

Holmlund, Helena

**Working Paper**

## How much does marital sorting contribute to intergenerational socio-economic persistence?

Working Paper, No. 2019:21

**Provided in Cooperation with:**

IFAU - Institute for Evaluation of Labour Market and Education Policy, Uppsala

*Suggested Citation:* Holmlund, Helena (2019) : How much does marital sorting contribute to intergenerational socio-economic persistence?, Working Paper, No. 2019:21, Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala

This Version is available at:

<https://hdl.handle.net/10419/227841>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# How much does marital sorting contribute to intergenerational socio-economic persistence?

Helena Holmlund

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website [www.ifau.se](http://www.ifau.se)

ISSN 1651-1166

# How much does marital sorting contribute to intergenerational socio-economic persistence?<sup>a</sup>

by

Helena Holmlund<sup>b</sup>

26 august 2019

## Abstract

This paper investigates to what extent assortative mating contributes to intergenerational earnings persistence. I use an errors-in-variables model to demonstrate how pooling of partners' 'potential' earnings affects intergenerational earnings persistence, and simulate persistence under different assumptions about assortative mating and women's earnings distribution. Using Swedish data on cohorts born 1945–1965, I show that a substantial decline in marital sorting has contributed little to lowering intergenerational persistence. Variations in marital sorting must be large to affect intergenerational mobility to a great extent. Instead, the relative earnings distributions of men and women, in combination with sorting, are important for intergenerational persistence.

Keywords: assortative mating, intergenerational mobility

JEL-codes: J1; J12; J62; I24

---

<sup>a</sup> The author wishes to thank Adrian Adermon, Matthew Lindquist, Martin Nybom, Håkan Selin and seminar participants at IFAU, University of Hamburg, University of Duisburg-Essen, EALE 2018 and the IZA World Labor Conference 2018 for valuable comments. Financial support from the Swedish Research Council (Vetenskapsrådet) [grant nr 2013-01992] is gratefully acknowledged.

<sup>b</sup> Helena Holmlund, [helena.holmlund@ifau.uu.se](mailto:helena.holmlund@ifau.uu.se). Institute for the Evaluation of Labour Market and Education Policy (IFAU), Uppsala Center for Labor Studies (UCLS), CESifo Munich

## Table of contents

1	Introduction .....	3
2	Conceptual framework .....	6
2.1	Gender differences in earnings distributions .....	9
2.2	Summary of conceptual framework .....	10
2.3	Estimating trends in marital sorting .....	10
3	Data .....	11
4	Trends in assortative mating.....	13
5	Assortative mating and intergenerational persistence among cohorts born in Sweden 1945–1955 .....	17
6	Intergenerational persistence under different assumptions about marital sorting and female labour supply .....	20
7	Conclusions .....	24

## 1 Introduction

Marital sorting is a persistent phenomenon that attracts the interest of economists and sociologists alike. Educational assortative mating may for example play a role in shaping both cross-sectional income inequality and intergenerational income persistence (Mare 1991; Fernández and Rogerson 2001; Fernández, Guner, and Knowles 2005; Ermisch, Francesconi, and Siedler 2006; Mare 2011; Greenwood et al. 2014; Gihleb and Lang 2016; Gonalons-Pons and Schwartz 2017; Eika, Mogstad, and Zafar *forthcoming*). In addition, if mating patterns lead to sorting on genetically transmitted skills and endowments (Mascie-Taylor 1989; Hugh-Jones et al. 2016), assortative mating ultimately also affects the distribution of such endowments in the next generation. As such, assortative mating is a phenomenon that, at least in theory, may have large implications for the development of society and the opportunities of future generations.

In the intergenerational mobility literature, marital sorting has been acknowledged as a potential contributor to intergenerational persistence (Chadwick and Solon 2002; Ermisch, Francesconi, and Siedler 2006; Raaum et al. 2008), but persistence in socio-economic status across generations is still often measured without explicitly distinguishing between the economic resources of the individual as compared to the household.<sup>1</sup> A large number of studies estimate earnings elasticities or correlations between fathers and sons (or daughters), without taking into account that family formation is likely to imply that husbands and wives pool their resources (among many studies, see for example Solon 1992; Björklund and Jäntti 1997; Mazumder 2005; Nybom and Stuhler 2016). Another set of studies uses household earnings in both generations, but does not always explicitly discuss the fact that intergenerational persistence may differ depending on the amount of sorting in household formation (see e.g. Lee and Solon 2009; Chetty et al. 2014; Chetty and Hendren 2018). If we take earnings in the parental generation as given (either fathers' or family earnings), the association with offspring's socioeconomic status, measured by pooled family earnings, will depend on the degree of assortative mating in the offspring generation in combination with the relative earnings distributions of spouses. Intuitively, if both women and men have positive earnings and if couples were formed randomly, the intergenerational association between fathers' earnings and their children's pooled household earnings would be lower than in the case of positive assortative mating on earnings. Marital sorting patterns thus play a role in shaping both cross-sectional income inequality and intergenerational persistence. Put differently, the link between father's earnings and offspring's individual earnings might not be a good representation of persistence in socio-economic status

---

<sup>1</sup> A similar argument is made in the sociology literature, where scholars discuss the the correct way of measuring social class (both origin and destination). Erikson and Goldthorpe (1992) propose the use of the 'dominant' social class at the family level as a relevant definition of a family's class. See also (Beller 2009) for a more recent discussion.

across generations, in particular if marital sorting is low and individuals pool resources with partners of a different ‘type’.

This purpose of this paper is to analyse and quantify to what extent assortative mating in the offspring generation contributes to intergenerational persistence in socio-economic status, by allowing economic status in the offspring generation to be represented by pooled family resources and allowing for different assumptions about women’s contribution to household income. Since earnings are endogenous to the realised match in the marriage market, the paper develops a measure of ‘potential’ earnings based on detailed information on levels and fields of education. I first use an adaptation of an errors-in-rank model (Haider and Solon 2006; Nybom and Stuhler 2016)<sup>2</sup> to demonstrate how marital sorting contributes to the intergenerational correlation in earnings ranks, and estimate the parameters of this model using population-wide Swedish register data. Next, a simulation exercise using complete information on earnings of fathers (generation  $t-1$ ) and potential earnings of husbands and wives in the offspring generation (generation  $t$ ) allows for the estimation of intergenerational rank correlations when invoking different degrees of sorting in the marriage market by generating hypothetical partnerships. I make different assumptions about assortative mating – from random couple formation to complete sorting – and in addition model what happens when we observe changes in women’s earnings distributions. The use of data from Swedish administrative registers allows me to identify partnership formation in adult age and at the same time link individuals to their fathers (and fathers-in-law) and to earnings.

The results are presented in two steps: first, a descriptive analysis shows how marital sorting patterns have evolved over time, for cohorts born 1945–1965 in Sweden. I observe a secular decline in marital sorting – the rank correlation between spouses’ potential earnings dropped from 0.43 (0.45) to 0.33 (0.35) for men (women) over cohorts born 1945–1955. I also demonstrate how the intergenerational rank correlation is affected when we allow offspring’s socio-economic position to be represented by pooled resources and show that the observed decline in marital sorting can explain about 7 (40) percent of the drop in the intergenerational rank correlation for men (women), although these changes are very small in absolute terms. The second part of the paper presents the simulations of intergenerational persistence under different assumptions about assortative mating and female labour supply. The main lessons from the simulation analysis follow a clear intuition, where the sensitivity of mobility to sorting depends on the relative earnings distributions of spouses. Intergenerational persistence among men is relatively insensitive to marital sorting with the currently observed potential earnings of women. However, assuming that women’s earnings increase to the level of men’s, marital sorting becomes

---

<sup>2</sup> See also Böhlmark and Lindquist (2006) for an application.

increasingly more important for men's mobility estimates. In contrast, estimates for women follow the reversed pattern: mobility measures are more sensitive to marital sorting the lower are women's earnings. Overall, elasticities are more sensitive to sorting than rank correlations, since ranks suppress differences in income inequality in the offspring generation induced by different degrees of sorting.

In quantitative terms, observed levels of the intergenerational rank correlation are 12 and 40 percent higher than under random sorting, for men and women respectively. I document a substantial secular decline in sorting among cohorts born in Sweden 1945–1955, but this had only a small impact on the intergenerational rank correlation – instead variations in women's earnings distribution can have large consequences for intergenerational persistence at a given level of sorting. This result is concordant with Eika, Mogstad, and Zafar (*forthcoming*), who demonstrate that increased marital sorting by education has not contributed to higher income inequality; instead changes in returns to education are more important drivers of income dispersion.

The link between assortative mating in the offspring generation and intergenerational earnings persistence is very intuitive but has (to the best of my knowledge) not previously been quantified over the full distribution of sorting. One limitation of previous studies is the endogeneity of observed (female) earnings, which has rarely allowed for similar exercises.<sup>3</sup> Using potential earnings, in combination with different assumptions about female labour supply, I overcome this issue and decompose and quantify the contribution of sorting to intergenerational mobility.

While the intuition relating marital sorting to intergenerational persistence is quite straightforward, the key contribution of the paper lies in quantifying its importance. The findings are informative for intergenerational processes both in societies with high and low female labour force participation (and transitions from one to the other), and also shed light on the mechanisms underlying differences in mobility across regions and time periods.

Marital sorting has been studied extensively in the social sciences, and a number of theoretical frameworks have been used to describe its existence (Becker 1993; Kalmijn 1994; Mare 2011). While factors such as opportunities, preferences and institutions often are suggested as potential mechanisms, it is also worth noting that marital sorting can be seen as endogenous to the degree of inequality in a society (Mare 2011). The greater is inequality, the larger is the 'social gap' between different groups in society, and the larger will be the returns from finding a high-income or high-status spouse. That said, in this paper I provide an empirical decomposition and accounting exercise, without modelling and endogenizing marital sorting. When interpreting

---

<sup>3</sup> Raaum et al. (2008) highlight this issue and approximate the intergenerational transmission of earnings potential by intergenerational correlations in years of schooling. Gonalons-Pons and Schwartz (2017) and Gihleb and Lang (2016) also estimate assortative mating based on potential earnings.



trends it is however important to consider that there may be feed-back mechanisms from trends in intergenerational mobility to trends in marital sorting.

The remainder of the paper is structured as follows. Section 2 presents a conceptual framework that links assortative mating to intergenerational mobility. Section 3 presents the data, and Sections 4–6 proceed to the analysis (description and simulation). Finally, Section 7 summarizes and discusses the main conclusions of the paper.

## 2 Conceptual framework

This section formulates a simple framework to understand the key concepts of the paper and the link between marital sorting and intergenerational mobility. The framework is an adaptation of the work on errors-in-variables by Haider and Solon (2006) and Nybom and Stuhler (2016). The approach is purely empirical and does not provide micro foundations that explain how and why assortative mating arises.

First, let  $\tilde{y}_t$  stand for life-time ‘potential’ earnings of child  $i$ , of gender  $g$  and of generation  $t$ , and let  $\tilde{y}_t^p$  represent the corresponding earnings measure of his/her partner/spouse. Subscripts  $i$  and  $g$  are omitted for simplicity. Let  $y_t = r\tilde{y}_t$  and  $y_t^p = r\tilde{y}_t^p$  represent the ranks of life time ‘potential’ earnings of child  $i$  and his/her partner, ranked separately by gender and cohort, and defined over the interval  $[0,100]$ . Note here that potential earnings conceptually represent the earnings the individual would have if his/her earnings corresponded to the median in cells defined by gender, birth year, county and detailed fields and levels of education (see Section 3 for more details). The rationale for using potential earnings rather than individuals’ actual earnings is that actual earnings are likely to be endogenous to the realised match in the marriage market.<sup>4</sup> Similar concepts measuring individual earnings capacity (exogenous to the marital match) have been used in the previous literature (see e.g. Raaum et al. 2008; Gihleb and Lang 2016; Gonalons-Pons and Schwartz 2017; Bratsberg et al. 2018).

