Online Appendix

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

Helena Holmlund
helena.holmlund@ifau.uu.se
Institute for the Evaluation of Labour Market and Education Policy (IFAU)

A. How does λ react to changes in the underlying earnings distributions?

The purpose of this section is to clarify how λ responds to changes in the means and variances of the underlying (non-transformed) earnings distributions of women and men, by presenting a simple simulation exercise using hypothetical distributions. The analysis demonstrates the behaviour of λ both under the rank and log transformation.

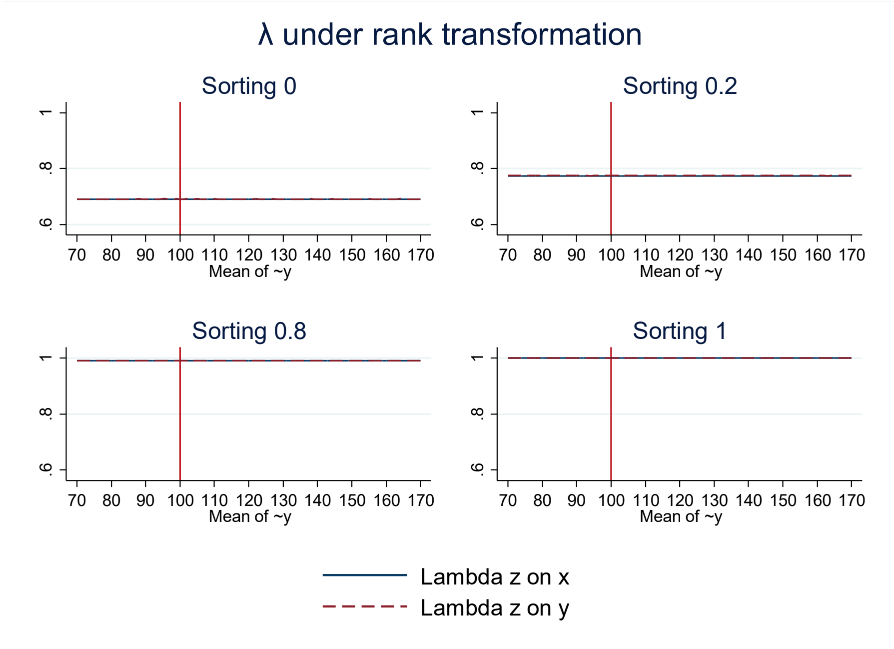
We have two random normal distributions, $\tilde{x}$ and $\tilde{y}$ where the mean is large in relation to the variance such that $\tilde{x}$ and $\tilde{y}$ rarely take on negative values. Next, we define $\tilde{z} = 0.5(\tilde{x}+\tilde{y})$. Throughout the simulation exercise, $E(\tilde{x})$ and $Var(\tilde{x})$ are held constant. Instead, we vary either i) the mean of $\tilde{y}$ (holding constant the variance); or ii) the variance of $\tilde{y}$, holding constant the mean.

The simulation first explores the rank transformation of $\tilde{x}$, $\tilde{y}$ and $\tilde{z}$, i.e. the rank-transformed distributions x, y and z. Figures A1–A2 show λ 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 $\tilde{x}$. The figures show that λ 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 λ should react, since λ is insensitive to changes in the mean.

Figures A3 and A4 examine the properties of λ under the log transformation (i.e. the transformed distributions x, y and z now refer to 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.

