Current Population Survey data from 1940 to 1987 in the United States, found that the association between spouses' schooling increased between the 1930s and the 1970s but was stable or even decreased in the 1980s.

In their study of assortative meeting and mating, Kalmijn and Flap (2001) focused their attention on five meeting settings (work, school, neighborhood, common family networks and voluntary associations) and five types of homogamy (with respect to age, education, class destinations, class origins and religious background). Using Dutch data, they concluded that "schools promote most forms of homogamy, while work places only promote homogamy with respect to class destinations. Neighborhood and common family networks promote religious homogamy but are not related to homogamy with respect to class origins". More generally the authors stressed that "mating requires meeting".

Using data from the National Longitudinal Survey of Young Women (NLSYW) and the National Longitudinal Survey of Youth (NLSY) covering the period from the late 1960s to the early 1980s in the United States, Sweeney and Cancian (2004) concluded that there had been an increase over time in the association of wives' wages with the occupational status of their husbands, findings that evidently have important implications for long-term levels of inequality.

Observing that during the last three decades of the twentieth century there was in the United States both a decline in the percentage of the population marrying and an increase in the human capital accumulation of women and in their labor force participation rates, and given that there is a tendency for hypergamy, Rose (2005) concluded that the marriage market penalized women for their achievements, with this penalty growing over time. Rose called this marriage market penalty that is associated with professional success the "success gap". She measured it as the difference between the likelihood that a woman with exactly twelve years of education gets married and the likelihood that a woman with more than sixteen years of education marries. Using the United States Census of Population Public Use Microdata Sample (PUMS) for the years 1980, 1990 and 2000, she concluded that hypergamy as a whole declined, but it increased at the lower end of education distribution, and that declines in marriage rates were stronger among the less educated.

Schoen and Cheng (2006) worked with a complete count of all marriages around 1990 from three states in the United States, Virginia, North Carolina and Wisconsin. They concluded that there was an important symmetry between the educational levels of brides and grooms, this being particularly true among white and well educated individuals.

Smits and Park (2009) studied trends in educational homogamy in ten Asian countries and concluded that since the 1950s educational homogamy decreased. Educational homogamy was also lower in more modern societies, in countries with higher female employment and less Confucian influence.

Torche (2010) found, for each country that she studied (Brazil, Chile and Mexico), that "barriers to intermarriage cross educational groups are highly isomorphic to the earnings gaps between these educational groups". In fact international variations in marital sorting seem to reflect differences in earnings gap across countries. Torche (2010) therefore concludes that economic inequality "widens the cultural and spatial distances that prevent interaction and romance between individuals with different level of education".

Hitsch et al. (2010), using data from a dating site and adopting the Gale-Shapley (1962) algorithm to predict stable matches, found that, "in online dating, assortative mating arises in the absence of search frictions, due primarily to preferences and the specific market mechanism by which matches are formed".

Using the Integrated Public Use Microdata Series (IPUMS) international database, Esteve et al. (2012) concluded that, though still prevalent, hypergamy decreased during the last decades in quite a few countries. In Brazil and the United States hypogamous couples even outnumber hypergamous couples.

Using US marriage data for individuals born between 1943 and 1972, Chiappori et al. (2017) find that the preference for partners of the same education significantly increased for white individuals, particularly for the highly educated, while there does not seem to be any evidence of such an increase for black individuals.

Eika et al. (2019) examined the extent of educational assortative mating and its evolution over time in the United States, but also in Denmark, Germany the United Kingdom and Norway. They concluded that there was positive assortative mating at all levels of education in each country. However while assortative mating declined over time among college graduates, it increased among low educated individuals. These authors also concluded that assortative mating accounted for a non-negligible part of the cross-section inequality in household income in each country. The latter conclusion was already reached by Schwartz (2010) who, working with data from the March Current Population Survey in the United States, concluded that earnings inequality would have been about 25–30% lower in the absence of association between spouses' earnings.

Almås et al. (2020), using complete multi-generational data for all offspring born from 1952 through 1975 in Norway, concluded that "there is a steeper positive relationship between own earnings-potential rank and the probability of finding a partner for men than for women; there are more unmatched men than women, particularly at the bottom of the rank distribution, and men with higher rank tend to mate multiple times; and the man's rank tends to exceed the woman's rank within couples".

3 Data sources and the marriage market in Thailand

The data used in this study are taken from the annual Labor Force Survey (LFS) of Thailand, from 1985 to 2019. They were collected by the National Statistical Office of Thailand. We assign the married individuals to three educational groups based on their attainments: low level (with no, some or completed primary education), medium level (with some or completed secondary level education) and high level (with some or completed university level of education). This classification into three educational groups is based on the way the Thai Labor Force Survey (LFS) codes the educational levels. We grouped the levels into three categories to capture the major changes in educational achievements that took place over time. In doing so we actually followed previous studies on education in Thailand using LFS data, like Nakavachara (2010) and Paweenawat and McNown (2018). In Table 4 in Appendix A we give the number of observations by gender and educational level, from 1985 to 2019. Note that these numbers refer to all men and women, whether they are married, single, widow(er)s, divorced or separated.

In Thailand, given the rapid economic development and the increase in educational levels, decisions regarding marriage have dramatically changed over the last few decades. The norms and traditions are no longer the most significant determinants of family formation. The main decisions are now based on economic opportunities, educational attainments and personal interest, and, combined with traditions, they characterize the modern marriage market in Thailand. In Fig. 1a we give, for

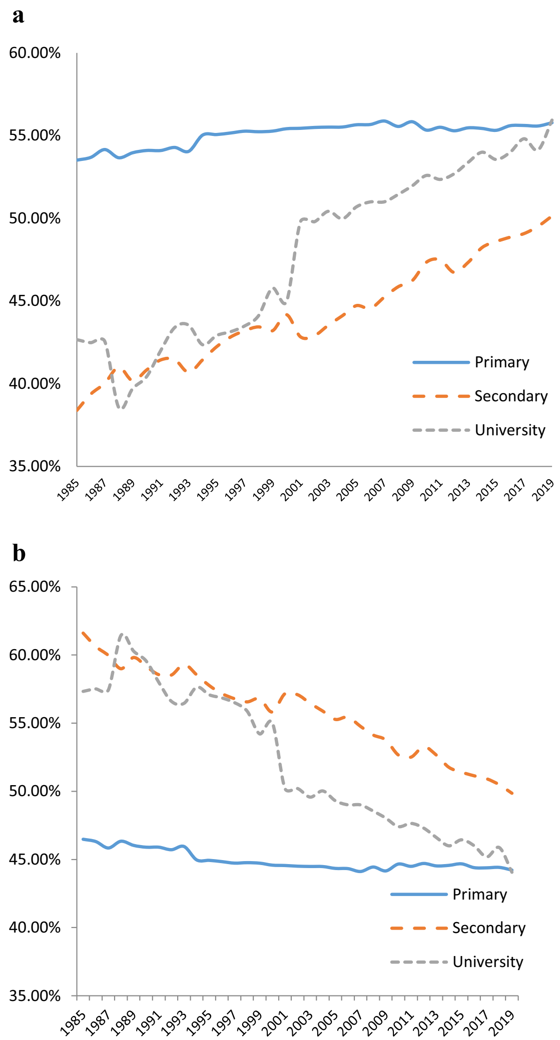

Fig. 1 **a** Share of married women by educational levels: 1985–2019. **b** Share of married men by educational levels: 1985–2019

each of the three educational levels distinguished, the percentage of women who are married. We observe that this percentage is generally highest among women with a low primary education and lowest among those with a University education. However whereas the share of married women among those with a primary education barely changed over time, the percentage of married women with a secondary or university education increased considerably over time. It is likely that in the mid to late 1980s there were relatively few women in these two groups and it was not easy for them to get married. In Fig. 1b we give, again for each of the three educational levels distinguished, the percentage of men who are married. Here we observe that the lowest share is observed for men with only primary education while the highest share is that of men with a secondary education. Note however that the percentage of married men declined over time, for those with a secondary as well as for those with a University education. It is likely that this downward trend in the percentage of married men among those with at least a secondary education is related to the increasing trend in the percentage of married women with at least a secondary education; this increase in female education is likely to imply that women prefer to complete their studies before getting married; moreover there must have been an increase in the financial independence of women who are much more likely now to have adopted Western patterns of marriage.

We also computed the percentage of single over the years, among men and women, separately for each of the three levels of education. These percentages are given in Table 5 in Appendix A and summarized graphically in Fig. 2. For men the percentage of single among those with a primary education only rose from 1.7 to

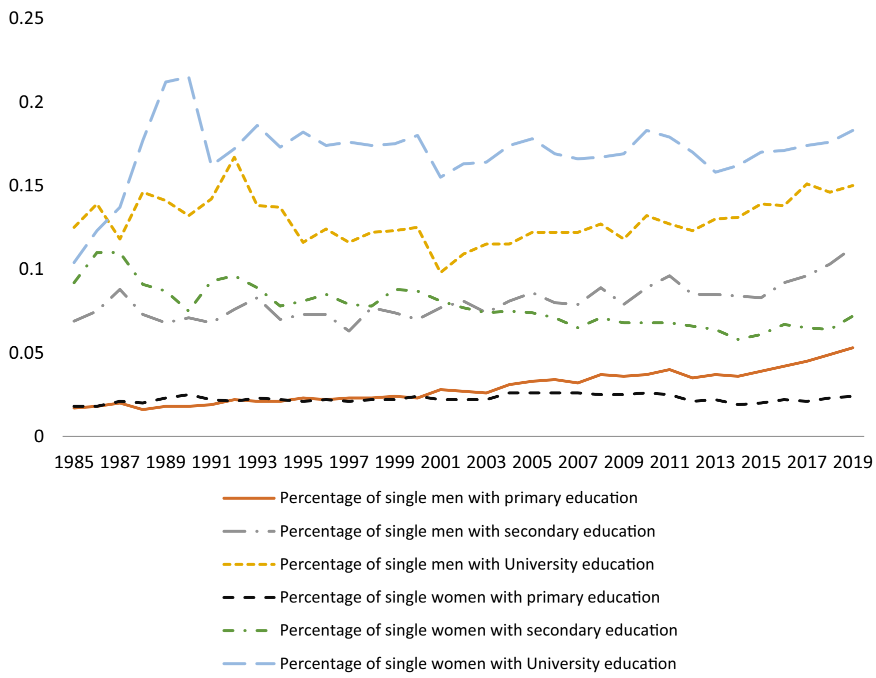

Fig. 2 Percentage of single individuals by gender and level of education

5.3% during the period 1985–2019. The increase during this same period among men with a secondary education was from 6.9 to 11.3%. Finally among men with a University education, the increase was from 12.5 to 15%. Note also that, as expected, whatever the year, the percentage of single men increases with the level of education.