Let us define the marital sorting parameter  $\alpha$  as the rank correlation between the potential earnings rank of an individual and his/her spouse:<sup>5</sup>

$$y_t^p = \mu_g + \alpha_g y_t + u_t; g = men; women \quad (1)$$

$$\alpha_g = \frac{Cov(y_t, y_t^p)}{Var(y_t)} \quad (2)$$

<sup>4</sup> I make the assumption that educational outcomes are exogenous to the marital match.

<sup>5</sup> Linear regression of ranks on ranks yields Spearman’s rank correlation, since the variance of the rank transformation is constant:  $Var(y_t) = Var(y_t^p)$ .

$\alpha$  is defined separately by gender.<sup>6</sup> In the following, I will use the term *assortative mating rank correlation* to describe  $\alpha$ . The rank correlation is bounded between 0 and 1, and  $\alpha = 1$  implies complete sorting on earnings, while  $\alpha = 0$  means that the earnings potentials of spouses are completely uncorrelated. This is the outcome we would expect under random matching of couples.

The recent literature on marital sorting has paid attention to the challenge of measuring trends in sorting, when presumably both marital sorting and the underlying distributions of individual characteristics are changing over time. The most relevant example is sorting by education. The literature proposes various methods to correct for changes in the education distribution; most recently Bratsberg et al. (2018) show that it is not enough to adjust for mechanical changes in matching probabilities between education groups that arise as a consequence of changes to the education distribution – it is also necessary to handle changes in the composition of educational groups. Bratsberg et al. (2018), among others, propose to use rank-based measures which implies that marginal distributions are held constant. This paper follows the tradition of using rank-based measures; the concept of marital sorting should in this context therefore be interpreted as sorting based on individuals' relative positions within the distribution of 'earnings potential'. Given the focus on income mobility, the rank-based sorting measure is particularly useful since it allows us to measure sorting and intergenerational mobility using the same metric.

Next, let  $y_{t-1} = r\tilde{y}_{t-1}$  represent the rank of life time earnings of the father of child  $i$  of generation  $t$ . In the intergenerational mobility literature, earnings mobility across generations is often represented by the rank correlation:

$$y_t = \theta_g + \beta_g y_{t-1} + \epsilon_t \quad (3)$$

$$\beta_g = \frac{Cov(y_t, y_{t-1})}{Var(y_{t-1})} \quad (4)$$

$\beta$  is an estimate of intergenerational earnings persistence between fathers and sons or daughters, with a higher  $\beta$  representing a stronger persistence and concordantly lower intergenerational earnings mobility.<sup>7</sup> However, the child's earnings  $y_t$  might be a noisy measure of his/her actual socio-economic status since most individuals do not remain single but instead live for extended periods of their adult lives with a partner, with whom they share some common goods. In the case of complete pooling of resources within the couple, the relevant measure of economic status will

---

<sup>6</sup>  $\alpha$  may differ between women and men for two reasons. First, there may be differential selection into the marriage market for women and men. Second, women and men of a given cohort are not married to each other, which means that at a given point in time, there is not perfect mirroring of marital sorting between women and men. The former mechanism is shut down in this study by ranking spouses within the distribution of observed spouses of the birth cohort of the index individual.

<sup>7</sup> Note here that  $\beta$  deviates slightly from the traditional interpretation of the intergenerational correlation: offspring's earnings are based on group-level medians and as such  $\beta$  shuts down intergenerational transmission of earnings through contacts and nepotism at the individual level.

be the rank of average household life time potential earnings, here defined as  $z_t = r[0.5(\tilde{y}_t + \tilde{y}_t^p)]$ .

The relationship between  $z_t$  and  $y_t$  can be expressed with the regression slope coefficient  $\lambda$ :

$$z_t = \gamma_g + \lambda_g y_t + \varepsilon_t \quad (5)$$

$$\lambda_g = \frac{\text{Cov}(z_t, y_t)}{\text{Var}(y_t)} \quad (6)$$

where  $\varepsilon_t$  is uncorrelated with  $y_t$  by construction, and  $\lambda \leq 1$  by definition. Intuitively, higher marital sorting implies that individual earnings  $y_t$  will be a better predictor of family earnings  $z_t$ , and  $\lambda$  will be higher. But as we shall explore below,  $\lambda$  is not only a function of the degree of sorting in the marriage market, but also of gender differences in the moments of the underlying potential earnings distributions ( $\tilde{y}_t$  and  $\tilde{y}_t^p$ ).

The focus of this paper is to understand how and to what extent intergenerational persistence is affected by sorting in the marriage market in the offspring generation. To this end, I am interested in the intergenerational rank correlation between household earnings rank  $z_t$  and the corresponding measure of father's earnings  $y_{t-1}$ :

$$\beta_g^* = \frac{\text{Cov}(z_t, y_{t-1})}{\text{Var}(y_{t-1})} \quad (7)$$

In principle, one could also use family earnings in the parent generation (i.e.  $z_{t-1}$ ) but this choice is innocuous, since in this framework earnings (whether fathers' or family earnings) in generation  $t-1$  are already realised, while mobility estimates may be affected by assortative mating and the choice of earnings measure in the offspring generation. Put differently, the paper focuses on and simulates marital sorting on potential earnings in the offspring generation and therefore explores variation in  $z_t$  due to sorting, while sorting in the parent generation is taken as given.<sup>8</sup>

The next step is to describe how  $\beta^*$  is affected by marital sorting. Using (4) and (5),  $\beta^*$  can be written as:

$$\beta_g^* = \lambda_g \beta_g + \left[ \frac{\text{Cov}(y_{t-1}, \varepsilon_t)}{\text{Var}(y_{t-1})} \right]_g \quad (8)$$

Holding constant the underlying potential earnings distributions of women and men,  $\beta^*$  is a function of the individual rank correlation  $\beta$ , by a fraction  $\lambda$  which captures how good individual earnings are as predictors of family earnings, i.e. the degree of sorting on earnings potential in the marriage market.  $\beta^*$  also includes a term which captures the correlation between father's earnings and  $\varepsilon_t$ , the error term in the linear projection of  $z_t$  on  $y_t$ . This latter term captures sorting on earnings between the spouse and the father, net of sorting on earnings between spouses. This type of sorting could arise if individuals search for partners with characteristics similar to their parents,

---

<sup>8</sup> Using  $z_{t-1}$  or  $y_{t-1}$  in the parent generation produces similar results. I choose  $y_{t-1}$  because in the empirical application it is possible to consistently measure fathers' earnings for a longer time period.

or if women's labour supply is very low such that sorting between husbands and their fathers-in-law becomes the relevant margin to express preferences for socio-economic status.

## 2.1 Gender differences in earnings distributions

Equation (8) describes the relationship between marital sorting and intergenerational rank mobility, holding constant the underlying earnings distributions. However,  $\beta^*$  is also sensitive to the moments of the earnings distributions underlying the rank transformation. For a given level of sorting, the properties of  $\lambda$  depend on how the covariance between  $z$  and  $y$  is affected, as the distributions of  $\tilde{y}_t$  and/or  $\tilde{y}_t^p$  change.<sup>9</sup>  $Cov(z_t, y_t)$  is insensitive to changes in means of the underlying distributions, but reacts to changes in the variances of  $\tilde{y}_t$  and  $\tilde{y}_t^p$ . For more detail, Appendix C explores in detail how  $\lambda$  responds to changes in the underlying earnings distributions.

To illustrate how gender differences in earnings distributions affect  $\lambda$ , let us consider two examples: a) complete sorting in the marriage market, and b) less than complete sorting in the marriage market. First, under a), assume that there is complete sorting on potential earnings ( $\alpha = 1$ ). The individual earnings rank will be an exact measure of the household earnings rank ( $z_t = y_t$ ) and  $\lambda = 1$  for both women and men, regardless of their earnings distributions. Under this scenario,  $\beta^* = \beta$  since the error term in the second expression of (5) is zero. In other words, at complete marital sorting, it makes no difference whether measuring the intergenerational rank correlation with individual or family earnings in the offspring generation.

Next we assume, as in b), that there is less than perfect sorting in the marriage market ( $\alpha < 1$ ), and that women's and men's earnings exhibit traditional gender patterns: women have lower average earnings  $E(\tilde{y}_{t,women}) < E(\tilde{y}_{t,men})$  and lower variances  $Var(\tilde{y}_{t,women}) < Var(\tilde{y}_{t,men})$ . At a given level of sorting, the rank transformation implies that  $Cov(z_t, y_t)$  and thus  $\lambda$  are sensitive to differences in the earnings distributions. Women's earnings rank will be a worse predictor of the family earnings rank:  $Cov(z_t, y_t)$  is higher for the spouse whose individual earnings has the highest variance, and  $\lambda_{men} > \lambda_{women}$ . An increase in  $Var(\tilde{y}_{t,women})$  implies that  $\lambda_{women}$  increases, while  $\lambda_{men}$  decreases, all else equal. Women's own earnings rank will become a better predictor of family earnings, and men's earnings rank will lose predictive power.

Using this result, we can predict how the intergenerational rank correlation will behave when we observe changes in the relative earnings distributions of women and men. When the mean and variance of women's earnings increase, for example as a consequence of increased female labour supply, or women gaining access to higher positions in the labour market, and men's distribution remains fixed,  $\beta_{women}^*$  will increase, and  $\beta_{men}^*$  will decrease, at any degree of marital sorting

---

<sup>9</sup> The variance of the percentile ranked transformation is constant and does not affect  $\lambda$ .

below complete sorting. Moreover, since  $\lambda$  converges to one at complete sorting regardless of the characteristics of the earnings distributions, we also have the intuition that the higher is the earnings variance for women, the less sensitive is the intergenerational persistence estimate  $\beta_{women}^*$  to marital sorting, and the more sensitive is  $\beta_{men}^*$ . Put differently, when women's earnings become better predictors of pooled family earnings, marital sorting patterns matter less for women, and more for men. The interplay between marital sorting and gender differences in earnings thus implies that it does make a difference whether we measure intergenerational mobility with individual or family earnings ranks.

## 2.2 Summary of conceptual framework

To summarize, the framework above introduces the following parameters that explain the links between assortative mating and intergenerational income persistence:

$\alpha$	the assortative mating rank correlation in potential earnings
$\beta$	the intergenerational rank correlation in offspring's <i>individual</i> potential earnings
$\beta^*$	the intergenerational rank correlation in offspring's pooled <i>family</i> potential earnings
$\lambda$	the coefficient from a regression of family potential earnings rank on individual potential earnings rank
$\frac{Cov(y_{t-1}, \varepsilon_t)}{Var(y_{t-1})}$	the coefficient from a regression of residuals from equation (5) on father's earnings rank.

On a final note, the conceptual framework applies also to the log transformation, with one exception:  $\lambda$  is responsive both to the mean of  $\tilde{y}_t$  and to  $Var(\tilde{y}_t)$ , and there is no unambiguous prediction of how  $\lambda$  will react to a joint shift in both mean and variance.<sup>10</sup> Results for the simulation analysis using the log transformation, i.e., the intergenerational elasticity, are reported in Appendix A.