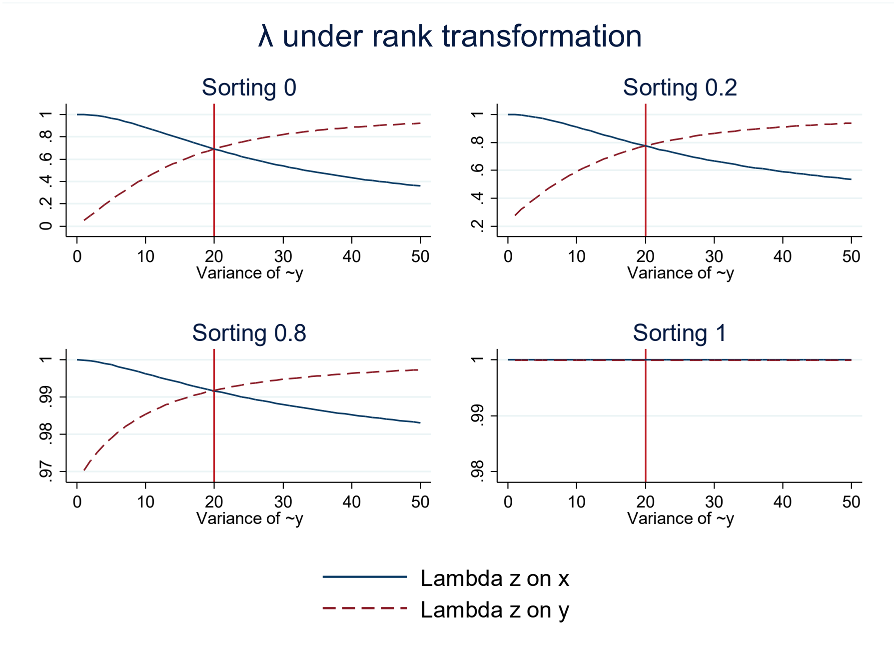
Next, in Figure A4, we see that λ is also sensitive to changes in the variance of $\tilde{y}$. As the variance increases, λ from z on y declines while λ from z on x increases. Unlike the case of the rank correlation, there is no a priori prediction of how λ will react to a simultaneous increase in both the mean and variance of the underlying distribution $\tilde{y}$, since the reactions to means and variances go in opposite directions.

Figure A 1 Lambda under rank transformation – reaction to change in mean



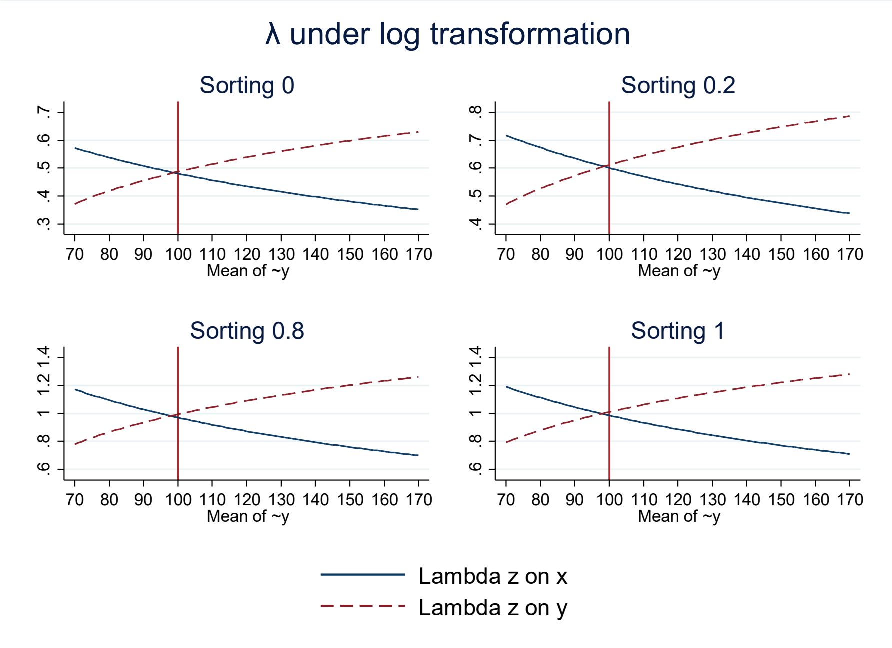
Notes: Simulations of lambda under different means of $\tilde{y}$. The vertical line shows the mean of the $\tilde{x}$-distribution, set to 100. The variance is held constant in both $\tilde{y}$ and $\tilde{x}$-distributions and set to 20. Sorting refers to the amount of sorting between x and y, using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