For women we observe that the percentage of single among those with only a primary education rose from 1.8 to 2.4%, a small change. But among those with a secondary education there was during the same period of 35 years a decrease in the percentage of single from 9.2 to 7.2%. Finally among women with a University education there was a strong increase in the percentage of single, from 10.4 to 18.3%.

According to NESDB (2015), in Thailand, the mean age at first marriage of both men and women has risen over the last four decades. Smits and Park (2009) suggest that the decreasing trend in educational homogamy corresponded to educational expansion in several Asian countries, including Thailand. Paweenawat and Liao (2018) find that the trend towards education homogamy in Thailand varies with the education level. It is stronger among couples with secondary and university level, but lower among those with only a primary level of education. The outcomes are mainly driven by the rise in educational attainments.

4 Measuring assortative mating: the methodology

4.1 Analyzing the degree of non-randomness of mating at a given time, via the Theil index

Let n_{ij}, as before, refer to the number of males of educational level i who are married to females with educational level j. We assume that i and j vary from 1 to L, L being the number of educational levels that are distinguished. The educational levels of the husbands will vary from the lowest ($i = 1$) to the highest ($i = L$) levels. Similarly the educational levels of the wives will vary from the lowest ($j = 1$) to the highest ($j = L$) levels. Define now m_{ij} as

$$m_{ij} = n_{ij} / \left(\sum_{i=1}^{L} \sum_{j=1}^{L} n_{ij} \right), \tag{1}$$

m_{ij} refers therefore to the share in the total number of couples of the couples where the males have educational level i and got married to females who have educational level j.

Define also $m_{i.}$ and $m_{.j}$ as

$$m_{i.} = \sum_{j=1}^{L} m_{ij}. \tag{2}$$

$$m_{.j} = \sum_{i=1}^{L} m_{ij}. \tag{3}$$

Using Theil's (1967) application of the concept of entropy to the analysis of economic and social phenomena, we can define an index TH_{tot} of non-random mating[1] as

$$TH_{tot} = \sum_{i=1}^{L}\sum_{j=1}^{L}(m_{ij})\log\left[\frac{(m_{ij})}{(m_{i.}m_{.j})}\right]. \tag{4}$$

When $(m_{i.}m_{.j}) > m_{ij}$, the actual probability m_{ij} that a *m*ale with educational level i gets marr*i*ed to a female with educational level j is smaller than what this probability would have been $(m_{i.}m_{.j})$ had the mating been random. In such a case $\log[\frac{(m_{ij})}{(m_{i.}m_{.j})}]$ will be negative and so will the expression $\{m_{ij}\log[\frac{(m_{ij})}{(m_{i.}m_{.j})}]\}$.

Similarly when $(m_{i.}m_{.j}) < m_{ij}$, the actual probability m_{ij} that a male with educational level i gets married to a female with educational level j is higher than what this probability would have been (that is, $(m_{i.}m_{.j})$) had the mating been random. In such a case $\log[\frac{(m_{ij})}{(m_{i.}m_{.j})}]$ will be positive and so will the expression $m_{ij}\log[\frac{(m_{ij})}{(m_{i.}m_{.j})}]$.

Finally when $(m_{i.}m_{.j}) = m_{ij}$, the probability m_{ij} that a male with educational level i gets married to a female with educational level j is identical to what this probability would have been $(m_{i.}m_{.j})$ had the mating been random. In such a case $\log[\frac{(m_{ij})}{(m_{i.}m_{.j})}]$ will be equal to 0, and so will the expression $\{m_{ij}\log[\frac{(m_{ij})}{(m_{i.}m_{.j})}]\}$. If this is true for all i and j, TH_{tot} will be equal to 0.

The three cases distinguished previously may naturally be also derived with this index TH'_{tot}. In what follows, we will use the index TH_{tot}.

We will now show that this index may be easily broken down into the sum of three indices $TH_{h=f}$, $TH_{h>f}$ and $TH_{h<f}$ These indices measure respectively the degree of non-random mating among the males who married females with the same level of education as theirs, among males who married females with a lower level of education than theirs and among males who married females with a higher level of education than theirs. More precisely, we write

$$TH_{h=f} = \sum_{i=1}^{L}\left\{(m_{ii})\log\left[\frac{(m_{ii})}{(m_{i.}m_{.i})}\right]\right\}. \tag{5}$$

$$TH_{h<f} = \sum_{i=1}^{L}\sum_{j>1}^{L}(m_{ij})\log\left[\frac{(m_{ij})}{(m_{i.}m_{.j})}\right]. \tag{6}$$

$$TH_{h>f} = \sum_{i=1}^{L}\sum_{j<i}^{L}(m_{ij})\log\left[\frac{(m_{ij})}{(m_{i.}m_{.j})}\right]. \tag{7}$$

It is then easy to verify that

$$TH_{tot} = TH_{h=f} + TH_{h<f} + TH_{h>f}. \tag{8}$$

[1] Note that Theil (1967) defined also an alternative index TH'_{tot} as $TH'_{tot} = \sum_{i=1}^{L}\sum_{j=1}^{L}(m_{i.}m_{.j})\log[\frac{(m_{i.}m_{.j})}{m_{ij}}]$.

The Theil index may be also broken down into the sum of a between and a within groups components, as will now be shown.

Assume we have the same educational categories for both husbands and wives and that there are L such categories. Using the notations of the text, call $m_{i.}$ the share of husbands having educational category i and $m_{.j}$ the share of wives having educational category j. The product $p_{ij} = (m_{i.}m_{.j})$ is hence the "expected" share of couples where the husband has educational category i and the wife educational category j. Such an expected share will be observed when there is independence between the educational categories of husbands and wives. In reality the actual share of couples where the husband has educational category i and the wife educational category j is m_{ij} with in general $m_{ij} \neq p_{ij}$.

The Theil index of non-assortative mating is then defined as

$$TH_{tot} = \sum_{i=1}^{L} \sum_{j=1}^{L} (m_{ij}) \ln \left(\frac{m_{ij}}{(m_{i.}m_{.j})} \right). \tag{9}$$

Assume now that we divide the couples into three categories ($L = 3$):

group 1: those couples where the husband and the wife have the same level of education, no matter which level
group 2: those couples where he husband has a higher level of education than the wife
group 3: those couples where the wife has a higher level of education than the husband.

The between groups Theil index TH_{bet} will then be expressed as

$$TH_{bet} = \sum_{k=1}^{3} e_k \ln \frac{e_k}{a_k}, \tag{10}$$

where e_k and a_k refer respectively to the expected and actual shares of group k in the total number of couples.

For example, for group1, we have:

$$a_1 = m_{11} + m_{22} + m_{33}$$

while

$$e_1 = (m_{1.}m_{.1}) + (m_{2.}m_{.2}) + (m_{3.}m_{.3})$$

The within group k Theil index will be written as

$$TH_k = \sum_{i \in k} \sum_{j \in k} m_{ij} \ln \left(\frac{m_{ij}}{(m_{i.}m_{.j})} \right). \tag{11}$$

The within groups Theil index TH_{with} is then expressed as

$$TH_{with} = \sum_{k=1}^{K} a_k TH_k. \tag{12}$$

It can be proven that the overall Theil index TH_{tot} may be decomposed into the sum of the between and within groups Theil indices. In other words

$$TH_{tot} = TH_{bet} + TH_{with}. \quad (13)$$

4.2 Properties of the Theil index of "non-random mating"

It is easy to observe that the Theil index TH_{tot} that was defined in (4) will be equal to 0 when $(m_{ij}) = (m_{i.}m_{.j})\forall i$ and $\forall j$. In such a case there will be perfect independence between the educational level of the husband and that of the wife, so that $(m_{ij}/m_{i.}) = (m_{.j}/1)$ $\forall i$ and $\forall j$ and $(m_{ij}/m_{.j}) = (m_{i.}/1)$, $\forall i$ and $\forall j$.

If a wife increases her level of education from level j to, say, level $j+1$, so that $(m_{ij}/m_{i.})$ decreases while $(m_{i,j+1}/m_{i.})$ increases, while there is no change in the other cells of the matrix $\{m_{ij}\}$ of the cells m_{ij}, then the extent of non-random mating will decrease (increase) if originally the ratio $[\frac{m_{ij}}{(m_{i.}m_{.j})}]$ was higher (lower) than the ratio $[\frac{m_{ij+1}}{(m_{i.}m_{.j+1})}]$. This property is known as the transfer principle.

Similarly, if a husband increases his level of education from level i to, say, level $i+1$, so that the ratio $(m_{ij}/m_{.j})$ decreases, while $(m_{i+1,j}/m_{.j})$ increases, while there is no change in the other cells of the matrix $\{m_{ij}\}$ of the cells m_{ij}, then the extent of non-random mating will decrease (increase) if originally the ratio $[\frac{m_{ij}}{(m_{i.}m_{.j})}]$ was higher (lower) than the ratio $[\frac{m_{i+1,j}}{(m_{i+1.}m_{.j})}]$.

If there is an increase in the population of spouses but if this increase is such that the new number of spouses n'_{ij} located in the cell (i, j) is expressed as $n'_{ij} = \lambda n_{ij}$, $\forall i$ and $\forall j$, there will be no change in the shares m_{ij} and henceforth in the value of the Theil index TH_{tot}. This property is known as the replication principle.

Finally since what matters for the computation of the Theil index TH_{tot} are the shares (m_{ij}) and $(m_{i.}m_{.j})$ and the ratios $(\frac{m_{ij}}{m_{i.}m_{.j}})$, a permutation of the lines corresponding to the educational levels of the husbands, or of the columns corresponding to the educational levels of the wives, will not affect the value of this Theil index. For example, assume that originally $i=1$ and $j=1$ referred to the lowest level of education of husbands and wives and $i=3$ and $j=3$ to the highest level, while $i=j=2$ correspond to the middle level of education. If we now decide to call $i=j=1$ the highest level of education and $i=j=3$ the lowest level, there will be no change in the Theil index. These assumptions concerning permutations correspond to what is called the anonymity or symmetry principle.

