Economics Letters 234 (2024) 111497

Contents lists available at ScienceDirect

Economics Letters

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

Optimal weights for marital sorting measures[☆]

Frederik Almar[a], Bastian Schulz[a,b,c,*]

[a] Aarhus University, Department of Economics and Business Economics, Denmark
[b] IZA, Germany
[c] CESifo, Germany

ARTICLE INFO

JEL classification:
C43
D10
J11
J12

Keywords:
Positive assortative mating
Marriage market sorting
Homophily
Educational attainment
Sorting measures
Aggregation

ABSTRACT

Changing distributions of male and female types affect the measurement of education-based marriage market sorting. We develop a weighting strategy that minimizes the distortion of sorting measures due to changing type distributions. The optimal weights reflect that female type distributions have changed relatively more in recent decades. Based on our weighted measure, we document increased sorting in Denmark between 1980 and 2018. Alternative measures suggest flat or decreasing trends.

1. Introduction

An increasing tendency of individuals to marry their like in terms of educational attainment, a phenomenon known as positive assortative mating (PAM, sorting), potentially increases inequality between households (e.g., Kremer, 1997; Fernández and Rogerson, 2001; Breen and Salazar, 2011; Chiappori et al., 2020a). However, the literature disagrees on whether PAM has increased (e.g., Greenwood et al., 2016; Eika et al., 2019; Almar et al., 2023). One reason is that shifting distributions of education-based types distort the measurement of sorting. While recent papers acknowledge this (e.g., Liu and Lu, 2006; Eika et al., 2019; Chiappori et al., 2020b, 2021), how this distortion can be compensated for is an open question.

To answer this question, we provide an in-depth analysis of a widely used sorting measure: the weighted sum of likelihood ratios. This measure captures marital sorting by comparing the observed probability that a man of a given type is married to a woman of the same type to that probability under random matching. The likelihood ratios for these same-type couples are aggregated using weights.

We derive optimal weights that minimize the distortion caused by changing type distributions. The optimal weights reflect that educational outcomes of females increased relatively more in recent decades (consistent with, e.g., Goldin, 2006) and eliminate the dominating effect of female-type-distribution changes on the sorting measure. The optimally weighted measure detects increasing PAM. Conventionally weighted measures suggest flat or decreasing trends because they confound increasing PAM with the increase in the supply of highly educated women. We conclude that it is important to take gender-specific trends in the underlying type distributions into account because they matter for conclusions about sorting trends.

2. Data and trends

We use Danish data to illustrate how changing type distributions affect the measurement of PAM. The population register contains demographic variables and person IDs for all residents and their (married or cohabiting) partners (Statistics Denmark BEF, 1980–2018).[1] We study

☆ We thank Audra Bowlus (the editor), Ana Reynoso, Rune Vejlin, and an anonymous referee for constructive comments and inspiring discussions. The Danish register data used in this study were accessed through ECONAU, Department of Economics and Business Economics, Aarhus University. Bastian Schulz thanks the Aarhus University Research Foundation (Grant no. AUFF-F-2018-7-6) and the Dale T. Mortensen Centre, Aarhus University, for financial support.

* Corresponding author. Fuglesangs Allé 4, DK-8210, Aarhus V, Denmark.
E-mail addresses: almar@econ.au.dk (F. Almar), bastian.schulz@econ.au.dk (B. Schulz).

[1] Cohabitation is identified based on a number of criteria: opposite sex, joint children, shared address, less than a 15-year age difference, no family relationship.

https://doi.org/10.1016/j.econlet.2023.111497
Received 3 August 2023; Received in revised form 7 December 2023; Accepted 13 December 2023
Available online 16 December 2023