Figure A 2 Lambda under rank transformation – reaction to change in variance



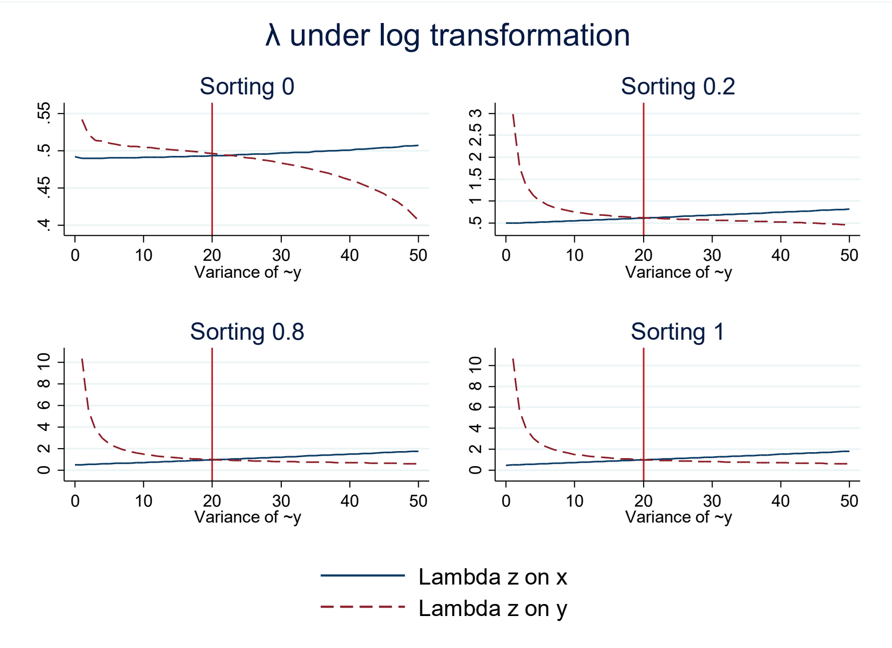
Notes: Simulations of lambda under different variances of $\tilde{y}$. The vertical line shows the variance of the $\tilde{x}$-distribution, set to 20. The mean is held constant in both $\tilde{y}$ and $\tilde{x}$-distributions and set to 150. Sorting refers to the amount of sorting between x and y,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

Figure A 3 Lambda under log transformation – reaction to change in mean



Notes: 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 A 4 Lambda under log transformation – reaction to change in variance



Notes: 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.

B. Appendix tables and figures

Table A1 Descriptive statistics of estimation sample born 1945–1965

	(1) Mean	(2) St. dev.	(3) N	(4) Mean	(5) St. dev.	(6) 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

Notes: 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.

Table A2 Assortative mating and intergenerational rank correlations for Swedish men and women born 1945 and 1955. Sensitivity analysis using fathers' ranks within own cohort.

	(1) AM rank corr. α	(2) λ	(3) Coef. resid - father's rank	(4) Intergen ind. rank corr. β	(5) Intergen pooled rank corr. $\beta*$
Outcome variable:	Spouse earnings rank	Family earnings rank	Residual from col (2)	Individual earnings rank	Family earnings rank
Men born 1945					
Individual earnings rank	0.430***	0.888***			
	(0.005)	(0.002)			
Father's earnings rank			0.047***	0.364***	0.370***
			(0.002)	(0.005)	(0.005)
Observations	33,852	33,852	33,852	33,852	33,852
R-squared	0.186	0.790	0.010	0.125	0.130
Men born 1955					
Individual earnings rank	0.328***	0.872***			
	(0.005)	(0.002)			
Father's earnings rank			0.060***	0.327***	0.346***
			(0.003)	(0.006)	(0.006)
Observations	31,029	31,029	31,029	31,029	31,029
R-squared	0.107	0.757	0.014	0.100	0.112
Women born 1945					
Individual earnings rank	0.445***	0.769***			
	(0.005)	(0.003)			
Father's earnings rank			0.105***	0.321***	0.352***
			(0.004)	(0.005)	(0.005)
Observations	32,083	32,083	32,083	32,083	32,083
R-squared	0.200	0.597	0.025	0.094	0.114
Women born 1955					
Individual earnings rank	0.347***	0.719***			
	(0.005)	(0.004)			
Father's earnings rank			0.099***	0.297***	0.312***
			(0.004)	(0.006)	(0.006)
Observations	30,883	30,883	30,883	30,883	30,883
R-squared	0.120	0.516	0.019	0.082	0.091

Notes: Rank correlations are calculated using ranks in potential earnings in the offspring generation, and ranks in actual earnings of fathers. Fathers' earnings are ranked within father's birth cohort. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Figure A 5 Fraction of each cohort in the estimation sample