5 An empirical illustration: non-random mating in Thailand between 1985 and 2019

In Table 6 in Appendix A we look separately at the three groups defined previously, namely the couples who have the same level of education, the couples where the husband has a lower level of education than his wife and the couples where the opposite is true. These data are summarized in Fig. 3. We observe that the actual

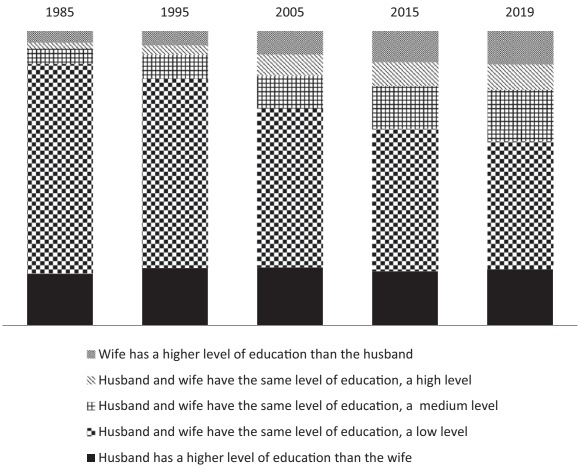

Fig. 3 Actual shares of the different categories of couples (selected years)

share of the couples with the same level of education decreased between 1985 and 2019, from around 79 to about 70% while the share of the group where the wife has a higher level of education increased, during this same period, from about 4 to 11%. As far as the share of the group, where the husband has a higher level of education than his wife, is concerned, the picture is not clear, as there is not much change over time during the 1985–2019 periods, although there are short term variations.

In Table 7 in Appendix A we give also the expected shares of the various categories of couples. While, as mentioned previously, the actual share of the couples where husband and wife have the same level of education decreased from 0.79 to 0.70, the expected share of couples with the same level of education decreased drastically between 1985 and 2019, from 66 to 43%. Therefore the ratio of the actual over the expected shares in this group increased a lot, from 1.19 in 1985 to 1.72 in 2019. Therefore, while a “naïve” look at the data would have led us to conclude that assortative mating decreased over time since the share of couples with the same level of education decreased, a closer look at the data shows that the ratio of the actual over the expected share of this group increased. The logarithm of this ratio which appears in the formulation of the Theil index of non-random mating is hence positive and it increased over time. This clearly indicates that this group of couples, where husband and wife have the same level of education, had a positive impact on the extent of non-random mating and this impact increased over time (see, Table 10).

For the group where the wife has a higher level of education than her husband, we observe an increase over time in the actual share of this group (from 3.8 to 11.3%) but also an increase in the expected share of this group (from 10.8 to 25.8%). As a consequence, the ratio of the actual over the expected shares of this group rose from 0.35 to 0.44 between 1985 and 2019. So here “naïve” conclusions based only the actual shares of the group would have led us to conclude that this group had a positive impact on non-assortative (random) mating and this impact increased over time. Such a conclusion however cannot be based only on what happened to the actual share of this group. As mentioned previously, we have to look at the ratio of the actual over the expected shares of this group and it then appears that this ratio remains smaller but increased over time. This means that the logarithm of this ratio remained negative and increased over time in absolute value. Given the formulation of the Theil index of non-random mating, this implies that the contribution of this group to the extent of non-random mating is negative and increased in absolute value between 1985 and 2019 (see, Table 10).

In short while for the group of couples where husband and wife have the same level of education looking at the actual share of the group and at the ratio of the actual over the expected share of the group led to opposite conclusions, this was not the case of the group where the wife has a higher level of education.

In Table 6 in Appendix A we make also a distinction, for the couples who have the same level of education, between the three levels of education. Table 8 in Appendix A completes this analysis by giving also the expected shares of these different categories of couples with the same level of education. For those with a low level of education, we then observe a decrease over time in both the actual (from 72 to 44%) and the expected (from 64 to 29%) shares of this group (in the total number of couples, whether they have the same or a different level of education), but the decrease in the expected share was stronger, since the ratio of the actual over the expected share increased from 1.12 to 1.50. So we cannot really conclude that there was an increase in the level of assortative mating among men and women with low levels of education. For couples with a medium level of education, we observe an increase in the actual (from 5% to almost 18%) and in the expected (from 1 to 9%) shares of these couples (again in the total number of couples) but the ratio of actual over expected shares decreased from 4.1to 1.9704 between 1985 and 2019. So here we should again not only look at what happens to the actual shares of couples with the same medium level of education, because then we would conclude that there was an increase in assortative mating in this group. Checking what happened to the ratio of the actual over the expected shares indicates on the contrary a strong decrease in this ratio, an observation that would rather lead us to conclude that there was a decrease in assortative mating. The data concerning couples with a high level of education indicate an increase in both the actual (from 2 to 8.5%) and the expected (from 0.6 to 2.5%) shares of this group in the total number of couple. The ratio of actual over expected shares in this group is however very instable, mainly because the expected as well as the actual shares are quite low so that it seems difficult to draw clear-cut conclusions concerning this group.

In Table 9 in Appendix A we look at the decomposition of the Theil index of non-random mating into two components. These data are summarized in Fig. 4. The first component gives the extent of between groups non-random mating, in so far as we ignore differences in the ratios of actual over expected shares within a given group, the groups referring respectively to couples with the same level of education, couples where the wife has a higher level of education and couples where she has a lower level of education than her husband. It then appears that there was a strong increase in the extent of between the three types of couples non-random mating (the between groups Theil index increased from 0.05 to 0.18 between 1985 and 2019). For the within groups evolution of the Theil index, we first observe an increase in the value of the index, from 0.063 in 1985 to 0.081 in 2003 but afterwards there was a decrease and in 2019 the value of this Theil index was 0.056. We also give in Table 9 bootstrap confidence intervals for the overall Theil index and the between and within groups Theil indices.

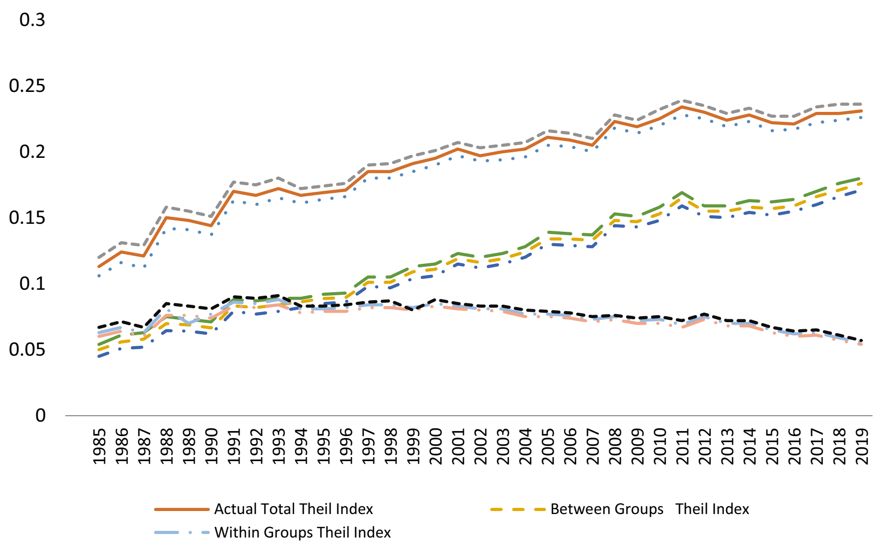

Fig. 4 Decomposition of the Theil index into between and within groups non-random mating, with bootstrap 5–95% confidence intervals

In Table 10 in Appendix A we give the contribution of the three categories of couples to the overall Theil index. The data are summarized in Fig. 5. It appears that there was a strong increase in the contribution of couples with the same level of education (from 0.18 to 0.40). Note that this contribution is higher than the overall Theil index because the contribution of the two other categories of couples is negative, indicating that the ratio of the actual over the expected shares in these two groups is smaller than one. In fact the contribution of these two groups became more negative over time (from −0.03 to −007 for the couples where the wife has a higher level of education than her husband, and from −0.3 to −0.10 for the couples where the husband has a higher level of education).

6 Analyzing changes over time in the degree of non-random mating in Thailand

6.1 The methodology

Assume now that we want to compare two "mating matrices" $\{m_{ij}\}$ and $\{v_{ij}\}$, referring to two different periods. On the basis of each of these two matrices, we can compute a Theil index of non-random mating, as defined in expression (4), and conclude in which case the degree of "non-random mating" was higher. We have however to be careful in drawing conclusions, as the educational composition of the male and female populations may have changed over time.

In fact the change in the degree of "non-random mating" may be the consequence of a variation over time in the degree of independence between the rows (educational levels of the husbands) and the columns (educational levels of the

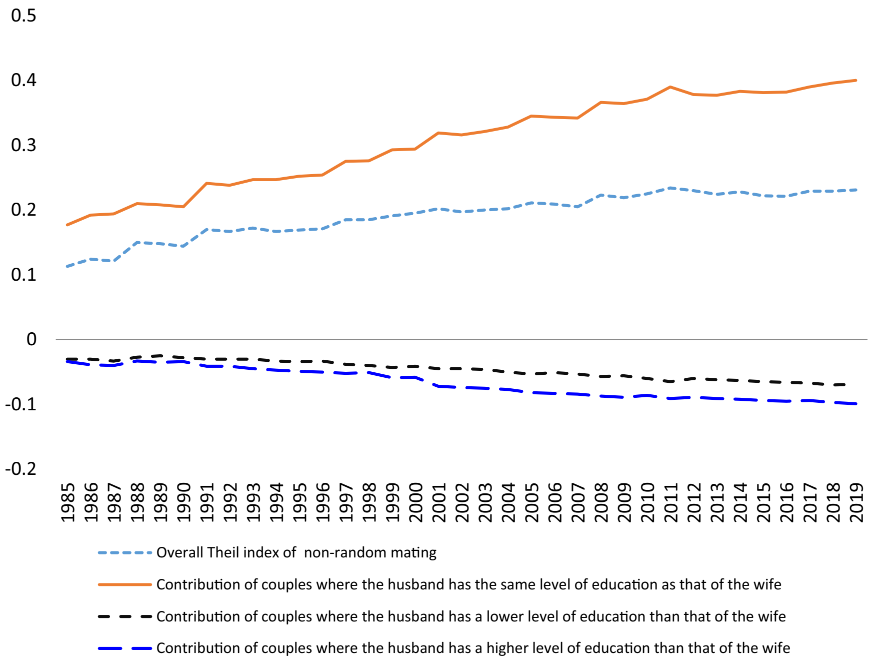

Fig. 5 Decomposition of the Theil index of non-random mating, with a breakdown by the three combinations of levels of education of the spouses

wives), but also of a change over time in the education composition of the husbands and wives. This important point was already stressed by Pencavel (1998) who, analyzing US data for the period 1940–1990, wrote that "the inference that schooling homogamy has increased since 1960 is subject to the objection that the measures do not effectively hold constant changes in the marginal distribution of schooling". Bratsberg et al. (2018) also emphasized the fact that "it is difficult to interpret trends in educational assortative mating, as they can arise from change in sorting into education as much as from change in sorting into partnership". Keeping constant marginal distributions, they concluded that assortative mating declined in Norway over the last 30 years.