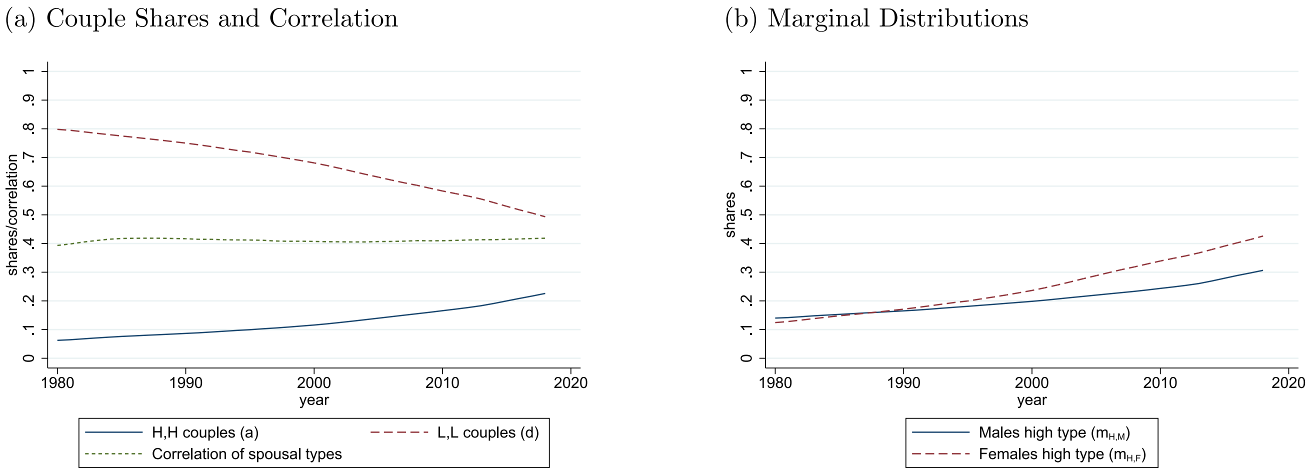

Fig. 1. Assortative matches and marginal distributions.
Note: Panel (a) shows how the shares of couples in which both spouses have either high or low educational attainment, (H, H) or (L, L), have evolved over time, along with the cross-sectional correlation of spousal types. Panel (b) shows the evolution of the fraction of highly educated males and females. Section 2 explains how the sample and education-based types are constructed. The symbols a, d, $m_{(H,M)}$, and $m_{(H,F)}$ are introduced in Section 3 and link the data series to the formal analysis of sorting measures.

the period 1980–2018 and observe on average 1,800,866 individuals in the age range 19–60 per year who are either married to or cohabiting with an individual of the opposite sex. The combined stock of couples is stable over time.[2]

We use the education register (Statistics Denmark UDDA, 1980–2018) to distinguish between highly educated individuals (bachelor's degrees and above, ISCED 6–8) and individuals with lower educational attainment (compulsory schooling, high school, vocational training, short-cycle tertiary programs, ISCED 1–5). Thus, the education-based type T is either H (high) or L (low). Gender is indexed M (male) and F (female).

Fig. 1a shows that the share of (H, H) couples (blue, solid line) increased between 1980 and 2018. Thus, it has become more common to observe couples in which both partners are highly educated. This alone, however, is not evidence of increasing PAM. The share of (L, L) couples (red, dashed line) has decreased at the same time. Moreover, educational attainment has increased, see Panel (b). In 2018, more than 40% (30%) of women (men) are highly educated, compared to approximately 15% in 1980. This shift in the marginal type distributions affects the share of (H, H) couples directly because it became more likely to meet highly educated individuals. Note also that the cross-sectional correlation of couple types in Panel (a) (green, dotted line) is essentially flat. The correlation coefficient conflates changes in couple shares and changes in marginal distributions. We show in Online Appendix A.2 that the correlation responds to such changes in a highly nonlinear way, which makes the trend of the correlation coefficient uninformative about PAM.[3]

In summary, we need a formal framework to measure PAM and disentangle it from changes in the marginal type distributions.

3. Measurement and optimal weights

3.1. The setup

Table 1a is a contingency table that describes the marriage market allocation. The share of couples in which both spouses have high (low) education is $a > 0$ ($d > 0$). Thus, $a + d$ is the share of sorted couples, while the sum of $b > 0$ and $c > 0$ denotes the share of couples with different levels of education. Intuitively, the higher $a + d$ is relative to $b + c$, the more pronounced PAM. To investigate how changing marginals affect sorting measures, we substitute the share of high-type men $m_{(H,M)}$ for $a+b$ and the share of high-type women $m_{(H,F)}$ for $a+c$ in Table 1b.