## 2.3 Estimating trends in marital sorting

As explained above, empirically observed time trends in  $\lambda$  can be the result of changes in marital sorting patterns, but can also originate from changes in the earnings distributions, holding constant the level of sorting. In order to adequately isolate how trends in assortative mating contribute to

---

<sup>10</sup> To exemplify, again assume gender differences in average earnings:  $E(\tilde{y}_{t,men}) > E(\tilde{y}_{t,women})$ . Intuitively, men's log earnings are better predictors of log family earnings, and we expect  $\lambda_{men} > \lambda_{women}$ . A closing of the gender earnings gap in terms of means implies that  $\lambda_{men}$  and  $\lambda_{women}$  converge, as  $\lambda_{men}$  decreases and  $\lambda_{women}$  increases. However, changes in earnings distributions are likely to affect both means and variances, and will have ambiguous effects on  $\lambda$ . The effects on  $\beta^*$  are less clear when considering also changes in the variances. See Appendix C for details.

intergenerational persistence it is therefore necessary to hold constant changes in the underlying earnings distributions. This can be achieved by imputing the same distribution of potential earnings to all cohorts, based on cohort-specific ranks in potential earnings (see Section 3 for details). The assumption of constant earnings distributions is relaxed in section 6.

### 3 Data

The paper uses data compiled from a number of registers held by Statistics Sweden. The multi-generation register identifies individuals residing in Sweden at some point after 1960, and all their parents and children. The population of interest, individuals born in Sweden between 1945 and 1965, is extracted from this register. The spouses/partners of these individuals are identified as the other biological parent of an individual's first-born child. In the Swedish context, where it is common that partners live together and have children without being formally married, this is the most reasonable definition of a spouse. This limits the analysis to couples with children, but given that the study of assortative mating is motivated by its potential consequences for income inequality and intergenerational transmission, it makes sense to study couples that are most likely to pool resources, that is, families with children.

After defining the population of interest and identifying spouses, tax register information on earnings of the individuals in the population, their spouses, parents and parents-in-law is merged to the data. The sample is restricted to couples with an age difference of a maximum of ten years, which allows me to observe completed education and mid-life earnings for both the oldest (born in 1935) and the youngest (born in 1975) spouses. Below I list the main variables used in the analysis.

*Offspring's potential earnings.* Earnings are derived from tax registers and represent the average of annual earnings when individuals are in the age range 40–45. Earnings include labour earnings and income replacement in the social insurance system related to illness and parental leave. Unemployment and social welfare benefits are not included. Using this earnings measure I make a prediction, based on the median earnings in cells defined by gender, birth cohort, county of residence in adolescence, and detailed levels and fields of education in the education register (about 3,000 different categories for each gender)<sup>11</sup>. Median earnings have previously been used to impute missing female earnings, see e.g. Neal (2004); Albrecht, van Vuuren, and Vroman (2015). In this setting, the median is preferred to the mean since it is more likely to represent full-

---

<sup>11</sup> I use the education register from 2013 or earlier if the individual is missing in 2013. The earliest education register that I use is from 2001. Due to a change in the education code in 2000, I am not able to use earlier versions of the register. This may be problematic for the oldest cohorts in my sample since some individuals might have died by 2001. However, using the earlier registers from the 1990s gives very similar trends in sorting and mobility for the earliest cohorts in my sample (but is unreliable for younger cohorts).

time labour supply for women.<sup>12</sup> For each birth cohort born 1935–1975 the prediction is based on a moving five-cohort window to allow enough observations in each cell.<sup>13</sup>

Cohort-specific predictions will generate earnings distributions that vary by cohort, for three reasons. First, if median earnings do not represent full-time labour supply, changes in labour supply over time will shift the potential earnings distribution. This is primarily an issue for women, whose labour supply has increased over the cohorts in the sample. Second, the distributions will also respond to changes in the education distribution (both levels and fields), and finally, to changes in the returns to education. When decomposing how much intergenerational mobility depends on marital sorting, it is (as explained in Section 2) necessary to separate changes to  $\lambda$  that are due to shifts in the relative earnings distributions of spouses, and that are due to differential marital sorting, holding constant the underlying earnings distributions. Below I explain how I proceed to separate the two mechanisms.

First, I generate a measure of ‘potential earnings’ that holds the moments of the earnings distribution constant over time. In other words, using ‘potential earnings’, I shut down any changes to  $\lambda$  that can be attributed to e.g. increased female labour supply over time. Using the cohort-specific earnings prediction (based on education, gender and county of residence as explained above), individuals are given a percentile rank by gender and cohort. To each gender- and cohort-specific percentile, I assign the mean earnings values associated with the corresponding percentiles in the distribution of *actual* earnings of the 1955 cohort. This allows me to hold the moments of the distribution constant over time, and also implies that I assess the role of sorting using a distribution that preserves the mean and variance of actual earnings, rather than the imputed distribution which exhibits much less variation.<sup>14</sup> Using this earnings measure, estimated trends in  $\lambda$  will isolate to what extent changes in sorting has affected intergenerational persistence.

Second, to demonstrate the interplay between sorting and the relative earnings distributions of spouses, in Section 6 I relax the assumption of constant earnings distributions and model the sorting-mobility relationship under different assumptions about women’s earnings distribution.

In Appendix B I present a validation analysis which provides more background and statistics related to the measure of potential earnings. It shows a) that the correlations between potential

---

<sup>12</sup> Calculations by Anna Sjögren based on Swedish women born in 1974 show that the fraction of full-time work is above 50 percent for women throughout women’s income distribution.

<sup>13</sup> I use tax registers covering the years 1985–2012. This means that for individuals born before 1945 and after 1968, I cannot observe earnings at age 40–45 and earnings will be observed later or earlier in life. The earliest earnings observations used are for the 1975 cohort, which are observed at age 34–37. For both fathers and offspring, earnings are expressed in 2006 values. See Appendix B for further sensitivity analysis.

<sup>14</sup> To fully hold distributions constant over time, ranks are taken by cohort and gender within the population of interest, i.e., individuals with a partner (i.e. those who have at least one child), and assigned to values of the 1955 distribution based on their rank in the estimation sample. Ranks in pooled household earnings are also computed by gender and cohort of the unit of observation (i.e. the index person born in 1945–1965)

earnings and actual earnings are stable over cohorts born 1935–1975, and b) that measures of intergenerational persistence (for men) exhibit similar trends over time using actual and potential earnings. These two findings support the use of potential earnings as an alternative to actual earnings.

*Father's earnings.* Father's earnings are observed in 1968, 1971, 1973, 1976, 1979, 1982 and 1985. I use averages over three earnings observations in years when fathers are aged 37–45. For older fathers, born before 1928, I observe earnings at an older age. Percentile ranks of father's earnings are obtained by ranking each father's average earnings within his child's cohort.<sup>15</sup>

*Sample restrictions.* The sample is restricted to individuals born in Sweden, since earnings histories for fathers of immigrated individuals to a large extent are incomplete. The age difference between partners is restricted to a maximum of ten years. In order to analyse intergenerational persistence, further restrictions apply: the fathers must be between 15 and 45 years of age at the time of birth of an individual. This is to ensure that the sample has the same age structure over time.<sup>16</sup>

Appendix Table A1 provides summary statistics of the estimation sample, based on cohorts born 1945–1965. Actual earnings exhibit higher means and variances than potential earnings (imputed with values of the 1955 cohort), which is explained by the real increase in earnings over time. By construction, there are no differences in means between actual and potential earnings of the 1955 cohort (not shown in table), but there is a small difference in the variance for men, which is attributed to the fact that imputing mean earnings of the top 1 percent of the distribution removes extreme top earnings observations and reduces the variance.<sup>17</sup>

#### 4 Trends in assortative mating

This section presents evidence on the development of assortative mating on potential earnings for cohorts born in Sweden 1945–1965. Figure 1a shows the assortative mating rank correlations by year of birth for men and women, respectively. Note here that men's spouses in Figure 1 are not fully overlapping with the population of women in Figure 1 (and vice versa for women's spouses): individuals born in 1945–1965 can form families with individuals born before 1945 or after 1965. However, spouses are ranked within the observed population of spouses, by cohort of the index individual. This means that the trends are not affected by differential selection into the marriage

---

<sup>15</sup> Results are unchanged when ranking fathers within their own cohort.

<sup>16</sup> Holding constant the age structure of the sample is important when estimating trends. Given that earnings are measured at the earliest in 1968, including very old fathers would introduce many missing observations among the early cohorts. The restriction on age at birth thus implies that the sample is comparable over time.

<sup>17</sup> As shown in Appendix Table A1, the standard deviation of men's *potential* earnings in the 1955 cohort is 163,773. The corresponding standard deviation in the *actual* earnings distribution of the 1955 cohort is 169,417. The difference is explained exclusively by extreme values in the top percentile of the distribution.



market over time (and by gender). The corresponding figure for the elasticity of spouses' earnings can be found in Appendix Figure A1.

The first observation to be made is the level of assortative mating: starting with the cohort born in 1945, the rank correlations are about 0.43 for men and 0.45 for women. One percentile higher potential earnings is on average associated with having a spouse with 0.43–0.45 percentile points higher potential earnings. There is clear evidence of positive marital sorting on potential earnings.

The second observation is that there is a negative trend in assortative mating over time in the first half of the study period. The trend indicates that among the observed cohorts, it has become more common to find a partner from a different position in the potential earnings distribution. This pattern holds for both women and men and indicates a relatively large drop, from 0.43 to 0.33 for men born 1945–1955, and from 0.45 to 0.35 for women from the same birth cohorts. The negative trend in sorting is further corroborated by sorting on i) years of schooling and ii) fathers' earnings rank, shown in Appendix Figure A2. This decline in marital sorting over birth cohorts in Sweden has previously been documented in Boschini et al. (2011), and is also in line with recent evidence from Norway which shows declining overall sorting using rank based measures (Bratsberg et al. 2018). While it is possible that the decline in part could be attributed to mechanical changes in potential earnings ranks related to changes in compulsory schooling laws (see Meghir and Palme 2005), the similar trend in sorting by fathers' earnings supports the interpretation that assortative mating by socio-economic status declined.<sup>18</sup>

Figure 1b presents the corresponding trends in  $\lambda$ , that is, the coefficients from regressions of rank of pooled household earnings on own earnings rank. As expected,  $\lambda$  is higher for men than for women since men's earnings exhibit higher variance. The decline in sorting observed in Figure 1a also translates into a reduction of  $\lambda$  over time: from 0.89 to 0.87 for men born 1955, and from 0.77 to 0.72 for women born 1955. In other words, lower marital sorting implies that own earnings rank becomes a worse predictor of household earnings rank. We also see that the drop in  $\lambda$  is larger for women, which is explained by the intuition that sorting is likely to have a greater impact on women's  $\lambda$  because of women's lower earnings variance (see Table A1 for descriptive statistics). Note that this drop is not driven by changes in female labour supply over time, since the correlations use potential earnings, holding constant the underlying earnings distributions over time. Instead, the trends should be attributed to reduced sorting by rank in earnings capacity.