In other words, we ought to make a distinction between the impact of a change in the margins of a matrix such as $\{m_{ij}\}$ and that of a change in the "internal structure" of this matrix. This is the terminology used by Karmel and MacLachlan (1988), when they analyzed changes over time in occupational segregation by gender, and in our case it reflects "pure" change in non-random assortative mating, "pure" meaning "net of changes in the margins".

It is in fact possible to apply the methodology proposed by Karmel and MacLachlan (1988) to analyze changes over time in the degree of "non-random mating". Such a methodology uses an algorithm originally proposed by Deming and Stephan (1940) and which is explained via a simple illustration in Appendix B. In addition it is possible, as was stressed by Deutsch et al. (2009), in the framework of occupational segregation analysis, to generalize this methodology by applying also the concept of Shapley decomposition (see, Chantreuil and Trannoy 2013; Shorrocks 2013). This concept of Shapley decomposition and its application to "non-random assortative mating" is presented in Appendix C. In other words, we will be able to derive the specific contributions of three components to the overall variation over time in the extent of "non-random mating". The first component corresponds to changes over time in the educational composition of the husbands while the second reflects changes over time in the educational composition of the wives, the sum of these two variations corresponding to changes in the margins of the original educational matching matrix. Finally the third component represents the "pure" variation in the extent of "non-random mating" and it reflects changes in the degree of independence between the lines and columns of the original educational matching matrix itself.

6.2 An empirical illustration: changes between 1985 and 2019 in the extent of non-random mating in Thailand

Combining the algorithm of Deming and Stephan (1940) with the concept of Shapley decomposition (see, Appendix B and Appendix C for more details) we can find out what the main determinants of the variation over time in the degree of non-random mating are. Such a decomposition appears in Table 1 for the change that took place in the overall Theil index over the whole period 1985–2019. We also give in Table 1 the bootstrap confidence intervals for the components of this change. Looking at the first column of Table 1, we observe that the increase in this Theil index is mainly due to variations between 1985 and 2019 in the margins of the 3 by 3 matrices giving the distribution of the couples by educational level of both spouses. More precisely the main factor of change is the change in the educational composition of the women, since this variation over time explains practically the totality of the increase in the degree of non-random mating.

But if we divide the period 1985–2019 into sub-periods we will observe that in some periods the increase in the overall Theil index is also due to a change in the "internal structure" of the matrices, that is to a modification in the "pure non-random matching" (degree of dependence between the rows and the columns of the matrix). This is, for example, the case of the period 1985–1988 (see Table 2) where most of the change in the overall Theil index is the consequence of a change in the "internal structure" of the matrices. The results for the other sub-periods are given in Tables 11 to 15 in Appendix A.

Finally in Table 3 we present the respective contribution of the three groups previously distinguished to the components of the breakdown of the change over time in non-random mating. It then appears that the magnitude of the contribution of spouses with the same level of education to the contribution of changes in the educational structure of the female population is 52% higher (it is equal to 0.175) than the total contribution of changes in the educational structure of the female population (which is equal to 0.115). The contribution of the two other groups (spouses with different levels of education) is actually negative and hence per se would actually have led to a decrease in the contribution of changes in the educational structure of the female population. In short this change in the educational

Table 1 Decomposition of the change in the overall Theil index of non-random mating during the period 1985–2019

Decomposition of changes in the Theil index of non-random mating	Actual values of the Theil index and of the components of its change over time	Lower bound of bootstrap confidence interval (5%)	Upper bound of bootstrap confidence interval (95%)
A-Degree of "non-random mating" in 1985	0.113	0.105	0.121
B-Degree of "non-random mating" in 2019	0.231	0.226	0.237
C = (B−A) = (D+E) Gross variation in "non-random mating"	0.118	0.109	0.128
D-Contribution of variation in the "internal structure of the "mating matrices"	0.000	−0.011	0.010
E = (F+G) - Contribution of variation in the margins of the "mating matrices"	0.118	0.113	0.123
F- Contribution of changes in the educational structure of the male population (horizontal margins)	0.003	0.001	0.006
G- Contribution of changes in the educational structure of the female population (vertical margins)	0.115	0.110	0.119

Table 2 Decomposition of the change in the overall Theil index of non-random mating during the period 1985–1988

Decomposition of changes in the Theil index of non-random mating	Actual values of the Theil index and of the components of its change over time	Lower bound of bootstrap confidence interval (5%)	Upper bound of bootstrap confidence interval (95%)
A- Degree of "non-random mating" in 1985	0.113	0.106	0.121
B-Degree of "non-random mating" in 1988	0.15	0.142	0.159
C = (B−A) = (D+E) Gross variation in "non-random mating"	0.037	0.026	0.047
D-Contribution of variation in the "internal structure of the "mating matrices"	0.036	0.027	0.045
E = F+G Contribution of variation in the margins of the "mating matrices"	0.001	¯0.005	0.007
F-Contribution of changes in the educational structure of the male population (horizontal margins)	0.001	0.000	0.003
G-Contribution of changes in the educational structure of the female population (vertical margins)	−0.000	−0.007	0.006

structure of the wives is mainly due to the increase in the extent of mating between spouses with the same level of education.

7 Concluding comments

In this paper we suggested a new way of measuring assortative mating and its change over time. More precisely we considered that a measure of non-assortative

Table 3 Decomposition of the change between 1985 and 2019 in the Theil index of non-random mating, with a breakdown by the three combinations of levels of education of the spouses

Decomposition of changes in the Theil index of non-random mating	Actual values of the Theil index and of the components of its change over time	Contribution of the couples where the husband and the wife have the same level of education	Contribution of the couples where the husband has a lower level of education than that of his wife	Contribution of the couples where the husband has a higher level of education than that of his wife
A-Degree of "non-random mating" in 1985	0.113	0.177	−0.0297	−0.034
B-Degree of "non-random mating" in 2019	0.231	0.4	−0.0691	−0.0992
C = (B−A) = (D+E) Gross variation in "non-random mating"	0.118	0.223	−0.0394	−0.0651
D-Contribution of variation in the "internal structure of the "mating matrices"	0.000309	0.0141	0.00545	−0.0193
E = (F+G) Contribution of variation in the margins of the "mating matrices"	0.118	0.209	−0.0449	−0.0459
F- Contribution of changes in the educational structure of the male population (horizontal margins)	0.00334	0.0335	0.00535	−0.0355
G-Contribution of changes in the educational structure of the female population (vertical margins)	0.115	0.175	−0.0502	−0.0103

mating should reflect the degree of dependence between the characteristics of the husband and of the wife. Focusing on the educational level of the spouses as the characteristic under study, we described an algorithm that has been applied in the past in other domains of social sciences, one that allows one to make a distinction between changes in the distribution of husbands and wives by educational level and a "pure change in assortative mating". The latter is defined as the consequence of a change in the degree of independence between the educational levels of husbands and wives. Our empirical illustration showed that, if one looks at the whole period 1985–2019, the increase in the Theil index of non-random mating was uniquely due to a change in the educational composition of the males and females (essentially of the female population). However there are several sub-periods where the "pure change in assortative mating", that is, where the degree of independence between the lines and the columns of the educational matrices that were analyzed, played an important role

Acknowledgements We would like to thank the National Statistical Office of Thailand, the Research Institute for Policy Evaluation and Design, and the University of the Thai Chamber of Commerce (UTCC), Thailand, for access to the data used in this paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

8 Appendix A Tables

Tables 4–15

Table 4 Yearly number of observations for men and women by educational level

	Men			Women		
Year	Primary	Secondary	University	Primary	Secondary	University
1985	8804	1809	805	11,197	1192	615
1986	8755	1937	842	11,151	1362	651
1987	7879	1801	740	10,252	1282	598
1988	8287	1778	659	10,618	1349	440
1989	11,392	2567	849	14,704	1856	622
1990	10,839	2439	764	14,112	1789	596
1991	10,961	2828	1083	14,234	2165	819
1992	10,826	2977	1097	14,010	2258	871
1993	10,436	2917	1113	13,456	2157	940
1994	20,704	6671	2,336	27,944	5022	1853
1995	20,345	6734	2404	27,259	5265	2038
1996	19,696	6714	2464	26,706	5391	2077
1997	19,312	6840	2544	26,215	5556	2177
1998	18,652	6990	2705	25,253	5728	2397
1999	18,190	7094	2844	24,690	5849	2684
2000	17,626	7073	2995	24,146	6093	2748
2001	22,763	9343	3778	31,085	7440	4237
2002	23,366	9541	3624	31,817	7560	4082
2003	22,106	9389	3600	30,378	7681	4116
2004	21,578	9596	3883	29,604	7938	4443
2005	21,242	10,108	4265	29,206	8601	4979
2006	22,538	11,267	4703	31,067	9527	5495
2007	21,971	11,112	4692	30,616	9614	5457
2008	22,466	11,496	4954	30,580	10,120	5891
2009	21,366	11,350	4,691	29,386	10,235	5777
2010	14,127	7702	3253	19,093	7169	4068
2011	20,208	11,485	4880	27,396	10,708	6115
2012	23,062	12,129	4556	30,752	11,009	5715
2013	22,230	11,874	4602	29,860	11,054	5789
2014	21,651	11,882	4410	29,034	11,328	5725
2015	20,523	11,799	4571	27,366	11,481	5825
2016	19,693	12,121	4586	26,543	11,913	5924
2017	19,133	12,049	4482	25,671	11,899	5970
2018	18,412	12,198	4373	24,390	12,021	5793
2019	17,244	12,174	4009	22,921	12,242	5664