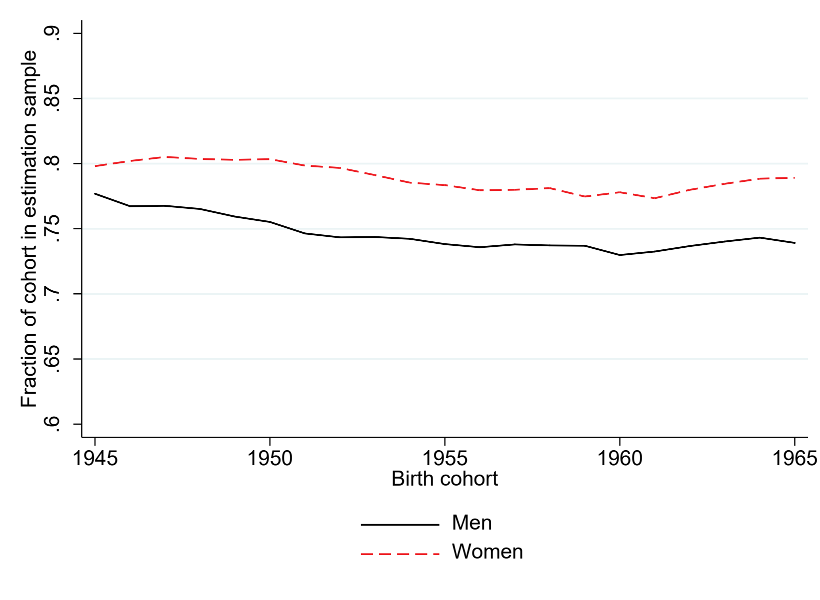


Figure A 6 Correlation between father's earnings rank and child being in estimation sample

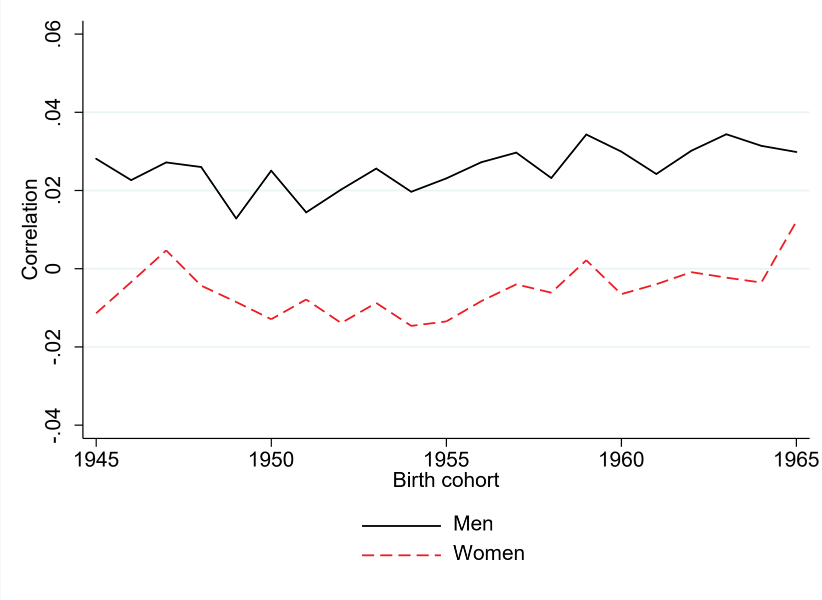
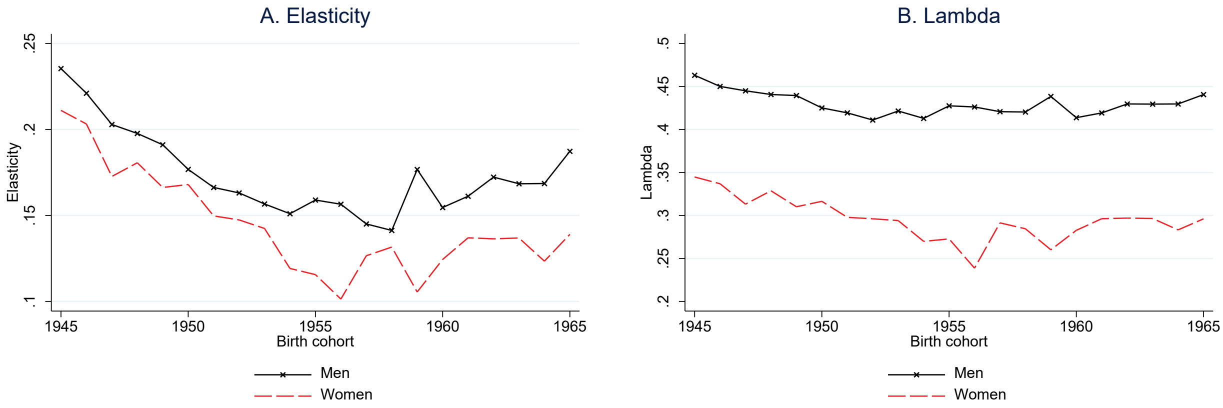
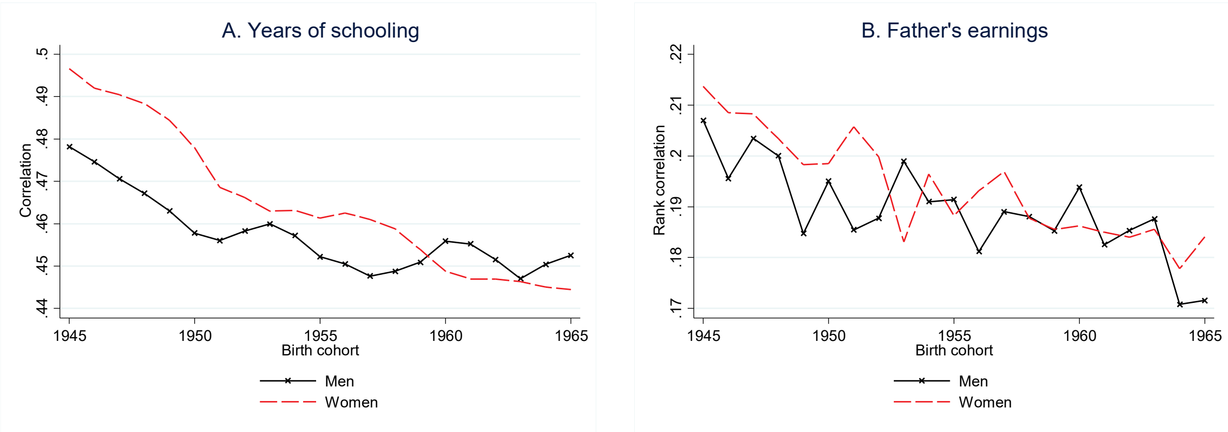


