sociological science

Supplement to:

Shi, Jiaxin, and Martin Kolk. 2023. “The Inequality of Lifetime Pensions” Sociological Science 7: 667-693.

Appendix

The Inequality of Lifetime Pensions

Jiaxin Shi and Martin Kolk

Appendix 1.

A1.1 Forecasting the ages at death of individuals survived to 2019

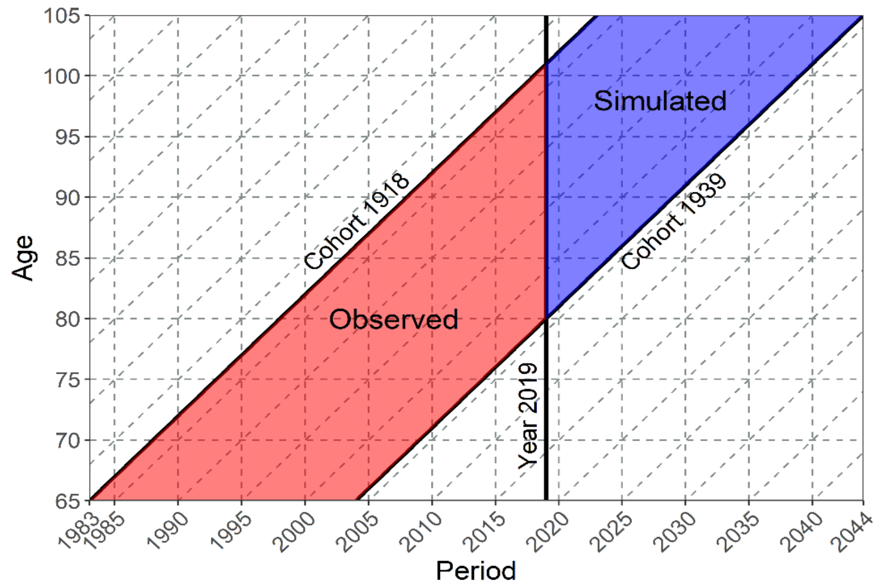

Figure A1. Lexis diagram for the data structure. *Source*: Authors' own.

We assume a Gompertz relationship for mortality rates at ages 65, 66, 67, …, 104, 105+. That is: $log\left(m(a)\right) = \alpha + \beta a$. We allowed both the intercept (α) and the slope (β) to vary across earnings and cohorts. Specifically, our imputation consisted of four steps summarized as follows:

Step 1. We fitted the linear models with the logarithm of mortality rates as the outcome variable, age (continuous), earnings quintile (ordinal, five levels), birth year (continuous), the interaction of age and earnings quintile, and the interaction of age and birth year as predictors. The models were estimated using the observed data and the ordinary least square method, separately for men and women.

Step 2. We predicted mortality rates for years of 2019 and onwards using the estimated coefficients from *Step 1*.

Step 3. We adjusted the estimated mortality rates from *Step 2* using mortality forecasts provided by Statistics Sweden (2020). This involved a proportional transformation for age-earnings-quintile-specific mortality rates so that the total mortality matched the official forecasts and the sizes of age-earnings quintiles were kept the same as empirically observed.

Step 4. We generated random numbers to simulate age at death for individuals who survived to 2019 using the adjusted mortality rates from *Step 3*.

Figure A1 illustrates the structure of the dataset used for subsequent analysis. The main results were highly robust when this procedure was repeated, or when education instead of earnings quintile was used as one of the predictors in *Step 1*.

A1.2 Standardization

We used a standardization approach to explore the roles of lifespan and pre-retirement earnings in lifetime pension inequality. By standardizing the age structure of one population to match that of a benchmark population, we can obtain the "standardized" income inequality value that can be seen as the counterfactual inequality level when the differences in age structures are "removed" (e.g., Lee 1994). We chose a set of benchmark years and then applied weights at the individual level to later-born cohorts so that the lifespan (or pre-retirement earnings) distributions of the later-born cohorts become identical to that of the benchmark cohort. The weights are the ratios between the density of the benchmark population and the later-born cohort at each age (income) position.

A1.3 The Partial R^2

The regression-based partial R^2 approach has also been used in the income and earnings inequality literature to disentangle the effects of different covariates on the variance of the outcome variable (e.g., Kim and Sakamoto 2008; Meng et al. 2013; Xie and Zhou 2014). The income variable is first regressed on a set of predictors. Then, a variable of interest is excluded from the regression model and the reduced model is re-estimated. The partial R^2 is calculated as:

$$Partial\ R^2 = \frac{R^2 - R^2_{-K}}{1 - R^2_{-K}}$$

where R^2 is the variance explained by all covariates in the full model and R^2_{-K} is variance explained by all covariates in the model where variable K is removed. This way, the partial R^2 can be interpreted as the proportion of the remaining variance that cannot be explained by other covariates but can be explained by variable K.