Table 5 Percentage of single by gender and level of education

Year	Percentage of single men with primary education	Percentage of single men with secondary education	Percentage of single men with University education	Percentage of single women with primary education	Percentage of single women with secondary education	Percentage of single women with University education
1985	0.017	0.069	0.125	0.018	0.092	0.104
1986	0.018	0.075	0.139	0.018	0.110	0.123
1987	0.020	0.088	0.118	0.021	0.110	0.137
1988	0.016	0.073	0.146	0.020	0.091	0.177
1989	0.018	0.068	0.141	0.023	0.087	0.212
1990	0.018	0.071	0.132	0.025	0.075	0.215
1991	0.019	0.068	0.142	0.022	0.093	0.162
1992	0.022	0.076	0.167	0.021	0.096	0.172
1993	0.021	0.083	0.138	0.023	0.089	0.186
1994	0.021	0.07	0.137	0.022	0.078	0.173
1995	0.023	0.073	0.116	0.021	0.081	0.182
1996	0.022	0.073	0.124	0.022	0.085	0.174
1997	0.023	0.063	0.116	0.021	0.079	0.176
1998	0.023	0.077	0.122	0.022	0.078	0.174
1999	0.024	0.074	0.123	0.022	0.088	0.175
2000	0.023	0.070	0.125	0.024	0.087	0.180
2001	0.028	0.077	0.098	0.022	0.081	0.155
2002	0.027	0.081	0.109	0.022	0.077	0.163
2003	0.026	0.074	0.115	0.022	0.074	0.164
2004	0.031	0.081	0.115	0.026	0.075	0.174
2005	0.033	0.086	0.122	0.026	0.074	0.178
2006	0.034	0.080	0.122	0.026	0.071	0.169
2007	0.032	0.079	0.122	0.026	0.065	0.166
2008	0.037	0.089	0.127	0.025	0.071	0.167
2009	0.036	0.079	0.118	0.025	0.068	0.169
2010	0.037	0.089	0.132	0.026	0.068	0.183
2011	0.040	0.096	0.127	0.025	0.068	0.179
2012	0.035	0.085	0.123	0.021	0.066	0.17
2013	0.037	0.085	0.130	0.022	0.064	0.158
2014	0.036	0.084	0.131	0.019	0.058	0.162
2015	0.039	0.083	0.139	0.020	0.061	0.170
2016	0.042	0.092	0.138	0.022	0.067	0.171
2017	0.045	0.096	0.151	0.021	0.065	0.174
2018	0.049	0.103	0.146	0.023	0.064	0.176
2019	0.053	0.113	0.150	0.024	0.072	0.183

Table 6 Actual shares of the three categories of couples

Year	Actual shares of the couples with same level of education	Actual shares of the couples where the wife has a higher level of education than the husband	Actual shares of the couples where the husband has a higher level of education than the wife	Actual share of couples with the same low level of education	Actual share of couples with the same medium level of education	Actual share of couples with the same high level of education
1985	0.787	0.038	0.175	0.716	0.051	0.021
1986	0.785	0.040	0.176	0.707	0.056	0.022
1987	0.779	0.038	0.183	0.698	0.057	0.024
1988	0.813	0.027	0.160	0.733	0.057	0.023
1989	0.808	0.027	0.164	0.728	0.059	0.022
1990	0.809	0.029	0.161	0.730	0.059	0.021
1991	0.792	0.034	0.174	0.694	0.066	0.032
1992	0.783	0.038	0.179	0.684	0.068	0.031
1993	0.787	0.040	0.174	0.684	0.068	0.035
1994	0.766	0.043	0.192	0.656	0.078	0.032
1995	0.758	0.047	0.195	0.644	0.081	0.033
1996	0.751	0.050	0.199	0.634	0.082	0.035
1997	0.757	0.050	0.194	0.630	0.089	0.038
1998	0.745	0.055	0.200	0.615	0.089	0.042
1999	0.744	0.059	0.197	0.602	0.094	0.048
2000	0.735	0.061	0.204	0.590	0.095	0.050
2001	0.738	0.065	0.196	0.580	0.093	0.065
2002	0.743	0.064	0.193	0.587	0.097	0.059
2003	0.738	0.070	0.191	0.576	0.101	0.061
2004	0.729	0.073	0.199	0.558	0.105	0.066
2005	0.725	0.079	0.197	0.540	0.111	0.073
2006	0.713	0.083	0.203	0.524	0.115	0.074
2007	0.708	0.085	0.206	0.516	0.117	0.075
2008	0.719	0.085	0.196	0.516	0.122	0.081
2009	0.711	0.089	0.200	0.503	0.127	0.080
2010	0.709	0.095	0.196	0.495	0.131	0.084
2011	0.715	0.095	0.191	0.487	0.139	0.088
2012	0.730	0.090	0.180	0.520	0.136	0.075
2013	0.722	0.094	0.184	0.508	0.136	0.078
2014	0.721	0.097	0.182	0.501	0.142	0.078
2015	0.711	0.105	0.184	0.485	0.146	0.080
2016	0.702	0.107	0.191	0.467	0.153	0.083
2017	0.705	0.107	0.188	0.465	0.158	0.084
2018	0.702	0.110	0.189	0.451	0.164	0.086
2019	0.696	0.113	0.191	0.436	0.176	0.085

Table 7 Ratios of actual over expected shares in each of the three groups

Year	Actual shares of the couples with same level of education	Expected shares of the couples with same level of education	Ratio of actual over expected shares (couples with same level of education)	Actual shares of the couples where the wife has a higher level of education than the husband	Expected shares of the couples where the wife has a higher level of education than the husband	Ratio of actual over expected shares (couples where the wife has a higher level of education than the husband)	Actual shares of the couples where the husband has a higher level of education than the wife	Expected shares of the couples where the husband has a higher level of education than the wife	Ratio of actual over expected shares (couples where the husband has a higher level of education than the wife)
1985	0.787	0.660	1.19	0.038	0.108	0.348	0.175	0.232	0.754
1986	0.785	0.646	1.22	0.040	0.115	0.348	0.176	0.239	0.736
1987	0.779	0.639	1.22	0.038	0.115	0.33	0.183	0.246	0.744
1988	0.813	0.667	1.22	0.027	0.108	0.25	0.160	0.225	0.711
1989	0.808	0.662	1.22	0.027	0.108	0.25	0.164	0.230	0.713
1990	0.809	0.665	1.22	0.029	0.109	0.266	0.161	0.226	0.712
1991	0.792	0.622	1.27	0.034	0.128	0.266	0.174	0.250	0.696
1992	0.783	0.611	1.28	0.038	0.133	0.286	0.179	0.256	0.699
1993	0.787	0.611	1.29	0.040	0.135	0.296	0.174	0.254	0.685
1994	0.766	0.586	1.31	0.043	0.141	0.305	0.192	0.274	0.701
1995	0.758	0.572	1.33	0.047	0.148	0.318	0.195	0.280	0.696
1996	0.751	0.562	1.34	0.050	0.152	0.329	0.199	0.286	0.696
1997	0.757	0.555	1.36	0.050	0.159	0.314	0.194	0.286	0.678
1998	0.745	0.542	1.37	0.055	0.166	0.331	0.200	0.292	0.685
1999	0.744	0.529	1.41	0.059	0.174	0.339	0.197	0.297	0.663
2000	0.735	0.518	1.42	0.061	0.178	0.343	0.204	0.304	0.671
2001	0.738	0.509	1.45	0.065	0.185	0.351	0.196	0.306	0.641
2002	0.743	0.517	1.44	0.064	0.181	0.354	0.193	0.301	0.641

Table 7 continued

Year	Actual shares of the couples with same level of education	Expected shares of the couples with same level of education	Ratio of actual over expected shares (couples with same level of education)	Actual shares of the couples where the wife has a higher level of education than the husband	Expected shares of the couples where the wife has a higher level of education than the husband	Ratio of actual over expected shares (couples where the wife has a higher level of education than the husband)	Actual shares of the couples where the husband has a higher level of education than the wife	Expected shares of the couples where the husband has a higher level of education than the wife	Ratio of actual over expected shares (couples where the husband has a higher level of education than the wife)
2003	0.738	0.508	1.45	0.070	0.189	0.37	0.191	0.303	0.63
2004	0.729	0.492	1.48	0.073	0.196	0.372	0.199	0.313	0.636
2005	0.725	0.476	1.52	0.079	0.207	0.382	0.197	0.317	0.621
2006	0.713	0.464	1.54	0.083	0.211	0.393	0.203	0.325	0.625
2007	0.708	0.460	1.54	0.085	0.213	0.399	0.206	0.328	0.628
2008	0.719	0.456	1.58	0.085	0.220	0.386	0.196	0.324	0.605
2009	0.711	0.448	1.59	0.089	0.223	0.399	0.200	0.329	0.608
2010	0.709	0.440	1.61	0.095	0.233	0.408	0.196	0.327	0.599
2011	0.715	0.435	1.64	0.095	0.238	0.399	0.191	0.328	0.582
2012	0.730	0.459	1.59	0.090	0.227	0.396	0.180	0.314	0.573
2013	0.722	0.451	1.6	0.094	0.231	0.407	0.184	0.318	0.579
2014	0.721	0.447	1.61	0.097	0.235	0.413	0.182	0.318	0.572
2015	0.711	0.437	1.63	0.105	0.243	0.432	0.184	0.321	0.573
2016	0.702	0.425	1.65	0.107	0.246	0.435	0.191	0.328	0.580
2017	0.705	0.422	1.67	0.107	0.250	0.430	0.188	0.328	0.573
2018	0.702	0.415	1.69	0.110	0.254	0.432	0.189	0.332	0.570
2019	0.696	0.405	1.72	0.113	0.258	0.437	0.191	0.337	0.566