Figure A 7 Trends in assortative mating, log potential earnings, cohorts born 1945–1965



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 (λ).

Figure A 8 Trends in assortative mating – alternative measures



Note: Figure A8A shows the Pearson correlation in spouse's years of schooling. Figure A8B shows the rank correlation in earnings of the father and the father-in-law.

C. Validation of 'potential' earnings

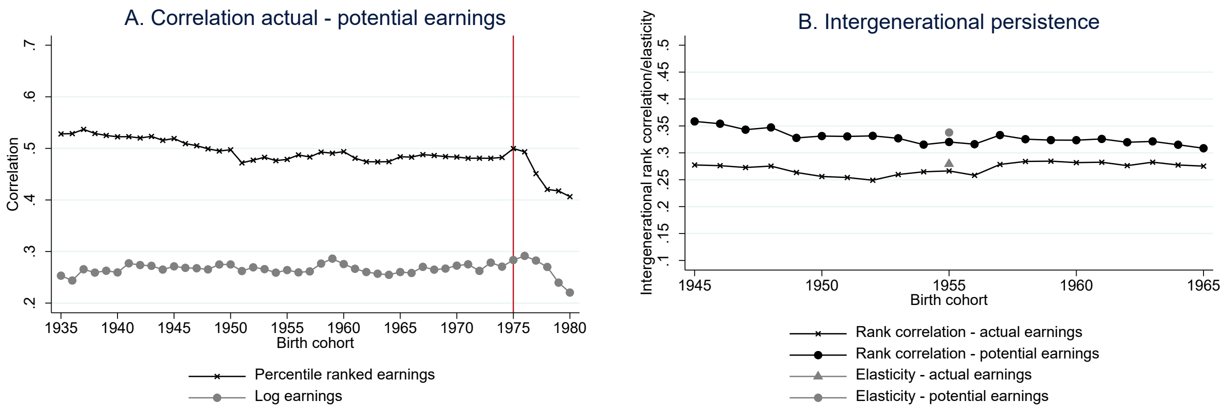
This paper consistently uses 'potential' earnings of the offspring generation. When studying marital sorting, it is necessary to use a measure of earnings potential rather than actual earnings, since actual earnings are endogenous to the realised match in the marriage market. One possible caveat with 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 A9a 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 0.25 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 +- 10 years from the index individual.

Figure A9b 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 A9a and A9b 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 A9b 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 A 9 Sensitivity analysis of 'potential earnings'



Notes: Figure A9a shows correlations between actual and potential earnings for men only. Percentile ranks are taken by cohort within the estimation sample. Figure A9b shows measures of intergenerational persistence (fathers – sons) using different definitions of son's earnings.

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 γ (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.