---

<sup>18</sup> Appendix Figures A2a and b however show diverging trends in the second half of the study period, where sorting on years of schooling increased but sorting on fathers' earnings rank continued to decrease. This pattern suggests that the increased sorting on offspring education may have been a result of increased social mobility where individuals from lower income households gained access to higher education, and also faced a wider set of marriage options in terms of family background. These two trends seem to offset each other when focusing on offspring's potential earnings (as in Figure 1), which essentially combines detailed information on education and home county to impute earnings.

Figure 1 is informative regarding the general trend, but remains silent about the underlying patterns of sorting. To shed light on this, Figure 2 presents 5-percent binned plots of average potential earnings ranks of partners over own potential earnings ranks, for women and men in the 1945 and 1955 cohorts. The figures also include fitted lines from kernel regressions. The figures indicate that over the observed decade, men and women with low earnings potential have become more likely to marry up, while there is a tendency for women at the upper part of the potential earnings distribution have become more likely to marry down.<sup>19</sup>

---

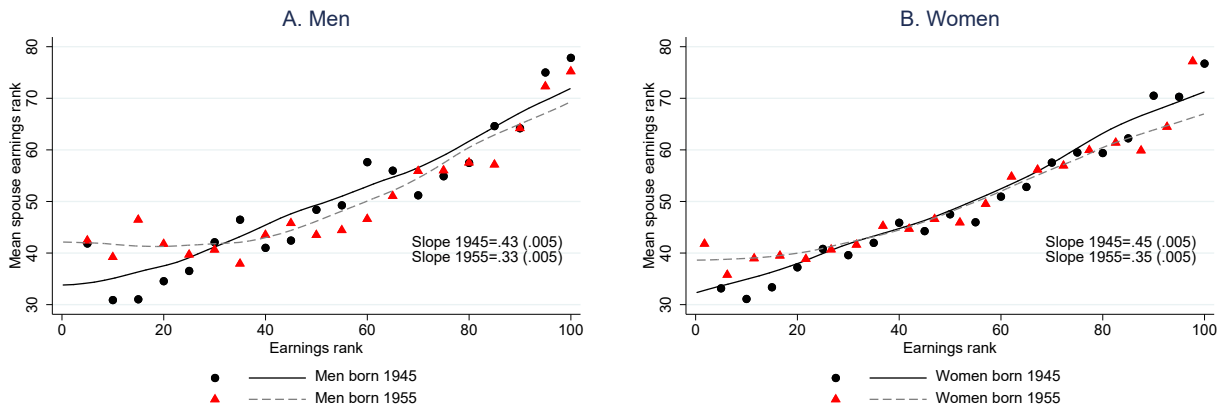
<sup>19</sup> This result is somewhat different to Boschini et al. (2011), who show that sorting on years of education decreased for the same Swedish birth cohorts, and that it decreased the most at the top of the distribution.

Figure 1 Trends in assortative mating on earnings ranks, cohorts born 1945–1965



*Note:* The figure is based on cohorts born in Sweden. In panel A, assortative mating is measured by the rank correlation in potential earnings. Potential earnings are predicted using detailed information on fields and levels of education (see Section 3 for details). Panel B shows regression coefficients of linear projections of rank of household potential earnings on rank of individual's potential earnings ( $\lambda$ ).

Figure 2 Binned plots and kernel regressions of spouse's earnings rank over own earnings rank, cohorts born 1945 and 1955



*Note:* Binned plots are based on 5 percent bins. Fitted lines are based on local linear kernel regressions using an Epanechnikov kernel with bandwidth 10. The x-axis shows the income rank of the index individual and the y-axis shows the average income rank among spouses.

To put the decline in marital sorting over time into context, it is worth noting that these cohorts were exposed to many reforms that built the foundations of the welfare system, such as the expansion of compulsory education from 7 to 9 years, increased opportunities for individuals with poor backgrounds to achieve higher education, more generous family policies, and in turn increased labour force participation among women (see e.g. Meghir and Palme (2005) and Björklund (2006) for descriptions of some of these policies). It is thus possible that these reforms, potentially in combination with a change of norms, affected the return to marrying an individual from a similar socio-economic background or with similar characteristics, or that preferences regarding homophily changed as norms on equality of opportunity grew stronger.

In the next section we take this question one step further, and ask to what extent the drop in assortative mating observed in Figures 1 and 2 can predict a reduction in intergenerational mobility.

## 5 Assortative mating and intergenerational persistence among cohorts born in Sweden 1945–1955

To begin the empirical investigation of the relationship between marital sorting and intergenerational mobility, Table 1 presents the parameters that were discussed in Section 2. The table is organised into four panels, presenting evidence for women and men born in 1945 and 1955, thus capturing the secular decline in marital sorting observed in Figure 1. The first two columns show estimates of the assortative mating rank correlation ( $\alpha$ ) and the linear projection of rank of pooled household earnings on rank of individual earnings ( $\lambda$ ), which are equivalent to those presented in Figure 1. The remaining columns show the regression coefficient that captures sorting between spouse's earnings residual and father's earnings rank (column 3), and the intergenerational rank correlations in individual earnings ( $\beta$  – column 4) and pooled family earnings ( $\beta^*$  – column 5), respectively. From Table 1 it is possible to decompose  $\beta^*$  into its two components: for men born in 1945, about 87 percent can be attributed to  $\beta$  through sorting on earnings ranks between partners, and the remaining 13 percent can be attributed directly to intergenerational sorting between wives and fathers.<sup>22</sup> For women, the corresponding decomposition shows that 70 percent of  $\beta^*$  can be attributed to  $\beta$  via sorting between partners, and the remaining part is due to sorting between husbands and fathers.

All else equal, a drop in  $\lambda$  should imply a decline in the intergenerational rank correlation based on household earnings. From column 5 it is clear that  $\beta^*$  is reduced slightly over time, from 0.37 to 0.34 for men, and from 0.34 to 0.30 for women. However, column 4 shows that also  $\beta$  declined over time, which also has contributed to lowering  $\beta^*$ . If we make the thought experiment of holding  $\beta$  and sorting between spouse and father (columns 3 and 4) constant, the drop in marital sorting observed over the 1945–1955 birth cohorts would imply a reduction of the rank correlation  $\beta^*$  by 0.006 for men and 0.016 for women. As such, the decline in marital sorting has lowered the intergenerational rank correlation fairly little, by about 1.6 percent for men, and 4.5 percent for women.<sup>23</sup> Put differently, assortative mating can explain about 7 percent of the observed 1945–1955 decline in  $\beta^*$  for men. For women, instead, 40 percent of the reduction in  $\beta^*$  can be attributed to the decline in marital sorting.

---

<sup>22</sup> This decomposition is based on the following calculation:  $(0.888 \cdot 0.359) / 0.366 = 0.87$ .

<sup>23</sup>  $((0.888 - 0.872) \cdot 0.359) / 0.366 = 0.016$ .

This empirical exercise sheds light on whether the decline in marital sorting observed in Sweden has contributed to lowering intergenerational persistence, and shows that a rather large reduction in marital sorting as estimated by the rank correlation in spouse's potential earnings, has had only a small impact on the intergenerational rank correlation in absolute terms. As will be clear in the next section, marital sorting will play a different role depending on the earnings distributions of women and men, and can under certain assumptions have large implications for the intergenerational rank correlation.

Table 1 Assortative mating and intergenerational rank correlations for Swedish men and women born 1945 and 1955

	(1)	(2)	(3)	(4)	(5)
	AM rank corr. $\alpha$		Coef. resid - father's rank $\lambda$	Intergen ind. rank corr. $\beta$	Intergen pooled rank corr. $\beta^*$
<i>Outcome variable:</i>					
	Spouse earnings rank	Family earnings rank	Residual from col (2)	Individual earnings rank	Family earnings rank
<b>Men born 1945</b>					
Individual earnings rank	0.430*** (0.005)	0.888*** (0.002)			
Father's earnings rank			0.047*** (0.002)	0.359*** (0.005)	0.366*** (0.005)
Observations	33,852	33,852	33,852	33,852	33,852
R-squared	0.186	0.790	0.011	0.129	0.134
<b>Men born 1955</b>					
Individual earnings rank	0.328*** (0.005)	0.872*** (0.002)			
Father's earnings rank			0.059*** (0.003)	0.320*** (0.005)	0.338*** (0.005)
Observations	31,029	31,029	31,029	31,029	31,029
R-squared	0.107	0.757	0.014	0.102	0.114
<b>Women born 1945</b>					
Individual earnings rank	0.445*** (0.005)	0.769*** (0.003)			
Father's earnings rank			0.102*** (0.003)	0.312*** (0.005)	0.342*** (0.005)
Observations	32,083	32,083	32,083	32,083	32,083
R-squared	0.200	0.597	0.026	0.096	0.117
<b>Women born 1955</b>					
Individual earnings rank	0.347*** (0.005)	0.719*** (0.004)			
Father's earnings rank			0.096*** (0.004)	0.288*** (0.005)	0.303*** (0.005)
Observations	30,883	30,883	30,883	30,883	30,883
R-squared	0.120	0.516	0.019	0.082	0.091

*Note:* Rank correlations are calculated using ranks in potential earnings in the offspring generation, and ranks in actual earnings of fathers. Robust standard errors in parentheses \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## 6 Intergenerational persistence under different assumptions about marital sorting and female labour supply

This section simulates intergenerational mobility under different assumptions about assortative mating and female labour supply. Given the nature of the data, which allows me to link individuals to their partners and parents (and parents-in-law) and to impute potential earnings, it is possible to assess the implications of marital sorting for intergenerational persistence by generating hypothetical partnerships and then simulate different degrees of sorting in the marriage market. In the simulations, the interplay between marital sorting and gender differences in earnings distributions is investigated by varying the earnings distribution of women. Table 2 presents the earnings distributions used in the simulations, which are based on the 1955 cohort. As expected, the observed distributions display traditional gender patterns where women's distribution has lower mean and variance than men's. The bottom two rows show the two hypothetical distributions, where in one case women's distribution is identical to men's, and in another women's distribution is more compressed and at a lower mean, reflecting a situation with lower female labour supply and/or a different occupational or educational structure.

Table 2 Potential earnings distributions

	(1) Mean	(2) St.Dev.	(3) Min	(4) Max
<i>Potential earnings distributions</i>				
Men observed	259,721	164,210	0.00	1,773,587
Women observed	179,287	92,910	0.00	779,945
Women using men's values	258,162	162,832	0.00	1,773,587
Women ½ of observed	89,643	46,455	0.00	389,973

The exercise entails estimating  $\lambda$  and  $\beta^*$  over marital sorting  $\alpha \in [0,1]$ . I maintain the assumption that  $Cov(y_{t-1}, \varepsilon_t) = 0$ , which means that I only model sorting on potential earnings of spouses, not additional sorting between the partners and their fathers-in-law. The sorting algorithm is presented in detail in Appendix D. The basic idea is to generate hypothetical partnerships using the pool of observed couples, by letting potential earnings of the spouse be represented by a weighted combination of potential earnings under complete and random sorting, and as such allow for different degrees of sorting between  $y_t$  and  $y_t^p$ .<sup>24</sup> This allows me to estimate  $\beta^*$ , the intergenerational rank correlation, under different degrees of marital sorting.