Table 1
Contingency tables.

(a)

M\F	H	L	Marginal
H	a	b	$a+b$
L	c	d	$c+d$
Marginal	$a+c$	$b+d$	1

(b)

M\F	H	L	Marginal
H	a	$m_{(H,M)} - a$	$m_{(H,M)}$
L	$m_{(H,F)} - a$	d	$1 - m_{(H,M)}$
Marginal	$m_{(H,F)}$	$1 - m_{(H,F)}$	1

3.2. The weighted sum of likelihood ratios

Based on Table 1b, we define the weighted sum of likelihood ratios as follows:

$$I_S = \frac{a}{m_{(H,M)} m_{(H,F)}} \times w_H + \frac{d}{(1 - m_{(H,M)})(1 - m_{(H,F)})} \times w_L. \tag{1}$$

PAM is captured by the ratio of the actual shares of sorted couples and the expected shares based on the "supply" of different types. This measure fulfills the formal criteria for sorting measures outlined by Chiappori et al. (2020b, 2021). It aggregates the shares of sorted couples along the diagonal using the weights w_H and w_L. Chiappori et al. (2020b) suggest that these weights can be thought of as a convex combination of the shares of males and females with the same level of education, which depend on the respective marginal distributions. Let I_S^{convex} denote the measure with these weights applied, where $\lambda \in [0, 1]$ is the coefficient on the male marginal distribution:

$$\begin{aligned} I_S^{convex} = {} & \frac{a}{m_{(H,M)} m_{(H,F)}} \times (\lambda m_{(H,M)} + (1-\lambda) m_{(H,F)}) \\ & + \frac{d}{(1 - m_{(H,M)})(1 - m_{(H,F)})} \times (\lambda (1 - m_{(H,M)}) + (1-\lambda)(1 - m_{(H,F)})). \end{aligned} \tag{2}$$

[2] Figure A.1 in the Online Appendix depicts the evolution of the stocks of couple types and their age composition.
[3] See also (Eika et al., 2019) and Chiappori et al. (2021).

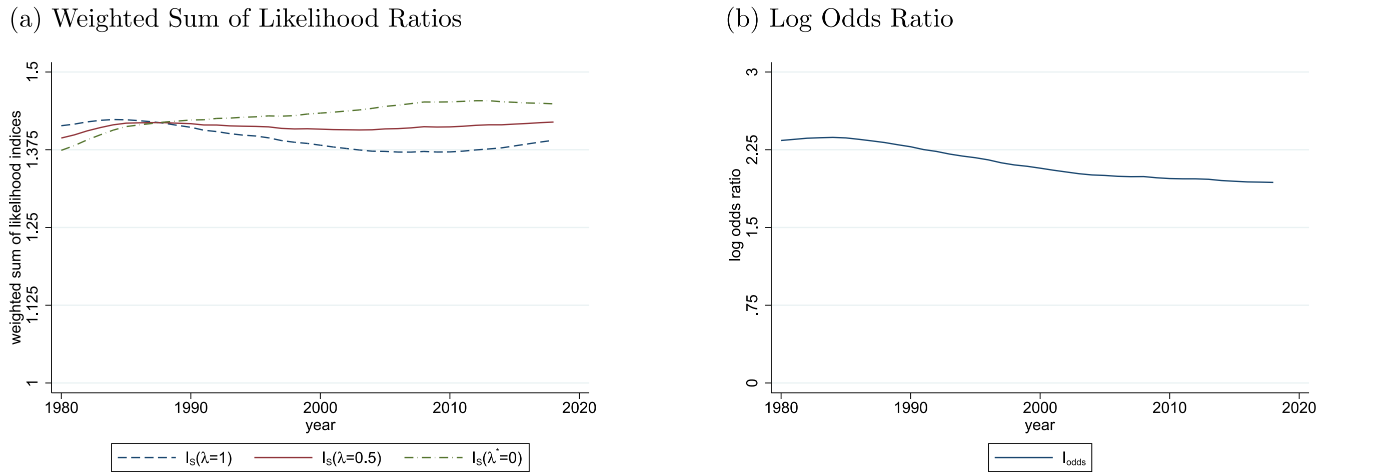

Fig. 2. Sorting trends depend on measures and weights.
Note: Panel (a) depicts the evolution of the weighted sum of likelihood ratios as defined in Eq. (1) for three different values of λ. Panel (b) depicts the evolution of the odds ratio as defined in Eq. (5). Section 2 explains how the sample and the education-based types are constructed.