Table 8 Non-random mating in Thailand among couples with same level of education

Year	Actual share of couples with the same low level of education	Expected share of couples with the same low level of education	Ratio of actual over expected shares among the couples with the same low level of education	Actual share of couples with the same medium level of education	Expected share of couples with the same medium level of education	Ratio of actual over expected shares among the couples with the same medium level of education	Actual share of couples with the same high level of education	Expected share of couples with the same high level of education	Ratio of actual over expected shares among the couples with the same high level of education
1985	0.716	0.642	1.12	0.051	0.012	4.10	0.021	0.006	3.8
1986	0.707	0.626	1.13	0.056	0.014	4	0.022	0.006	3.67
1987	0.698	0.618	1.13	0.057	0.015	3.8	0.024	0.006	4
1988	0.733	0.649	1.13	0.057	0.015	3.8	0.023	0.003	7.67
1989	0.728	0.643	1.13	0.059	0.016	3.69	0.022	0.003	7.33
1990	0.730	0.646	1.13	0.059	0.016	3.69	0.021	0.003	7
1991	0.694	0.596	1.16	0.066	0.020	3.3	0.032	0.005	6.4
1992	0.684	0.584	1.17	0.068	0.022	3.09	0.031	0.006	5.17
1993	0.684	0.584	1.17	0.068	0.021	3.24	0.035	0.006	5.83
1994	0.656	0.552	1.19	0.078	0.027	2.89	0.032	0.006	5.33
1995	0.644	0.536	1.2	0.081	0.029	2.79	0.033	0.007	4.71
1996	0.634	0.524	1.21	0.082	0.031	2.65	0.035	0.008	4.38
1997	0.630	0.514	1.23	0.089	0.033	2.7	0.038	0.008	4.75
1998	0.615	0.497	1.24	0.089	0.035	2.54	0.042	0.009	4.67
1999	0.602	0.481	1.25	0.094	0.037	2.54	0.048	0.011	4.36
2000	0.590	0.466	1.27	0.095	0.040	2.38	0.050	0.012	4.17
2001	0.580	0.457	1.27	0.093	0.038	2.45	0.065	0.015	4.33
2002	0.587	0.466	1.26	0.097	0.039	2.49	0.059	0.013	4.54

Table 8 continued

Year	Actual share of couples with the same low level of education	Expected share of couples with the same low level of education	Ratio of actual over expected shares among the couples with the same low level of education	Actual share of couples with the same medium level of education	Expected share of couples with the same medium level of education	Ratio of actual over expected shares among the couples with the same medium level of education	Actual share of couples with the same high level of education	Expected share of couples with the same high level of education	Ratio of actual over expected shares among the couples with the same high level of education
2003	0.576	0.453	1.27	0.101	0.041	2.46	0.061	0.013	4.69
2004	0.558	0.432	1.29	0.105	0.045	2.33	0.066	0.015	4.4
2005	0.540	0.410	1.32	0.111	0.048	2.31	0.073	0.018	4.06
2006	0.524	0.393	1.33	0.115	0.053	2.17	0.074	0.019	3.89
2007	0.516	0.386	1.34	0.117	0.054	2.17	0.075	0.019	3.95
2008	0.516	0.380	1.36	0.122	0.055	2.22	0.081	0.020	4.05
2009	0.503	0.368	1.37	0.127	0.059	2.15	0.080	0.021	3.81
2010	0.495	0.355	1.39	0.131	0.063	2.08	0.084	0.022	3.82
2011	0.487	0.346	1.41	0.139	0.066	2.11	0.088	0.023	3.83
2012	0.520	0.381	1.36	0.136	0.061	2.23	0.075	0.017	4.41
2013	0.508	0.370	1.37	0.136	0.062	2.19	0.078	0.019	4.11
2014	0.501	0.362	1.38	0.142	0.065	2.18	0.078	0.019	4.11
2015	0.485	0.346	1.4	0.146	0.070	2.09	0.080	0.021	3.81
2016	0.467	0.328	1.42	0.153	0.075	2.04	0.083	0.022	3.77
2017	0.465	0.323	1.44	0.156	0.077	2.02	0.084	0.023	3.70
2018	0.451	0.309	1.46	0.164	0.082	2.00	0.086	0.024	3.65
2019	0.436	0.291	1.50	0.176	0.089	1.97	0.085	0.025	3.41

Table 9 Decomposition of the theil index into between and within groups non-random mating, with bootstrap 5–95% confidence intervals

Year	Actual Total Theil Index	Lower Bound of Bootstrap Confidence Interval for Total Theil Index	Higher Bound of Bootstrap Confidence Interval for Total Theil Index	Between Groups Theil Index	Lower Bound of Bootstrap Confidence Interval for Between Groups Theil Index	Higher Bound of Bootstrap Confidence Interval for Between Groups Theil Index	Within Groups Theil Index	Lower Bound of Bootstrap Confidence Interval for Within Groups Theil Index	Upper Bound of Bootstrap Confidence Interval for Within Groups Theil Index
1985	0.113	0.106	0.120	0.050	0.045	0.054	0.063	0.060	0.067
1986	0.124	0.116	0.131	0.056	0.051	0.061	0.067	0.064	0.0713
1987	0.121	0.112	0.129	0.058	0.052	0.063	0.063	0.060	0.067
1988	0.15	0.142	0.158	0.0697	0.0645	0.075	0.081	0.076	0.085
1989	0.148	0.141	0.155	0.0688	0.064	0.073	0.07	0.076	0.083
1990	0.144	0.137	0.151	0.0666	0.062	0.071	0.077	0.074	0.081
1991	0.170	0.163	0.177	0.0833	0.079	0.088	0.086	0.083	0.090
1992	0.167	0.160	0.175	0.0819	0.077	0.087	0.085	0.082	0.089
1993	0.172	0.165	0.180	0.084	0.079	0.089	0.088	0.084	0.091
1994	0.167	0.161	0.172	0.0863	0.082	0.0890	0.081	0.078	0.083
1995	0.169	0.164	0.174	0.0887	0.085	0.092	0.081	0.079	0.083
1996	0.171	0.166	0.176	0.0895	0.086	0.093	0.082	0.079	0.084
1997	0.185	0.180	0.190	0.101	0.098	0.105	0.084	0.082	0.086
1998	0.185	0.18	0.191	0.101	0.097	0.105	0.084	0.082	0.087
1999	0.191	0.185	0.197	0.109	0.104	0.113	0.082	0.080	0.08
2000	0.195	0.190	0.201	0.111	0.106	0.115	0.085	0.083	0.088
2001	0.202	0.197	0.207	0.119	0.115	0.123	0.083	0.081	0.085
2002	0.197	0.193	0.203	0.116	0.112	0.120	0.081	0.080	0.083

Table 9 continued

Year	Actual Total Theil Index	Lower Bound of Bootstrap Confidence Interval for Total Theil Index	Higher Bound of Bootstrap Confidence Interval for Total Theil Index	Between Groups Theil Index	Lower Bound of Bootstrap Confidence Interval for Between Groups Theil Index	Higher Bound of Bootstrap Confidence Interval for Between Groups Theil Index	Within Groups Theil Index	Lower Bound of Bootstrap Confidence Interval for Within Groups Theil Index	Upper Bound of Bootstrap Confidence Interval for Within Groups Theil Index
2003	0.200	0.194	0.205	0.119	0.115	0.123	0.081	0.079	0.083
2004	0.202	0.196	0.207	0.124	0.120	0.128	0.077	0.075	0.080
2005	0.211	0.205	0.216	0.134	0.130	0.139	0.077	0.075	0.079
2006	0.209	0.204	0.214	0.134	0.129	0.138	0.076	0.074	0.078
2007	0.205	0.200	0.210	0.133	0.128	0.137	0.073	0.071	0.075
2008	0.223	0.218	0.228	0.148	0.144	0.153	0.075	0.0728	0.076
2009	0.219	0.214	0.224	0.147	0.143	0.151	0.072	0.070	0.074
2010	0.225	0.22	0.232	0.153	0.148	0.158	0.073	0.070	0.075
2011	0.234	0.228	0.239	0.165	0.159	0.169	0.069	0.067	0.072
2012	0.230	0.225	0.235	0.155	0.151	0.159	0.075	0.073	0.077
2013	0.224	0.219	0.229	0.155	0.150	0.159	0.070	0.068	0.072
2014	0.228	0.223	0.233	0.158	0.154	0.163	0.070	0.068	0.072
2015	0.222	0.216	0.227	0.157	0.152	0.162	0.065	0.063	0.067
2016	0.221	0.217	0.227	0.159	0.155	0.164	0.062	0.060	0.064
2017	0.229	0.222	0.234	0.166	0.160	0.170	0.063	0.061	0.065
2018	0.229	0.224	0.236	0.171	0.166	0.176	0.059	0.057	0.061
2019	0.231	0.226	0.236	0.176	0.171	0.180	0.056	0.054	0.057

Table 10 Theil index of non-random mating for the three categories of couples

Year	Theil index of non-random mating within couples with same level of education	Theil index of non-random mating within couples where the wife has a higher level of education than that of the husband	Theil index of non-random mating within couples where the husband has a higher level of education than that of the wife
1985	0.177	−0.0297	−0.034
1986	0.192	−0.0297	−0.0386
1987	0.194	−0.0334	−0.0396
1988	0.210	−0.0269	−0.033
1989	0.208	−0.0254	−0.0352
1990	0.205	−0.0279	−0.0336
1991	0.241	−0.0301	−0.0412
1992	0.238	−0.03	−0.0411
1993	0.247	−0.03	−0.0451
1994	0.247	−0.0334	−0.0468
1995	0.252	−0.0336	−0.0491
1996	0.254	−0.0328	−0.05
1997	0.275	−0.0383	−0.0516
1998	0.276	−0.0398	−0.0505
1999	0.293	−0.0427	−0.059
2000	0.294	−0.0411	−0.0577
2001	0.319	−0.0447	−0.0719
2002	0.316	−0.045	−0.0738
2003	0.321	−0.0463	−0.0753
2004	0.328	−0.0496	−0.077
2005	0.345	−0.0527	−0.0816
2006	0.343	−0.0511	−0.0825
2007	0.342	−0.053	−0.0835
2008	0.366	−0.0569	−0.0867
2009	0.364	−0.0559	−0.0888
2010	0.371	−0.0603	−0.0855
2011	0.390	−0.0648	−0.0911
2012	0.378	−0.0596	−0.0891
2013	0.377	−0.0617	−0.0913
2014	0.383	−0.0633	−0.0915
2015	0.381	−0.065	−0.0937
2016	0.382	−0.0657	−0.0946
2017	0.39	−0.0673	−0.0936
2018	0.396	−0.07	−0.0971
2019	0.400	−0.0691	−0.0992

Table 11 Decomposition of the change in the overall Theil index of non-random mating during the period 1988–1997

Decomposition of changes in the Theil index of non-random mating	Actual values of the Theil index and of the components of its change over time	Lower bound of bootstrap confidence interval (5%)	Upper bound of bootstrap confidence interval (95%)
A-Degree of "non-random mating" in 1988	0.150	0.141	0.159
B-Degree of "non-random mating" in 1997	0.185	0.18	0.191
C = (B−A) = (D+E) Gross variation in "non-random mating"	0.035	0.025	0.046
D-Contribution of variation in the "internal structure of the "mating matrices"	−0.023	−0.032	−0.013
E = (F+G) Contribution of variation in the margins of the "mating matrices"	0.058	0.052	0.064
F-Contribution of changes in the educational structure of the male population (horizontal margins)	−0.014	−0.015	−0.012
G-Contribution of changes in the educational structure of the female population (vertical margins)	0.071	0.065	0.078