Figures 3 and 4 show simulations of  $\lambda$  and  $\beta^*$  under three different scenarios. Figure 3a shows that for men,  $\lambda$  is relatively insensitive to changes in marital sorting. At observed potential

<sup>24</sup> I make the simplified assumption that there are no exits or new entrants on the marriage market. That is, the simulation does not model what would happen if the unmarried would enter, and married individuals would exit the marriage market. Instead, I assume that for a given birth cohort  $t$ , the set of potential partners consists only of those already observed to have matched with an individual in cohort  $t$ .

earnings distributions, a  $\lambda$  of 0.79 at random sorting and observed labour supply shows that men's earnings dominate women's and therefore men's individual earnings are good predictors of pooled household earnings at any level of sorting. As a consequence of little variation in  $\lambda$ , the intergenerational rank correlation of men is relatively insensitive to marital sorting (see Figure 4a). Going from random sorting to complete sorting between spouses moves the rank correlation only from 0.25 to 0.32. If we instead assume that women earn as much as men ("women high potential earnings"), sorting has larger consequences for  $\lambda$ , and in turn  $\beta^*$  becomes more sensitive to marital sorting. However, moving from 0.22 to 0.32 when going from random to complete sorting is still a relatively modest change, given that we are comparing extremes.

The shaded areas of the figures highlight the range of  $\lambda$ 's observed in the 1945 to 1955 cohorts (see Table 1), and the simulated intergenerational rank correlations corresponding to these values of  $\lambda$ . By comparing the simulated intergenerational rank correlation at  $\lambda$  in 1955 with the rank correlation at random sorting, we find that men's  $\beta^*$  is 12 percent higher than what would have been the case under random sorting.<sup>25</sup> However, the variation in  $\lambda$  observed over the 1945–1955 cohorts implies a very small change in  $\beta^*$ , as shown by the variation in  $\beta^*$  within the shaded area.

Figures 3b and 4b display the corresponding results for women. Women's individual earnings are not as good predictors of pooled household earnings as are men's, as demonstrated by lower levels of  $\lambda$  at random sorting and consequently a steeper slope of  $\lambda$  as sorting increases. Marital sorting plays a larger role in explaining intergenerational persistence of women as compared to men: when moving from random to complete sorting,  $\beta^*$  almost doubles from 0.15 to 0.29. Comparing the simulated rank correlation at observed  $\lambda$  in 1955 with the scenario under random sorting, women's  $\beta^*$  is 40 percent higher than under random sorting, but the variation in  $\lambda$  over the 1945–1955 cohorts (indicated by the shaded area) has a modest impact on  $\beta^*$ .<sup>26</sup>

When comparing the different scenarios with respect to women's earnings distributions, we observe that the lower is women's earnings variance, the more sensitive is the intergenerational rank correlation to marital sorting. Intuitively, when the variation of women's earnings is lower than that of men's, her earnings will contribute less to the rank of household earnings, and the choice of partner will matter more for the intergenerational persistence between women and their fathers. Under the assumption of low potential earnings, the intergenerational rank correlation ranges from 0.09 to 0.29, which points to substantially different levels of mobility.

Quantifying the contribution of sorting to the intergenerational rank correlation as in Figure 4 shows that variations in marital sorting must be large in order to have economically meaningful consequences for intergenerational mobility in terms of ranks. Instead, for a given level of sorting,

---

<sup>25</sup>  $(0.28-0.25)/0.25=0.12$ .

<sup>26</sup>  $(0.21-0.15)/0.15=0.4$ .

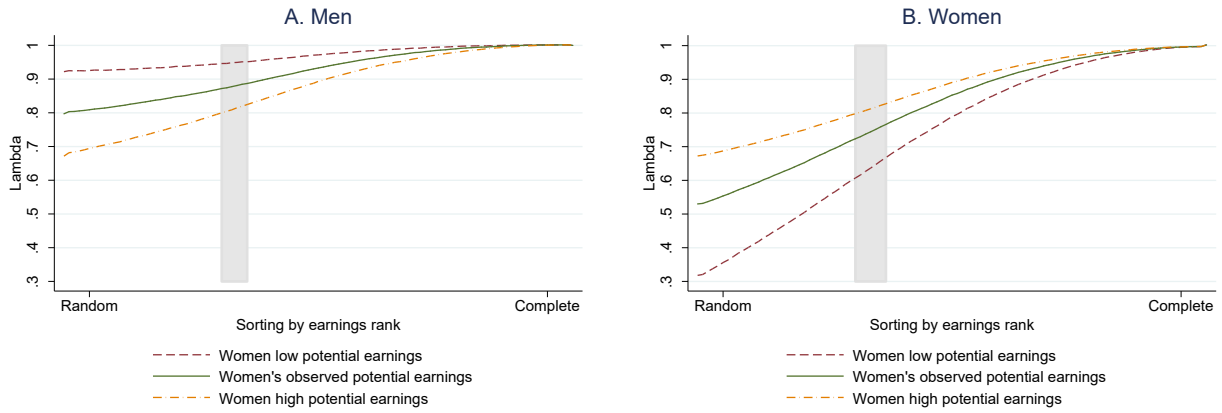


in particular at the lower end, differences in the relative earnings distributions of women and men imply substantial shifts in the level of intergenerational persistence. This result squares well with the recent literature on assortative mating and income inequality, which points to the importance of returns to education rather than increased sorting as explanations for increasing income inequality (Gonalons-Pons and Schwartz 2017; Eika, Mogstad, and Zafar, *forthcoming*).<sup>27</sup>

---

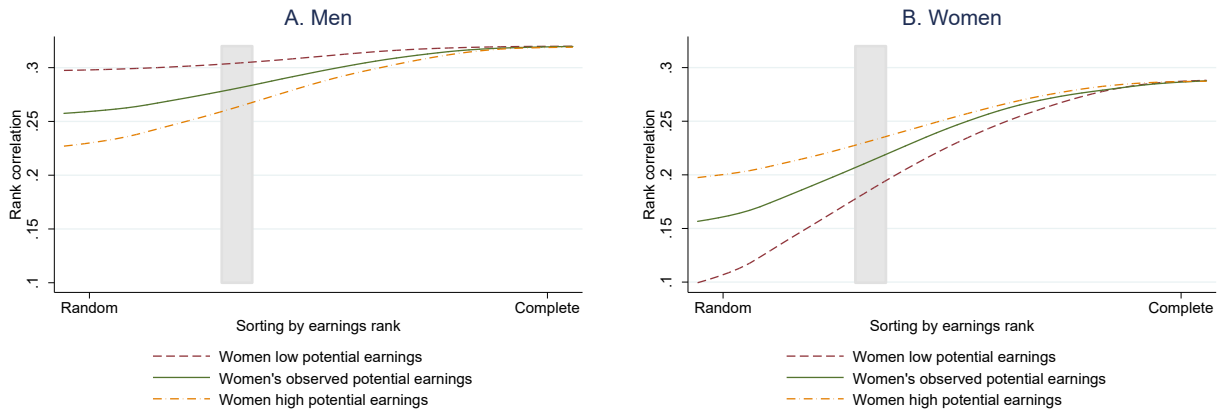
<sup>27</sup> Earlier work on marital sorting and intergenerational mobility has estimated the contribution of sorting by the covariance of spouse's and parents' earnings as a fraction of the covariance between household earnings and parents' earnings (Ermisch, Francesconi, and Siedler 2006). In this context, it is equivalent to  $Cov(y_t^p, y_{t-1})/Cov(z_t, y_{t-1})$ . Given the rank transformation, this fraction goes from 0 to 1 when moving from random to perfect sorting in the offspring generation. The simulations in this paper however make it clear that large variations in the contribution of spouse's earnings as measured by  $Cov(y_t^p, y_{t-1})/Cov(z_t, y_{t-1})$  do not necessarily translate into large variations in the intergenerational rank correlation, since at all levels of sorting own earnings ( $y_t$ ) is a relatively good predictor of family earnings ( $z_t$ ) for both men and women at observed earnings distributions.

Figure 3 Simulation of lambda under different assumptions on marital sorting and women's earnings distribution



Note: Figures show the coefficients from a linear projection of the family earnings rank on individual earnings rank ( $\lambda$ , see equation 5) under different assumptions about marital sorting and female labour supply. Marital sorting is simulated using the algorithm described in Appendix D. “Low potential earnings” assumes that women’s earnings are 50 percent of observed earnings. “High potential earnings” imputes women’s earnings at each percentile of women’s earnings distribution with the potential earnings of men at the corresponding percentiles in men’s earnings distribution. The grey shaded areas indicate the range of  $\lambda$ s empirically observed in the 1945–1955 cohorts.

Figure 4 Simulation of intergenerational rank correlations under different assumptions on marital sorting and women's earnings distribution



Note: Figures show the intergenerational rank correlation using pooled household earnings in the offspring generation ( $\beta^*$ , assuming  $Cov(y_{t-1}, \epsilon_t) = 0$ ), under different assumptions about marital sorting and female labour supply. Marital sorting is simulated using the algorithm described in Appendix D. The grey shaded areas indicate the range of simulated intergenerational correlations corresponding to the  $\lambda$ s empirically observed in the 1945–1955 cohorts. See notes to Figure 3 for details on women’s earnings distributions.

A relevant concern with the results presented above is that ranks of pooled household earnings suppress differences in income inequality in the offspring generation as generated by different degrees of sorting. As a sensitivity test, Appendix Figures A3 and A4 therefore present the corresponding intergenerational elasticity under different assumptions about marital sorting. Using a potential earnings distribution that preserves the mean and variance of the actual earnings distribution is here a key contribution.

As shown in Figure A3, the elasticity exhibits a steeper gradient of  $\lambda$  for both women and men, indicating that the variation in household income inequality generated by sorting is attenuated when using ranks, as compared to log potential earnings.<sup>28</sup>

Figure A4 presents the simulated intergenerational elasticities and shows (as expected) that the elasticity is more sensitive to marital sorting than is the rank correlation.<sup>29</sup> Moving from random to complete sorting, men's elasticity increases from 0.17 to 0.32, and women's from 0.10 to 0.27 when using women's observed earnings distribution. Empirically observed variations in sorting, indicated by the shaded bars, however have only minor implications for the intergenerational elasticity for men, but somewhat larger effects for women.

On a final note, the analysis ignores the possibility that sorting between spouse and in-laws plays an important role in passing on socioeconomic status to the offspring generation. This may be particularly true when women's earnings are low, and/or when social position is determined by family background rather than individual characteristics.

## 7 Conclusions

This paper clarifies the link between marital sorting and intergenerational mobility, when allowing offspring's earnings to be represented by pooled household earnings. Quantifying the contribution of marital sorting to intergenerational persistence shows that variations in marital sorting must be large in order to affect intergenerational mobility to a great extent. Instead, simulations show that the relative earnings distributions of men and women, in combination with sorting, are important for our understanding of intergenerational persistence.

The findings of this paper shed light on the processes underlying intergenerational mobility and are informative for understanding and comparing mobility patterns across regions or countries, and over time. In particular, part-time work among women differs substantially across countries, with part-time rates among OECD countries varying from well below 10 percent in eastern Europe, to above 50 percent in the Netherlands (OECD 2018). These cross-country differences in women's labour supply are likely to affect estimates of intergenerational persistence across countries. This is of particular importance if, as in the most recent literature, intergenerational mobility is estimated using household earnings.

---

<sup>28</sup> Unlike the case for ranked earnings,  $\lambda$  is not a priori bounded to one, but when men's and women's earnings distributions are the same  $\lambda$  converges to 1 at complete sorting.

<sup>29</sup> Figure A4b presents  $\beta^* = \lambda\beta$  for women. At high levels of sorting the ordering of  $\beta^*$  for different earnings distributions is different from the ordering of  $\lambda$  (for women). This is because the individual elasticity  $\beta$  differs for the three different earnings distributions of women:  $\beta$  is higher for "women high earnings potential" than for the other two distributions.