To investigate the impact of changing shares of sorted couples and marginal distributions, we totally differentiate (2):

$$\begin{aligned}\Delta I_S^{convex} = &\underbrace{\left(\frac{\lambda m_{(H,M)} + (1-\lambda) m_{(H,F)}}{m_{(H,M)} m_{(H,F)}}\right)}_{>0}\Delta a \\ &+ \underbrace{\left(\frac{\lambda(1-m_{(H,M)}) + (1-\lambda)(1-m_{(H,F)})}{(1-m_{(H,M)})(1-m_{(H,F)})}\right)}_{>0}\Delta d \\ &+ \underbrace{(1-\lambda)\left(\frac{d}{(1-m_{(H,M)})^2} - \frac{a}{m_{(H,M)}^2}\right)}_{\gtreqless 0}\Delta m_{(H,M)} \\ &+ \underbrace{\lambda\left(\frac{d}{(1-m_{(H,F)})^2} - \frac{a}{m_{(H,F)}^2}\right)}_{\gtreqless 0}\Delta m_{(H,F)}.\end{aligned} \tag{3}$$

The sorting measure I_S^{convex} is increasing in the shares of sorted couples (a, d) because the coefficients in the first two lines are unambiguously positive. However, the impact of changing marginal distributions is ambiguous and depends on both the configuration of the contingency table and λ. Thus, the choice of λ allows us to account for gender differences in the effect of changing marginals on measured sorting.

We plot the weighted sum of likelihood ratios I_S^{convex} for different values of λ in Fig. 2a. It indicates PAM in all cases because $I_S^{convex} > 1$. However, different values of λ lead to different trends. With weight on changes in the male type distribution ($\lambda = 1$), sorting is decreasing. With weight on changes in the female type distribution ($\lambda = 0$), sorting is increasing. For $\lambda = 0.5$, the trend is flat. Thus, the choice of λ is crucial for conclusions about the trend of PAM.

We propose to choose λ to minimize the impact of marginal-distribution changes on the sorting measure. This can be achieved by setting $\lambda \in [0, 1]$ such that the absolute value of the sum of the $\Delta m_{(H,M)}$ and $\Delta m_{(H,F)}$ terms in Eq. (3) is minimized:

$$\min_{\lambda}\left|(1-\lambda)\underbrace{\left(\frac{d}{(1-m_{(H,M)})^2} - \frac{a}{m_{(H,M)}^2}\right)\Delta m_{(H,M)}}_{=\gamma_1} + \lambda\underbrace{\left(\frac{d}{(1-m_{(H,F)})^2} - \frac{a}{m_{(H,F)}^2}\right)\Delta m_{(H,F)}}_{=\gamma_2}\right|.$$

This objective function is a convex combination of the two endpoints γ_1 and γ_2. If $sign(\gamma_1) \neq sign(\gamma_2)$, then zero lies between the two endpoints and the optimal λ^* solves $(1-\lambda)\gamma_1 + \lambda\gamma_2 = 0$. If, on the other hand, $sign(\gamma_1) = sign(\gamma_2)$, then the optimal λ^* is the endpoint with the smallest absolute value, either $|\gamma_1|$ or $|\gamma_2|$. In summary:

$$\lambda^* = \begin{cases} 0 & \text{if} \quad sign(\gamma_1) = sign(\gamma_2), \quad |\gamma_2| > |\gamma_1| \\ \frac{\gamma_1}{\gamma_1-\gamma_2} & \text{if} \quad sign(\gamma_1) \neq sign(\gamma_2) \\ 1 & \text{if} \quad sign(\gamma_1) = sign(\gamma_2), \quad |\gamma_1| > |\gamma_2|. \end{cases} \tag{4}$$

From Fig. 1, we know that $\Delta m_{(H,F)} > \Delta m_{(H,M)} > 0$ and that in the base year 1980, $m_{(H,M)} \approx m_{(H,F)}$. Thus, $sign(\gamma_1) = sign(\gamma_2)$ for all years, and λ^* must be either zero or one. In the data, $|\gamma_2| > |\gamma_1|$ because of the larger change in the female marginal type distribution. Thus, $\lambda^* = 0$ is optimal for all years.