Table 12 Decomposition of the change in the overall Theil index of non-random mating during the period 1997–2004

Decomposition of changes in the Theil index of non-random mating	Actual values of the Theil index and of the components of its change over time	Lower bound of bootstrap confidence interval (5%)	Upper bound of bootstrap confidence interval (95%)
A-Degree of "non-random mating" in 1997	0.185	0.180	0.190
B-Degree of "non-random mating" in 2004	0.202	0.197	0.207
C = (B−A) = (D+E) Gross variation in "non-random mating"	0.017	0.009	0.025
D-Contribution of variation in the "internal structure of the "mating matrices"	−0.023	−0.030	−0.016
E = (F+G) Contribution of variation in the margins of the "mating matrices"	0.040	0.036	0.044
F-Contribution of changes in the educational structure of the male population (horizontal margins)	−0.010	−0.012	−0.009
G-Contribution of changes in the educational structure of the female population (vertical margins)	0.050	0.047	0.054

Table 13 Decomposition of the change in the overall Theil index of non-random mating during the period 2004–2012

Decomposition of changes in the Theil index of non-random mating	Actual values of the Theil index and of the components of its change over time	Lower bound of bootstrap confidence interval (5%)	Upper bound of bootstrap confidence interval (95%)
A-Degree of "non-random mating" in 2004	0.202	0.197	0.206
B-Degree of "non-random mating" in 2012	0.230	0.225	0.235
C = (B−A) = (D+E) Gross variation in "non-random mating"	0.028	0.021	0.035
D-Contribution of variation in the "internal structure of the "mating matrices"	0.006	−0.000	0.013
E = (F+G) Contribution of variation in the margins of the "mating matrices"	0.022	0.019	0.025
F-Contribution of changes in the educational structure of the male population (horizontal margins)	−0.005	−0.006	−0.004
G-Contribution of changes in the educational structure of the female population (vertical margins)	0.027	0.023	0.030

Table 14 Decomposition of the change in the overall Theil index of non-random mating during the period 2012–2016

Decomposition of changes in the Theil index of non-random mating	Actual values of the Theil index and of the components of its change over time	Lower bound of bootstrap confidence interval (5%)	Upper bound of bootstrap confidence interval (95%)
A-Degree of "non-random mating" in 2012	0.230	0.224	0.235
B-Degree of "non-random mating" in 2016	0.221	0.216	0.227
C = (B−A) = (D+E) Gross variation in "non-random mating"	−0.008	−0.016	−0.000
D-Contribution of variation in the "internal structure of the "mating matrices"	−0.021	−0.028	−0.014
E = (F+G) Contribution of variation in the margins of the "mating matrices"	0.013	0.010	0.016
F-Contribution of changes in the educational structure of the male population (horizontal margins)	−0.007	−0.008	−0.006
G-Contribution of changes in the educational structure of the female population (vertical margins)	0.020	0.017	0.024

Table 15 Decomposition of the change in the overall Theil index of non-random mating during the period 2016–2019

Decomposition of changes in the Theil index of non-random mating	Actual values of the Theil index and of the components of its change over time	Lower bound of bootstrap confidence interval (5%)	Upper bound of bootstrap confidence interval (95%)
A-Degree of "non-random mating" in 2016	0.221	0.216	0.227
B-Degree of "non-random mating" in 2019	0.231	0.226	0.237
C = (B−A) = (D+E) Gross variation in "non-random mating"	0.010	0.002	0.017
D-Contribution of variation in the "internal structure of the "mating matrices"	0.005	−0.003	0.011
E = (F+G) Contribution of variation in the margins of the "mating matrices"	0.005	0.002	0.008
F-Contribution of changes in the educational structure of the male population (horizontal margins)	−0.005	−0.006	−0.004
G-Contribution of changes in the educational structure of the female population (vertical margins)	0.010	0.007	0.013

9 Appendix B: A simple illustration of the approach of Deming and Stephan (1940)

Assume the following 3 by 3 matrix $a1$ of assortative mating at time 0

$$a1 = \begin{pmatrix} 4 & 7 & 1 \\ 5 & 7 & 3 \\ 2 & 8 & 6 \end{pmatrix}$$

and a matrix $b1$ of assortative mating at time 1, with

$$b1 = \begin{pmatrix} 5 & 1 & 3 \\ 6 & 7 & 9 \\ 10 & 8 & 2 \end{pmatrix}$$

Let us define the following matrices $a2$ and $b2$ of the shares of each category of couple in the total number of couples at times 0 and 1. We get

$$a2 = \begin{pmatrix} 0.093 & 0.163 & 0.023 \\ 0.116 & 0.163 & 0.070 \\ 0.047 & 0.186 & 0.140 \end{pmatrix}$$

and

$$b2 = \begin{pmatrix} 0.098 & 0.020 & 0.059 \\ 0.118 & 0.137 & 0.176 \\ 0.196 & 0.157 & 0.039 \end{pmatrix}$$

It then turns out that the Theil index of a_2 is equal to 0.057 while that of b_2 is 0.081.

The difference between these two indices is hence equal to 0.024.

The horizontal margins of $a2$ and $b2$ are then (0.283 0.340 0.378) for $a2$ and (0.176 0.431 0.392) for $b2$.

Let us now multiply each element on a given line of the matrix $b2$ by the ratio of the horizontal margin of this line in $a2$ over the horizontal margin of this line in $b2$ and call $b4$ the matrix you obtain. It is then easy to check that

$$b4 = \begin{pmatrix} 0.157 & 0.031 & 0.094 \\ 0.093 & 0.108 & 0.139 \\ 0.189 & 0.151 & 0.038 \end{pmatrix}$$

For example, in $b4$, $0.093 = 0.118 \times \left(\frac{0.340}{0.431}\right)$. It is easy to check that the horizontal margins of $b4$ are identical to those of $a2$.

Compute now the vertical margins of $a2$ and $b4$. They are respectively equal to (0.264 0.510 0.227) and (0.439 0.291 0.271).

Multiply now each element on a given line of the matrix $b4$ by the ratio of the vertical margin of this line in $a2$ over the vertical margin of this line in $b4$ and call $b6$ the matrix you obtain. It is then easy to check that

$$b6 = \begin{pmatrix} 0.094 & 0.055 & 0.079 \\ 0.056 & 0.190 & 0.116 \\ 0.114 & 0.265 & 0.032 \end{pmatrix}$$

For example, in 6, $0.265 = 0.151 \times \left(\frac{0.510}{0.291}\right)$. It is also easy to check that the vertical margins of $b6$ are identical to those of $a2$.

Now we start again and multiply, as before, each element of $b6$ by the ratio of the corresponding horizontal margin of $a2$ divided by the corresponding margin of $b6$. Then we continue the procedure working this time with the vertical margins. Deming and Stephan (1940) have then shown that we end up, after only a few iterations, with a matrix b' that will have the horizontal and vertical margins of $a2$ but the "internal structure", that is, the pure degree of assortative mating, of $b2$.

In fact it is easy to find out that

$$b' = \begin{pmatrix} 0.115 & 0.073 & 0.095 \\ 0.050 & 0.186 & 0.104 \\ 0.099 & 0.251 & 0.027 \end{pmatrix}$$

The Theil index of b' turns out to be equal to 0.088. In other words a matrix that would have the internal structure of matrix $b2$ but the margins of $a2$ would have a Theil index equal to 0.088. Given that the Theil index of matrix $b2$ was equal to 0.081 while that of $a2$ was equal to 0.057, we can conclude that this difference of

(0.081–0.057) = 0.024 may be decomposed into the sum of two components. The first one is the difference (0.081–0.088) = −0.007 between the Theil index of $b2$ and that of b' and it reflects the impact of differences in the margins of $a2$ and $b2$ since the matrix b' has the same margins as the matrix $a2$, and the matrices $b2$ and b' have the same internal structure. The second component, equal to the difference (0.088–0.057) = −0.031 between the Theil index of b' and that of $a2$, reflects differences in the "internal structure" ("pure non-assortative matching") of the matrices $a2$ and $b2$, since the matrix b' has the internal structure of the matrix $b2$ but the margins of the matrix $a2$.

The algorithm of Deming and Stephan (1940) allows one to further decompose this difference in the margins of −0.007 into a component due to differences in the horizontal margins (differences in the educational composition of the husbands) and one reflecting differences in the educational composition of the wives (for more details, see, Deutsch et al. 2009).

Naturally we could have started from $b2$ and look for a matrix *a'* that would have the margins of the matrix $b2$ and the "internal structure" of the matrix $a2$. One would then find that

$$a' = \begin{pmatrix} 0.095 & 0.068 & 0.013 \\ 0.216 & 0.114 & 0.101 \\ 0.100 & 0.131 & 0.161 \end{pmatrix}$$

The idea of Shapley decomposition (see, Appendix C and Deutsch et al. 2009, for more details on the latter procedure) is then to combine these two procedures: going from $a2$ to $b2$, and from $b2$ to $a2$.

10 Appendix C: A short summary of the simplest application of the concept of Shapley decomposition

Let $F(a, b)$ be a function depending on two variables, a and b. Such a function need not be linear. Although Chantreuil and Trannoy (1999) and Sastre and Trannoy (2002) limited their application of the Shapley value to the decomposition of income inequality, Shorrocks (1999) has shown that such a decomposition could be applied to any function.

The idea of the Shapley value is to consider all the possible sequences allowing us to eliminate the variables a and b. Let us start with the elimination of the variable a. This variable may be the first one or the second one to be eliminated. If it is eliminated first, the function $F(a, b)$ will become equal to $F[(a = 0), (b \neq 0)]$ since the variable a has been eliminated, so that in this case the contribution of a to the function $F(a, b)$ is equal to $F[(a \neq 0), (b \neq 0)] - F[(a = 0), (b \neq 0)]$. If the variable a is the second one to be eliminated, the function F will then be equal to $F[(a \neq 0), (b = 0)]$. Since both elimination sequences are possible and assuming the probability of these two sequences is the same, we may conclude that the contribution $C(a)$ of the variable a to the function $F(a, b)$ is equal to

$$\begin{aligned} C(a) &= (1/2)\{F[(a \neq 0), (b \neq 0)] - F[(a = 0), (b \neq 0)]\} \\ &\quad + (1/2)\{F[(a \neq 0), (b = 0)] - F[(a = 0), (b = 0)]\} \\ &= (1/2)\{F[(a \neq 0), (b \neq 0)] - F[(a = 0), (b \neq 0)]\} \\ &\quad + (1/2)\{F[(a \neq 0), (b = 0)]\} \end{aligned} \tag{C1}$$

since we assume that $F[(a = 0), (b = 0)] = 0$.