## References

- Albrecht, James, Aico van Vuuren, and Susan Vroman. 2015. 'The Black–White Wage Gap among Young Women in 1990 vs. 2011: The Role of Selection and Educational Attainment'. *Labour Economics* 33 (April): 66–71. <https://doi.org/10.1016/j.labeco.2015.02.006>.
- Becker, Gary S. 1993. *A Treatise on the Family: Enlarged Edition*. Enlarged edition. Cambridge, Mass.: Harvard University Press.
- Beller, Emily. 2009. 'Bringing Intergenerational Social Mobility Research into the Twenty-First Century: Why Mothers Matter'. *American Sociological Review* 74 (4): 507–28.
- Björklund, Anders. 2006. 'Does Family Policy Affect Fertility?' *Journal of Population Economics* 19 (1): 3–24. <https://doi.org/10.1007/s00148-005-0024-0>.
- Björklund, Anders, and Markus Jäntti. 1997. 'Intergenerational Income Mobility in Sweden Compared to the United States'. *The American Economic Review* 87 (5): 1009–18.
- Black, Sandra E., and Paul J. Devereux. 2011. 'Chapter 16 - Recent Developments in Intergenerational Mobility\*'. In *Handbook of Labor Economics*, edited by David Card and Orley Ashenfelter, 4, Part B:1487–1541. Elsevier. [https://doi.org/10.1016/S0169-7218\(11\)02414-2](https://doi.org/10.1016/S0169-7218(11)02414-2).
- Böhlmark, Anders, and Matthew J. Lindquist. 2006. 'Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden'. *Journal of Labor Economics* 24 (4): 879–96. <https://doi.org/10.1086/506489>.
- Boschini, Anne, Chirstina Håkanson, Åsa Rosén, and Anna Sjögren. 2011. 'Trading It off or Having It All? Completed Fertility and Mid-Career Earnings of Swedish Men and Women'. IFAU Working paper 2011:15.
- Bratsberg, Bernt, Simen Markussen, Oddbjørn Raaum, Knut Røed, and Ole Røgeberg. 2018. 'Trends in Assortative Mating and Offspring Outcomes', 49.
- Chadwick, Laura, and Gary Solon. 2002. 'Intergenerational Income Mobility Among Daughters'. *American Economic Review* 92 (1): 335–44. <https://doi.org/10.1257/000282802760015766>.
- Chetty, Raj, and Nathaniel Hendren. 2018. 'The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects'. *The Quarterly Journal of Economics* 133 (3): 1107–62. <https://doi.org/10.1093/qje/qjy007>.

- Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. 'Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States'. *The Quarterly Journal of Economics* 129 (4): 1553–1623. <https://doi.org/10.1093/qje/qju022>.
- Eika, Lasse, Magne Mogstad, and Basit Zafar. n.d. 'Educational Assortative Mating and Household Income Inequality'. *Journal of Political Economy* forthcoming.
- Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux: Study of Class Mobility in Industrial Societies*. Oxford England : New York: Clarendon Press.
- Ermisch, John, Marco Francesconi, and Thomas Siedler. 2006. 'Intergenerational Mobility and Marital Sorting\*'. *The Economic Journal* 116 (513): 659–79. <https://doi.org/10.1111/j.1468-0297.2006.01105.x>.
- Fernández, Raquel, Nezih Guner, and John Knowles. 2005. 'Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality'. *The Quarterly Journal of Economics* 120 (1): 273–344. <https://doi.org/10.1162/0033553053327498>.
- Fernández, Raquel, and Richard Rogerson. 2001. 'Sorting and Long-Run Inequality'. *The Quarterly Journal of Economics* 116 (4): 1305–41. <https://doi.org/10.1162/003355301753265589>.
- Gihleb, Rania, and Kevin Lang. 2016. 'Educational Homogamy and Assortative Mating Have Not Increased'. Working Paper 22927. National Bureau of Economic Research. <https://doi.org/10.3386/w22927>.
- Gonalons-Pons, Pilar, and Christine R. Schwartz. 2017. 'Trends in Economic Homogamy: Changes in Assortative Mating or the Division of Labor in Marriage?' *Demography* 54 (3): 985–1005. <https://doi.org/10.1007/s13524-017-0576-0>.
- Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos. 2014. 'Marry Your Like: Assortative Mating and Income Inequality'. *American Economic Review* 104 (5): 348–53. <https://doi.org/10.1257/aer.104.5.348>.
- Haider, Steven, and Gary Solon. 2006. 'Life-Cycle Variation in the Association between Current and Lifetime Earnings'. *American Economic Review* 96 (4): 1308–20. <https://doi.org/10.1257/aer.96.4.1308>.
- Hugh-Jones, David, Karin J. H. Verweij, Beate St. Pourcain, and Abdel Abdellaoui. 2016. 'Assortative Mating on Educational Attainment Leads to Genetic Spousal Resemblance for Polygenic Scores'. *Intelligence* 59 (November): 103–8. <https://doi.org/10.1016/j.intell.2016.08.005>.

- Kalmijn, Matthijs. 1994. 'Assortative Mating by Cultural and Economic Occupational Status'. *American Journal of Sociology* 100 (2): 422–52.
- Lee, Chul-In, and Gary Solon. 2009. 'Trends in Intergenerational Income Mobility'. *The Review of Economics and Statistics* 91 (4): 766–72. <https://doi.org/10.1162/rest.91.4.766>.
- Mare, Robert D. 1991. 'Five Decades of Educational Assortative Mating'. *American Sociological Review* 56 (1): 15–32. <https://doi.org/10.2307/2095670>.
- . 2011. 'A Multigenerational View of Inequality'. *Demography* 48 (1): 1–23.
- Mascie-Taylor, C. G. N. 1989. 'Spouse Similarity for IQ and Personality and Convergence'. *Behavior Genetics* 19 (2): 223–27. <https://doi.org/10.1007/BF01065906>.
- Mazumder, Bhashkar. 2005. 'Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data'. *The Review of Economics and Statistics* 87 (2): 235–55. <https://doi.org/10.1162/0034653053970249>.
- Meghir, Costas, and Mårten Palme. 2005. 'Educational Reform, Ability, and Family Background'. *American Economic Review* 95 (1): 414–24. <https://doi.org/10.1257/0002828053828671>.
- Neal, Derek. 2004. 'The Measured Black-White Wage Gap among Women Is Too Small'. *Journal of Political Economy* 112 (S1): S1–28. <https://doi.org/10.1086/379940>.
- Nyblom, Martin, and Jan Stuhler. 2016. 'Biases in Standard Measures of Intergenerational Income Dependence'. *Journal of Human Resources*, December, 0715-7290R. <https://doi.org/10.3368/jhr.52.3.0715-7290R>.
- OECD. 2018. 'Part-Time Employment Rate (Indicator). Doi: 10.1787/F2ad596c-En (Accessed on 09 July 2018)'. TheOECD. 2018. <http://data.oecd.org/emp/part-time-employment-rate.htm>.
- Raaum, Oddbjørn, Bernt Bratsberg, Knut Røed, Eva Österbacka, Tor Eriksson, Markus Jäntti, and Robin A Naylor. 2008. 'Marital Sorting, Household Labor Supply, and Intergenerational Earnings Mobility across Countries'. *The B.E. Journal of Economic Analysis & Policy* 7 (2). <https://doi.org/10.2202/1935-1682.1767>.
- Solon, Gary. 1992. 'Intergenerational Income Mobility in the United States'. *The American Economic Review* 82 (3): 393–408.

## Appendix

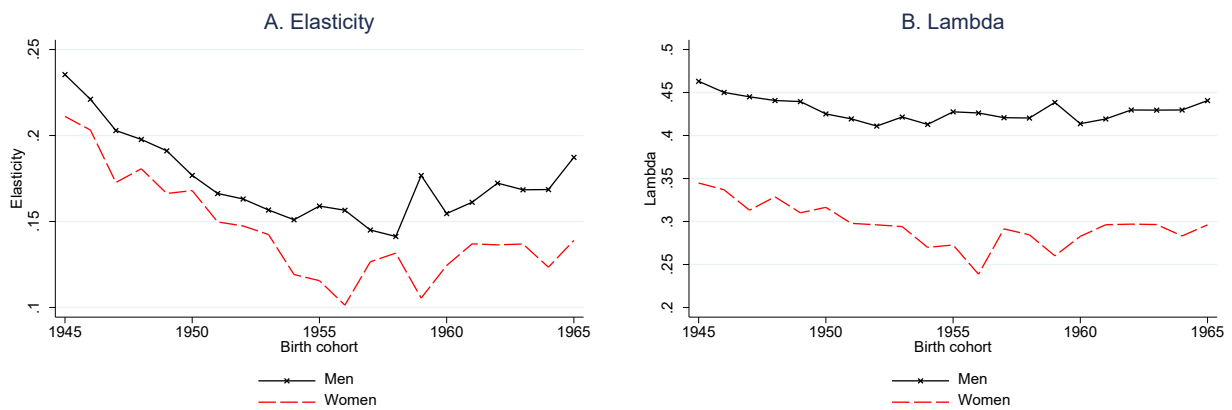
### A. Appendix tables and figures

Table A1 Descriptive statistics of estimation sample born 1945–1965

	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	St. dev.	N	Mean	St. dev.	N
		Men			Women	
Birth year	1955	6.19	681,928	1955	6.18	673,038
Father birth year	1924	9.15	681,928	1924	9.15	673,038
Earnings	305,192	217,774	681,033	198,866	115,298	672,453
Earnings spouse	204,813	121,001	681,072	294,105	208,236	672,182
Earnings father	236,405	175,257	681,928	232,883	181,164	673,038
Potential earnings (1955)	260,009	163,773	681,928	179,924	92,430	673,038
Potential earnings spouse (1955)	179,564	92,468	681,928	260,751	164,399	673,038
Pooled potential earnings (1955)	219,787	107,406	681,928	220,337	107,916	673,038
Log earnings	12.45	0.73	674,254	12.03	0.79	662,262
Log potential earnings (1955)	12.23	0.90	678,777	11.87	1.01	669,401
Log earnings spouse	12.05	0.82	669,653	12.42	0.72	665,680
Log potential earnings spouse (1955)	11.87	1.02	678,332	12.24	0.91	669,319
Log pooled potential earnings (1955)	12.18	0.52	681,911	12.19	0.53	673,007
Log earnings father	12.22	0.59	677,925	12.20	0.59	669,054

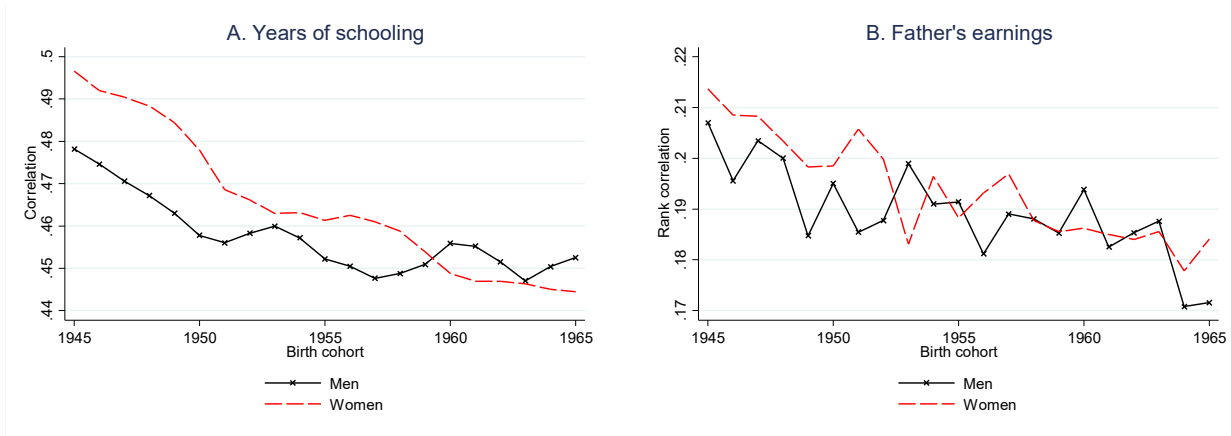
*Note:* Earnings and potential earnings are expressed in Swedish krona, 2006 values. Individuals with missing individual earnings observations are in a few cases assigned potential earnings, since the variables used for the prediction are observed. Potential earnings are imputed using the 1955 gender-specific earnings distributions for all cohorts.