We know from Fig. 2a that $\lambda^* = 0$ implies increasing sorting. This is because the positive contribution of more sorted high-type couples (term one in Eq. (3) is positive) outweighs the negative contributions from fewer sorted low-type couples (term two in Eq. (3) is negative) and changing marginal type distributions (term three is negative, and term four drops out with $\lambda^* = 0$ in Eq. (3)).[4]

An advantage of the weighted sum of likelihood ratios is that it can be defined for any number of types. In Online Appendix A.3, we generalize the decision rule (4) for more than two types.

3.3. Alternative measures

An alternative measure that also fulfills the formal criteria for sorting measures outlined by Chiappori et al. (2020b, 2021) is the log odds ratio. Based on Table 1, it is defined as follows:

$$I_{odds} = \ln\left(\frac{ad}{bc}\right) = \ln\left(\frac{ad}{(m_{(H,M)} - a)(m_{(H,F)} - a)}\right), \tag{5}$$

where the denominator can be written in terms of the marginal distributions using $m_{(H,M)}$ and $m_{(H,F)}$. As before, we totally differentiate I_{odds}.

$$\Delta I_{odds} = \underbrace{\left(\frac{m_{(H,M)} m_{(H,F)} - a^2}{a(m_{(H,M)} - a)(m_{(H,F)} - a)}\right)}_{>0}\Delta a + \underbrace{\left(\frac{1}{d}\right)}_{>0}\Delta d \tag{6}$$

[4] Totally eliminating the effect of changing marginal distributions on the sorting measure—the sum of terms three and four in Eq. (3)—would require λ^* to be outside the unit interval. The measure would no longer fulfill the monotonicity property stated in Chiappori et al. (2021) because the measure would decrease in the share of sorted couples; see terms one and two in Eq. (3).

$$- \underbrace{\left(\frac{1}{m_{(H,M)} - a}\right)}_{>0} \Delta m_{(H,M)} - \underbrace{\left(\frac{1}{m_{(H,F)} - a}\right)}_{>0} \Delta m_{(H,F)}.$$

Increasing shares of sorted couples a and d imply higher sorting, while increasing shares of high-type individuals $m_{(H,M)}$ and $m_{(H,F)}$ imply lower sorting. Thus, I_{odds} can decrease over time even with increasing shares a and d if the increase in $m_{(H,M)}$ or $m_{(H,F)}$ is sufficiently large.

We plot I_{odds} in Fig. 2b. A log odds ratio > 0 indicates PAM. This sorting measure is indeed decreasing over time. The increasing share of (H, H) couples ($\Delta a > 0$) is dominated by a decreasing share of (L, L) couples ($\Delta d < 0$) and increasing shares of highly educated males and females ($\Delta m_{(H,M)} > 0$, $\Delta m_{(H,F)} > 0$); recall Fig. 1. Note that the coefficients of the male and female high-type shares are symmetric in (6). Therefore, the measure does not allow for gender-specific effects of changing marginals on measured PAM.

An advantage of the (log) odds ratio is that standardization algorithms can be used to generate uniform marginal distributions while preserving the association patterns in the contingency table (Mosteller, 1968). Tan et al. (2004) show that this procedure leaves the odds ratio unchanged, so sorting patterns can be compared over time despite changing type distributions (see, e.g., Greenwood et al., 2014). However, a limitation of the odds ratio is that it is defined for two types only. The weighted sum of likelihood ratios, which can be defined for any number of types, is, however, not invariant to standardization.[5] Our optimal weighting strategy is a viable alternative to standardization because it facilitates the analysis of sorting trends with more than two types and changing marginal distributions.