Similarly one can prove that the contribution $C(b)$ of the variable b to the function $F(a, b)$ is

$$\begin{aligned} C(b) &= (1/2)\{F[(a \neq 0), (b \neq 0)] - F[(a \neq 0), (b = 0)]\} \\ &\quad + (1/2)\{F[(a = 0), (b \neq 0)] - F[(a = 0), (b = 0)]\} \\ &= (1/2)\{F[(a \neq 0), (b \neq 0)] - F[(a \neq 0), (b = 0)]\} \\ &\quad + (1/2)\{F[(a = 0), (b \neq 0)]\} \end{aligned} \tag{C2}$$

Combining (C1) and (C2) we observe that

$$C(a) + C(b) = F[(a \neq 0), (b \neq 0)] \tag{C3}$$

Applying Shapley's decomposition to the analysis of variations over time in the value of the Theil index

Using expressions (C1) to (C3) we may express the contribution $C_{\Delta m}$ of the variations of the margins to the overall change ΔI in the extent of non-random mating as

$$\begin{aligned} C_{\Delta m} &= \left(\tfrac{1}{2}\right) f[(\Delta m \neq 0), (\Delta is = 0)] + \left(\tfrac{1}{2}\right) f[(\Delta m \neq 0), (\Delta is \neq 0)] \\ &\quad - \left(\tfrac{1}{2}\right) f[(\Delta m = 0), (\Delta is \neq 0)] \end{aligned} \tag{C4}$$

where Δm and Δis refer respectively to the change in the margins and to that in the internal structure of the original matrix.

Similarly the contribution $C_{\Delta is}$ of the variation in the internal structure of the matrix to the overall change ΔI in the extent of non-random mating will be

$$\begin{aligned} C_{\Delta is} &= \left(\tfrac{1}{2}\right) f[(\Delta m = 0), (\Delta is \neq 0)] + \left(\tfrac{1}{2}\right) f[(\Delta m \neq 0), (\Delta is \neq 0)] \\ &\quad - \left(\tfrac{1}{2}\right) f[(\Delta m \neq 0), (\Delta is = 0)] \end{aligned} \tag{C5}$$

It is easy to observe that $C_{\Delta m} + C_{\Delta is} = \Delta I$

References

Almås, I., A. Kotsadam, E. Moen & K. Røed (2020). The economics of hypergamy, *Journal of Human Resources*. https://doi.org/10.3368/jhr.58.3.1219-10604R1.

Becker, G. (1973). A theory of marriage: Part I. *Journal of Political Economy*, *81*(4), 813–846.

Becker, G. (1974). A theory of marriage: Part II. *Journal of Political Economy*, *82*(2), S11–S26.

Becker, G. S. (1981). *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Bibby, J. (1975). Methods of measuring mobility. *Quality and Quantity*, *9*, 107–136.

Bibby, J. (1980). Measures of social mobility based on income inequality measures. *Quality and Quantity*, *14*, 619–633.

Bourguignon, F., & Chiappori, P.-A. (1992). Collective models of household behavior: an introduction. *European Economic Review*, *36*(2–3), 355–364.

Bratsberg, B., Markussen, S., Raaum, O., Røed, K., & Røgeberg, O. (2018). Trends in assortative mating and offspring outcomes, IZA Discussion Paper No. 11753.
Cao, J., Shi, X., & Shum, M. (2019). On the empirical content of the Beckerian Marriage Model. *Economic Theory*, *67*(2), 349–362.
Chantreuil, F., & Trannoy, A. (2013). Inequality decomposition values: the trade-off between marginality and consistency. *Journal of Economic Inequality*, *11*(1), 83–98.
Chantreuil, F., & Trannoy, A. (1999). Inequality decomposition values: The trade-off between marginality and consistency. THEMA Discussion Paper, Universite' de Cergy-Pontoise.
Chiappori, P.-A., Salanié, B., & Weiss, Y. (2017). Partner choice, investment in children, and the marital college premium. *American Economic Review*, *107*(8), 2109–2167.
Choo, E., & Siow, A. (2006). Who marries whom and why. *Journal of Political Economy*, *114*(1), 175–201.
Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginals are known. *Annals of Mathematical Statistics*, *11*, 427–444.
Deutsch, J, Flückiger, Y., & Silber, J. (2009). Analyzing changes in occupational segregation: the case of Switzerland (1970–2000). In: Y. Flückiger, S. Reardon, J. Silber (Eds), *Research on economic inequality entitled occupational and residential segregation*. (volume 17, pp. 171–202). Bingley, UK: Emerald.
Eika, L., Mogstad, M., & Zafar, B. (2019). Educational assortative mating and household income inequality. *Journal of Political Economy*, *127*(6), 2795–2835.
Esteve, A., García-Román, J., & Permanyer, I. (2012). The gender-gap reversal in education and its effect on union formation: the end of hypergamy? *Population and Development Review*, *38*(3), 535–546.
Fields, G. S. (2008). Income mobility. In L. Blume & S. Durlauf (Eds), *The new Palgrave dictionary of economics*. New York, NY: Palgrave Macmillan.
Fields G. S., & Ok E. A. (1999). The measurement of income mobility: an introduction to the literature. In: Jacques Silber (Ed.), *Handbook on income inequality measurement*, Dordrecht, The Netherlands: Kluwer Academic Publishers.
Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, *69*(1), 9–15.
Galichon, A., & Salanié, B. (2015). Cupid's invisible hand: social surplus and identification in matching models. https://doi.org/10.2139/ssrn.1804623.
Grossbard-Shechtman, S. (1993). *On the economics of marriage: a theory of marriage, labor and divorce*. Boulder, CO: Westview.
Hitsch, G. J., Hortaçsu, A., & Ariely, D. (2010). Matching and sorting in online dating. *American Economic Review*, *100*(1), 130–163.
Jaffe, S., & Weber, S. (2019). The effect of meeting rates on matching outcomes. *Economic Theory*, *67*(2), 363–378.
Kalmijn, M., & Flap, H. (2001). Assortative meeting and mating: unintended consequences of organized settings for partner choices. *Social Forces*, *79*, 1289–312.
Karmel, T., & MacLachlan, M. (1988). Occupational sex segregation—increasing or decreasing. *Economic Record*, *64*, 187–195.
Lundberg, S. J., Pollak, R. A., & Wales, T. J. (1997). Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit. *Journal of Human Resources*, *32*(3), 463–480.
Manser, M., & Brown, M. (1980). Marriage and household decision-making: a bargaining analysis. *International Economic Review*, *21*(1), 31–44.
Mare, R. D. (1991). Five decades of educational assortative mating. *American Sociological Review*, *56*, 15–32.
McElroy, M., & Horney, M. J. (1981). Nash-bargained household decisions: toward a generalization of the theory of demand. *International Economic Review*, *22*, 333–349.
McFarland, D. D. (1972). *Comparison of alternative marriage models*. New York, NY: Academic Press.
Mourifié, I. (2019). A marriage matching function with flexible spillover and substitution patterns. *Economic Theory*, *67*(2), 421–461.
Mourifié, I., & Siow, A. (2014). Cohabitation vs marriage: marriage matching with peer effects, Working paper tecipa-531, University of Toronto, Department of Economics.
Mourifié, I., & Siow, A. (2017). The Cobb Douglas marriage matching function: marriage matching with peer effects, SSRN: https://ssrn.com/abstract=2541895 or https://doi.org/10.2139/ssrn.2541895.
Nakavachara, V. (2010). Superior female education: explaining the gender earnings gap trend in Thailand. *Journal of Asian Economics*, *21*(2), 198–218.

NESDB. (2015). Features of Thai families in the era of low fertility and longevity, National Economic and Social Development Board. Available at https://thailand.unfpa.org/en/publications.

Paweenawat, S.W., & Liao, L. (2018). Educational assortative mating and income inequality in Thailand, PIER Discussion Papers 92, Puey Ungphakorn Institute for Economic Research, revised June 2018.

Paweenawat, S. W., & McNown, R. (2018). A synthetic cohort analysis of female labour supply: the case of Thailand. *Applied Economics*, *50*(5), 527–544.

Pencavel, J. (1998). Assortative mating by schooling and the work behavior of wives and husbands. *American Economic Review*, *88*(2), 326–329.

Prais, S. J. (1955). Measuring Social Mobility. *Journal of the Royal Statistical Society*, *A118*(1), 56–66.

Rose, E. (2005) Education, hypergamy and the 'Success Gap', University of Washington, unpublished manuscript.

Saint-Paul, G. (2015). Genes, legitimacy and hypergamy: another look at the economics of marriage. *Journal of Demographic Economics*, *81*(4), 331–377.

Sastre, M., & Trannoy, A. (2002). Shapley inequality decomposition by factor components: some methodological issues. *Journal of Economics*, *Supplement 9*, 51–89.

Schoen, R., & Cheng, Y.-H. A. (2006). Partner choice and the differential retreat from marriage. *Journal of Marriage and Family*, *68*, 1–10.

Schwartz, C. R. (2010). Earnings inequality and the changing association between spouses' earnings. *American Journal of Sociology*, *115*, 1524–57.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the shapley value. *Journal of Economic Inequality*, *11*(1), 99–126.

Shorrocks, A. F. (1999). Decomposition procedures for distributional analysis: A unified framework based on the Shapley value. Mimeo, University of Essex.

Smits, J., & Park, H. (2009). Five decades of educational assortative mating in 10 east Asian societies. *Social Forces*, *88*(1), 227–256.

Sweeney, M., & Cancian, M. (2004). The changing importance of white women's economic prospects for assortative mating. *Journal of Marriage and Family*, *66*, 1015–28.

Theil, H. (1967). *Economics and information theory*, North-Holland.

Torche, F. (2010). Educational assortative mating and economic inequality: a comparative analysis of three Latin American countries. *Demography*, *47*(2), 481–502.