Figure A1 Trends in assortative mating, log potential earnings, cohorts born 1945–1965



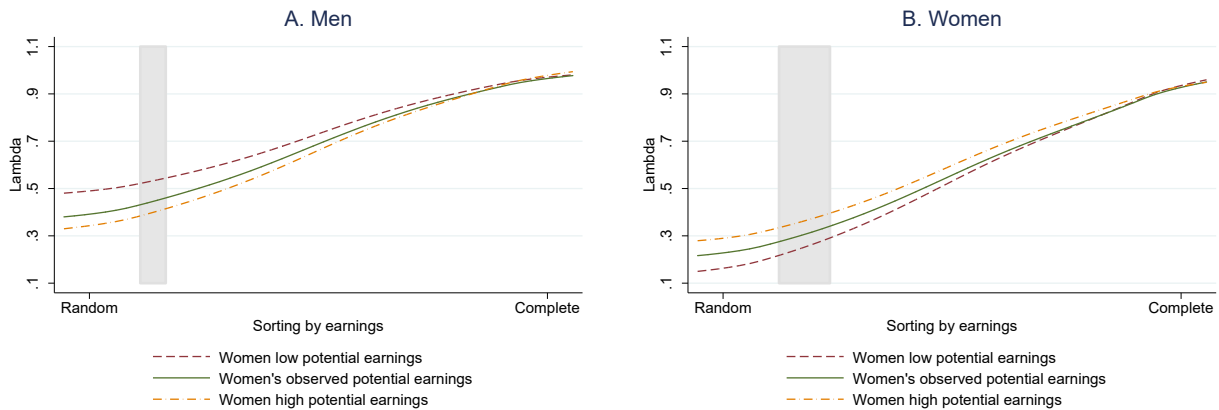
*Note:* The figure is based on cohorts born in Sweden. In panel A, assortative mating is measured by the elasticity in potential earnings. Panel B shows regression coefficients of linear projections of log household potential earnings on log individual's potential earnings ( $\lambda$ ).

Figure A2 Trends in assortative mating – alternative measures



Note: Figure A2A shows the Pearson correlation in spouse's years of schooling. Figure A2B shows the rank correlation in earnings of the father and the father-in-law.

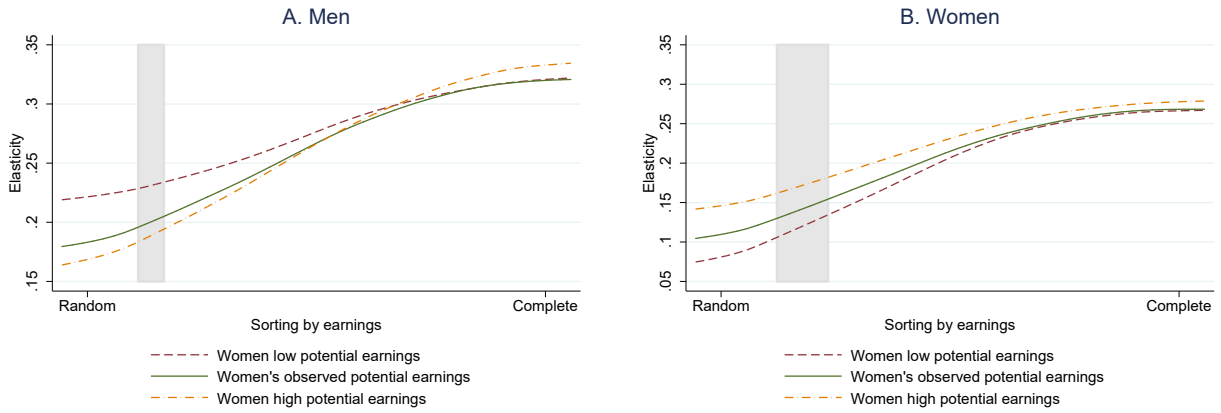
Figure A3 Simulation of lambda (based on log earnings) under different assumptions on marital sorting and women's earnings distribution



Note: Figures show the coefficients from a linear projection of log family earnings rank on log individual earnings ( $\lambda$ , see equation 5) under different assumptions about marital sorting and women's earnings. Marital sorting is simulated using the algorithm described in the Appendix D. The grey shaded areas indicate the range the of  $\lambda$ s empirically observed in the 1945–1955 cohorts. See notes to Figure 3 for details on women's earnings.



Figure A4 Simulation of intergenerational elasticities under different assumptions on marital sorting and women's earnings distribution



*Note:* Figures show the intergenerational elasticity using pooled household earnings in the offspring generation ( $\beta^*$ , assuming  $Cov(y_{t-1}, \varepsilon_t) = 0$ ), under different assumptions about marital sorting and women's earnings. Marital sorting is simulated using the algorithm described in Appendix D. The grey shaded areas indicate the range of simulated intergenerational elasticities corresponding to the  $\lambda$ s empirically observed in the 1945–1955 cohorts. See notes to Figure 3 for details on female labour supply.

## **B. Validation of ‘potential’ earnings**

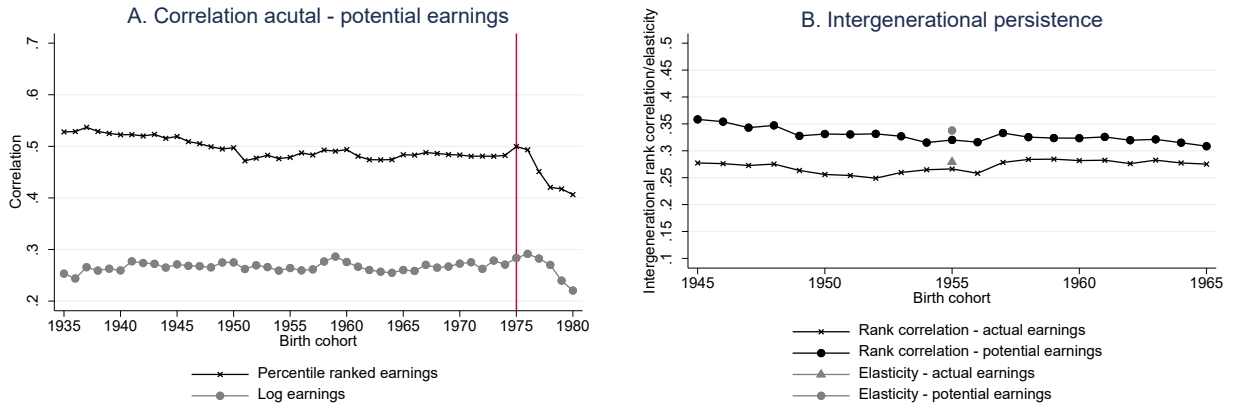
This paper consistently uses ‘potential’ earnings of the offspring generation. When studying marital sorting, it is necessary to use a measure of earnings potential rather than actual earnings, since actual earnings are endogenous to the realised match in the marriage market. One possible caveat with using ‘potential’ earnings is however that we measure a different intergenerational parameter than the one typically proposed in the literature: the variation in potential earnings is derived directly from schooling outcomes, and shuts down intergenerational links due to e.g. contacts and nepotism. We should thus expect to find a parameter of different size. Another concern is that the predictive power of fields and levels of education for earnings may vary over time, which can introduce spurious trends in measures of marital sorting and intergenerational mobility using potential earnings.

In order to validate the use of potential earnings, I make the assumption that men’s labour supply has been constant over time, and estimate (for men only) a) correlations between actual earnings and potential earnings over time, and b) intergenerational rank correlations and elasticities using actual and potential earnings.

Figure A4a shows correlations between percentile ranks of actual and predicted earnings, and between logs of actual and predicted earnings for men born 1935–1980. The correlations are largely stable over time, at around 0.5 for percentile ranks and 0.4 for logs. For cohorts born after 1975, there is a sharp drop in the correlation between actual and potential earnings. These are cohorts for which we can at the latest observe earnings in their early 30s, and as such their cohort-specific rank in potential earnings turns out to be less reliable. For this reason, the estimation sample is restricted to index individuals born 1945–1965, allowing partners to be born  $\pm 10$  years from the index individual.

Figure A4b compares annual intergenerational rank correlations when using sons’ actual and potential earnings. The rank correlations are higher when using potential earnings, which likely reflects that the prediction uses schooling, and intergenerational correlations in schooling tend to be higher than those of earnings or income (Black and Devereux 2011). Importantly, however, both series show stability over time and there is no obvious deviation in long-run trends. The similarity of trends in Figures A4a and A4b indicates that the relationship between actual and predicted earnings is stable over time in the full-time working population, and thus supports the use of potential earnings as a proxy for earnings. Figure A4b also shows intergenerational elasticities measured in the 1955 cohort. Since fathers’ earnings are observed at different ages over child cohorts, I refrain from studying trends in the elasticity.

Figure A5 Sensitivity analysis of 'potential earnings'



*Note:* Figure A4a shows correlations between actual and potential earnings for men only. Percentile ranks are taken by cohort within the estimation sample. Figure A4b shows measures of intergenerational persistence (fathers – sons) using different definitions of son’s earnings.

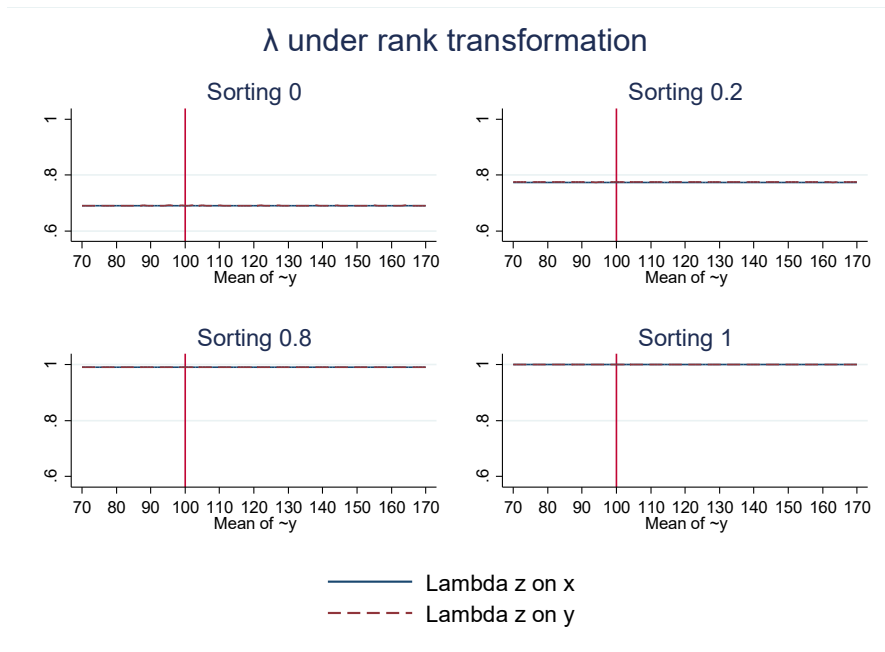
### C. How does $\lambda$ react to changes in the underlying earnings distributions?