Finally, Greenwood et al. (2016), Eika et al. (2019), and Almar et al. (2023) use versions of the measure I_S with alternative weights. In Online Appendix A.5, we show that those weights are not necessarily a convex combination of the male and female marginals. The effect of more sorted couples is thus not guaranteed to be positive. In our data, conclusions based on the optimal λ and the alternative weights used in the literature are similar, i.e., sorting has increased. However, this is coincidental and not guaranteed to hold in other settings.

4. Conclusion

We show how to use optimal weights, which compensate for changes in the underlying type distributions, to improve the measurement of education-based marriage market sorting. Because the female type distribution has changed more than its male counterpart in recent decades, attaching the weight to the female side minimizes the distortion of the sorting measure.

We find increasing PAM, while conventionally-weighted measures, which confound increasing PAM with the pronounced increase in the supply of highly educated women, suggest flat or decreasing trends. Thus, the weighting scheme is important, and researchers should use context-dependent weights that are disciplined by the data to study sorting trends.

Data availability

The authors do not have permission to share data.

Online Appendix

Supplementary material related to this article can be found in an Online Appendix at https://doi.org/10.1016/j.econlet.2023.111497.

References

Almar, F., Friedrich, B., Reynoso, A., Schulz, B., Vejlin, R., 2023. Marital Sorting and Inequality: How Educational Categorization Matters. Discussion Paper 15912, IZA Institute of Labor Economics.

Breen, R., Salazar, L., 2011. Educational assortative mating and earnings inequality in the United States. Am. J. Sociol. 117 (3), 808–843.

Chiappori, Pierre-Andre, Costa-Dias, Monica, Crossman, Sam, Meghir, Costas, 2020a. Changes in assortative matching and inequality in income: Evidence for the UK. Fiscal Stud. 41 (1), 39–63.

Chiappori, P.-A., Costa Dias, M., Meghir, C., 2020b. Changes in Assortative Matching: Theory and Evidence for the US. Working Paper 26932, National Bureau of Economic Research.

Chiappori, P.-A., Costa Dias, M., Meghir, C., 2021. The Measuring of Assortativeness in Marriage: a Comment. Cowles Foundation Discussion Paper.

Eika, L., Mogstad, M., Zafar, B., 2019. Educational assortative mating and household income inequality. J. Polit. Econ. 127 (6), 2795–2835.

Fernández, R., Rogerson, R., 2001. Sorting and long-run inequality. Q. J. Econ. 116 (4), 1305–1341.

Goldin, C., 2006. The quiet revolution that transformed women's employment, education, and family. Amer. Econ. Rev. 96 (2), 1–21.

Greenwood, J., Guner, N., Kocharkov, G., Santos, C., 2014. Marry your like: Assortative mating and income inequality. Am. Econ. Rev.: Pap. Proc. 104 (5), 348–353.

Greenwood, J., Guner, N., Kocharkov, G., Santos, C., 2016. Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation. Am. Econ. J.: Macroecon. 8 (1), 1–41.

Kremer, M., 1997. How Much does Sorting Increase Inequality? Q. J. Econ. 112 (1), 115–139.

Liu, H., Lu, J., 2006. Measuring the degree of assortative mating. Econom. Lett. 92 (3), 317–322.

Mosteller, F., 1968. Association and estimation in contingency tables. J. Amer. Statist. Assoc. 63 (321), 1–28.

Statistics Denmark BEF, 1980–2018. BEF - Befolkningen, statistics denmark: population and education, social statistics. https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/the-population.

Statistics Denmark UDDA, 1980–2018. UDDA - Uddannelser (BUE), statistics denmark: Population and education, social statistics. https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/highest-education-attained.

Tan, P.-N., Kumar, V., Srivastava, J., 2004. Selecting the right objective measure for association analysis. Inf. Syst. 29 (4), 293–313.

[5] As we illustrate in Online Appendix A.4, standardization changes the sorting measure I_S, defined in Eq. (1). The reason is that the diagonal elements of the contingency table, which the measure is based on, change while the offsetting changes to the off-diagonal elements are not taken into account.