This section explores the behaviour of  $\lambda$  as means and variances of the underlying earnings distributions change. We have two random normal distributions,  $\tilde{x}$  and  $\tilde{y}$  where the mean is large in relation to the variance such that  $\tilde{x}$  and  $\tilde{y}$  rarely take on negative values.  $E(\tilde{x})$  and  $Var(\tilde{x})$  are held constant throughout. We now vary either i) the mean of  $\tilde{y}$  (holding constant the variance); or ii) the variance of  $\tilde{y}$ , holding constant the mean.

Define  $\tilde{z} = 0.5(\tilde{x} + \tilde{y})$  and take ranks or logs of  $\tilde{x}$ ,  $\tilde{y}$  and  $\tilde{z}$  to get the distributions  $x$ ,  $y$  and  $z$ . Figures A6–A9 show  $\lambda$  from estimates of  $z$  on  $x$  and  $z$  on  $y$ , as the mean or variance of  $\tilde{y}$  changes, at different levels of sorting between  $\tilde{y}$  and  $\tilde{x}$ , for the rank and log transformations, respectively. Figures A6 and A7 present results for the rank transformation, and show that  $\lambda$  is insensitive to changes in means, but varies with changes in the underlying variance of  $\tilde{y}$ . As the variance of  $\tilde{y}$  increases, the predictive power of  $y$  for  $z$  increases, while the opposite is true for  $x$ . At complete sorting,  $\lambda = 1$  for all distributions of  $\tilde{y}$ . A simultaneous increase in both the mean and the variance of  $\tilde{y}$  thus has unambiguous predictions for how  $\lambda$  should react, since  $\lambda$  is insensitive to changes in the mean.

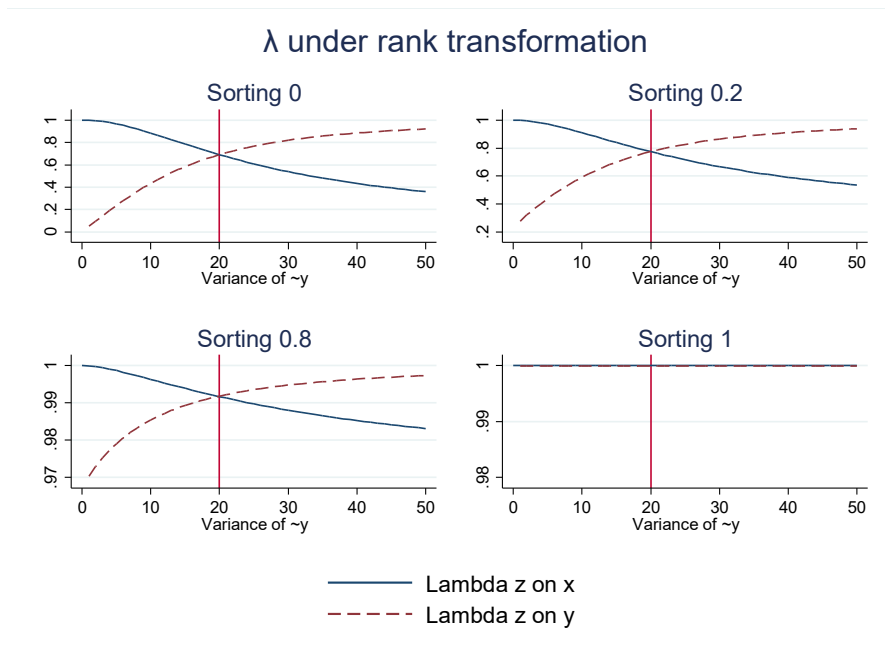
Figures A8 and A9 examine the properties of  $\lambda$  under the log transformation. First, the predictive power of  $y$  for  $z$  increases as the mean of  $\tilde{y}$  goes up, and simultaneously  $x$  loses explanatory power. In contrast to the rank transformation, changes in means of the underlying distributions are important for the relationship between  $z$  and  $y$ , and  $z$  and  $x$ . Next, in Figure A9, we see that  $\lambda$  is also sensitive to changes in the variance of  $\tilde{y}$ . As the variance increases,  $\lambda$  from  $z$  on  $y$  declines while  $\lambda$  from  $z$  on  $x$  increases. Unlike the case of the rank correlation, there is no a priori prediction of how  $\lambda$  will react to a simultaneous increase in both the mean and variance of the underlying distribution  $\tilde{y}$ , since the reactions to means and variances go in opposite directions.

Figure A6 Lambda under rank transformation – reaction to change in mean



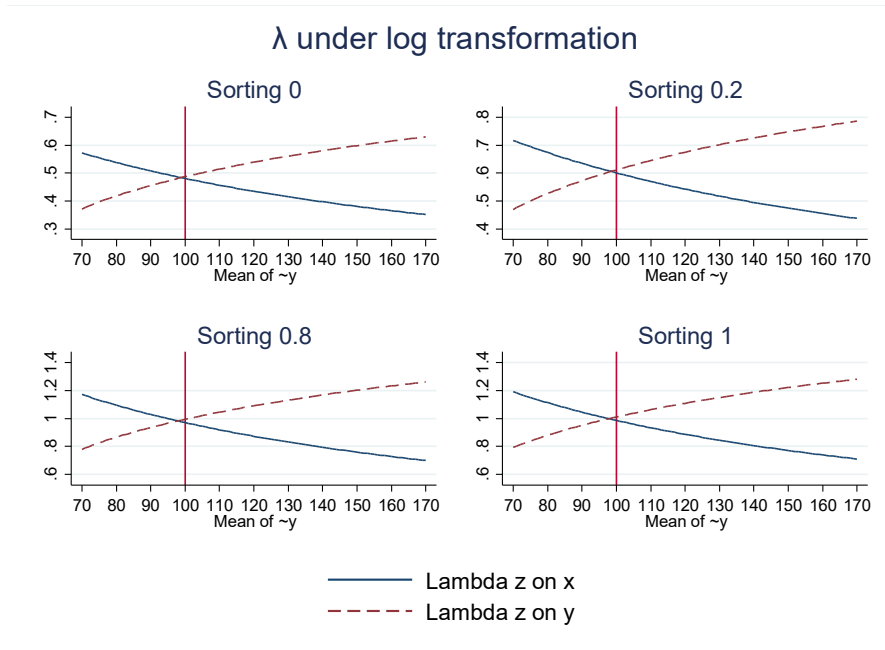
*Note:* Simulations of lambda under different means of  $\tilde{y}$ . The vertical line shows the mean of the  $\tilde{x}$ -distribution, set to 100. The variance is held constant in both  $\tilde{y}$  and  $\tilde{x}$ -distributions and set to 20. Sorting refers to the amount of sorting between x and y, using the algorithm described in Appendix D.

Figure A7 Lambda under rank transformation – reaction to change in variance



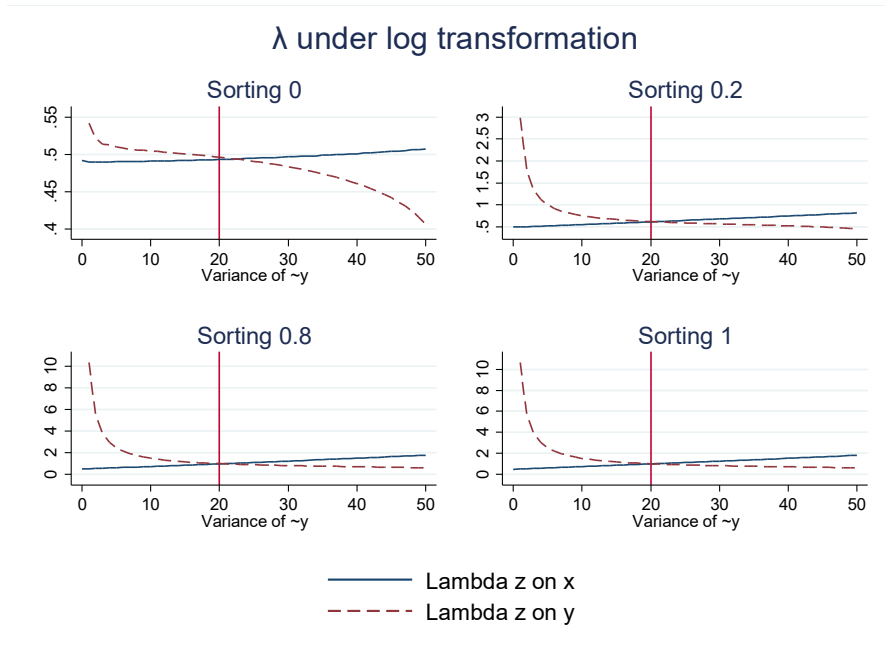
*Note:* Simulations of lambda under different variances of  $\tilde{y}$ . The vertical line shows the variance of the  $\tilde{x}$ -distribution, set to 20. The mean is held constant in both  $\tilde{y}$  and  $\tilde{x}$ -distributions and set to 150. Sorting refers to the amount of sorting between x and y, using the algorithm described in Appendix D.

Figure A8 Lambda under log transformation – reaction to change in mean



*Note:* Simulations of lambda under different means of  $\tilde{y}$ . The vertical line shows the mean of the  $\tilde{x}$ -distribution, set to 100. The variance is held constant in both  $\tilde{y}$  and  $\tilde{x}$ -distributions and set to 20. Sorting refers to the amount of sorting between x and y, using the algorithm described in Appendix D.

Figure A9 Lambda under log transformation – reaction to change in variance



*Note:* Simulations of lambda under different variances of  $\tilde{y}$ . The vertical line shows the variance of the  $\tilde{x}$ -distribution, set to 20. The mean is held constant in both  $\tilde{y}$  and  $\tilde{x}$ -distributions and set to 150. Sorting refers to the amount of sorting between x and y, using the algorithm described in Appendix D.

## D. Sorting algorithm

Hypothetical matching of spouses under different degrees of marital sorting is done using a re-weighting procedure. For  $\gamma \in [0,1]$ , I create  $\hat{y}_t^p = \gamma \tilde{y}_t^{p-s} + (1 - \gamma) \tilde{y}_t^{p-r}$ , where  $\tilde{y}_t^{p-s}$  stands for partners' potential earnings sorted from the lowest to the highest value, and  $\tilde{y}_t^{p-r}$  represents random sorting of partners' potential earnings. Since the re-weighting procedure compresses the variance, I next take percentiles of  $\hat{y}_t^p$  and to each percentile assign the original values of the distribution of  $\tilde{y}_t^p$  to maintain the moments. For each draw of  $\gamma$  (increased in each draw by an increment of 0.01),  $\hat{y}_t^p$  is paired with  $\tilde{y}_t$  and  $\tilde{y}_{t-1}$ , sorted from the lowest to the highest value of  $\tilde{y}_t$ . After this procedure, I compute pooled household earnings and take ranks or logs of individual and household earnings.