While this approach can show the relative importance of different variables in determining the total variance of the outcome variable, it has several limitations. First, it only shows the relative role of the determinants without accounting for the absolute level of inequality. Policy interventions are more concerned about the actual magnitude of inequality that is caused by certain sources. Second, partial R^2 may drift in either direction on some occasions when the covariate actually leads to an increase in the total inequality (see discussions in Zhou 2014). Third, variance as a measure of distributional dispersion is much less used in the income inequality literature, making results difficult to be compared across studies.

Appendix 2. Figures

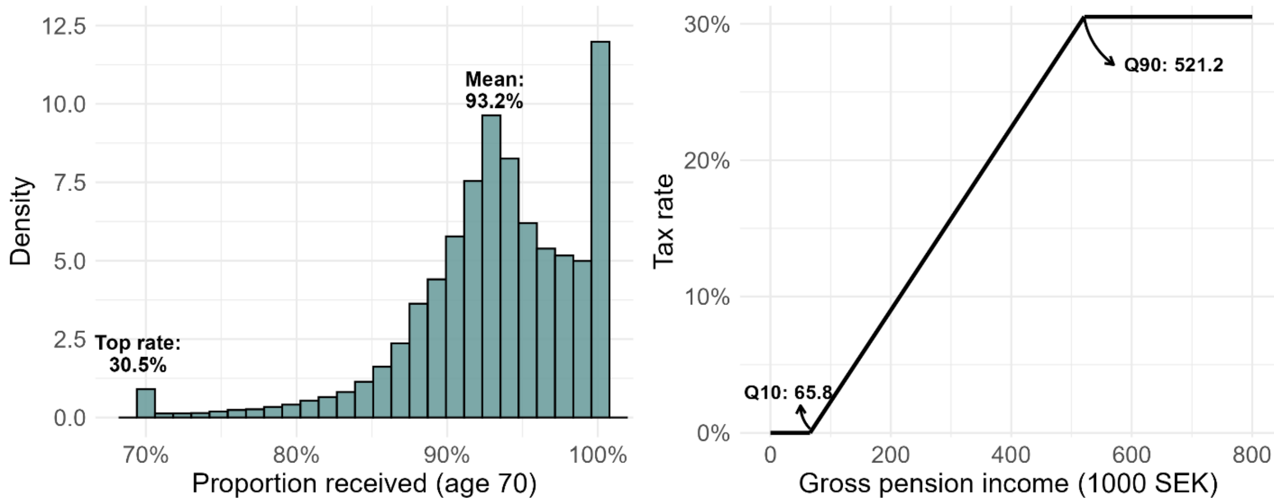

Figure A2. Illustrations for the progressive taxation scenario used for hypothetical pension calculations, 1928 cohort. Left: proportion of income received after tax. Right: tax rate by gross annual pension income. *Source*: Authors' calculations based on Swedish register data.

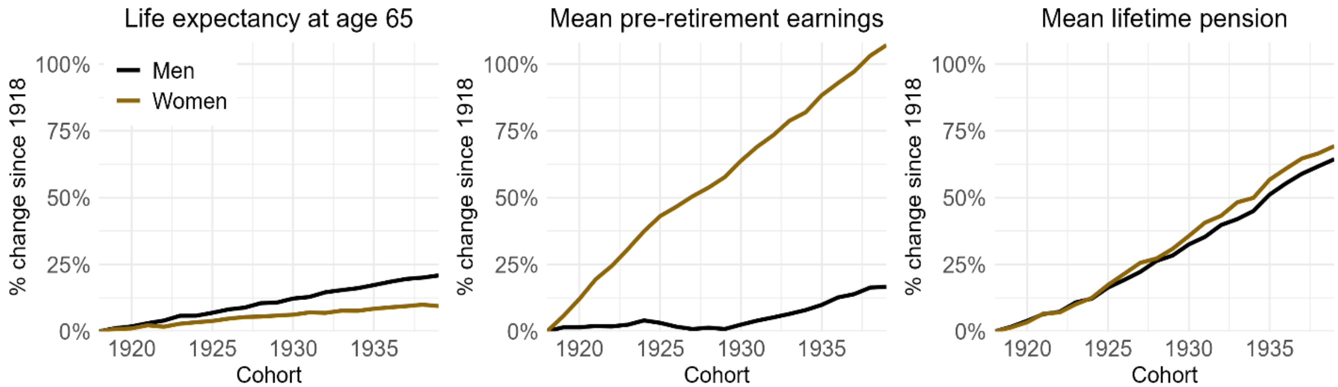

Figure A3. Proportional change in the means of three main variables as compared to 1918 by sex. Left: Life expectancy. Middle: Pre-retirement earnings (over ages 50–59). Right: Lifetime pension.
Source: Authors' calculations based on Swedish register data.

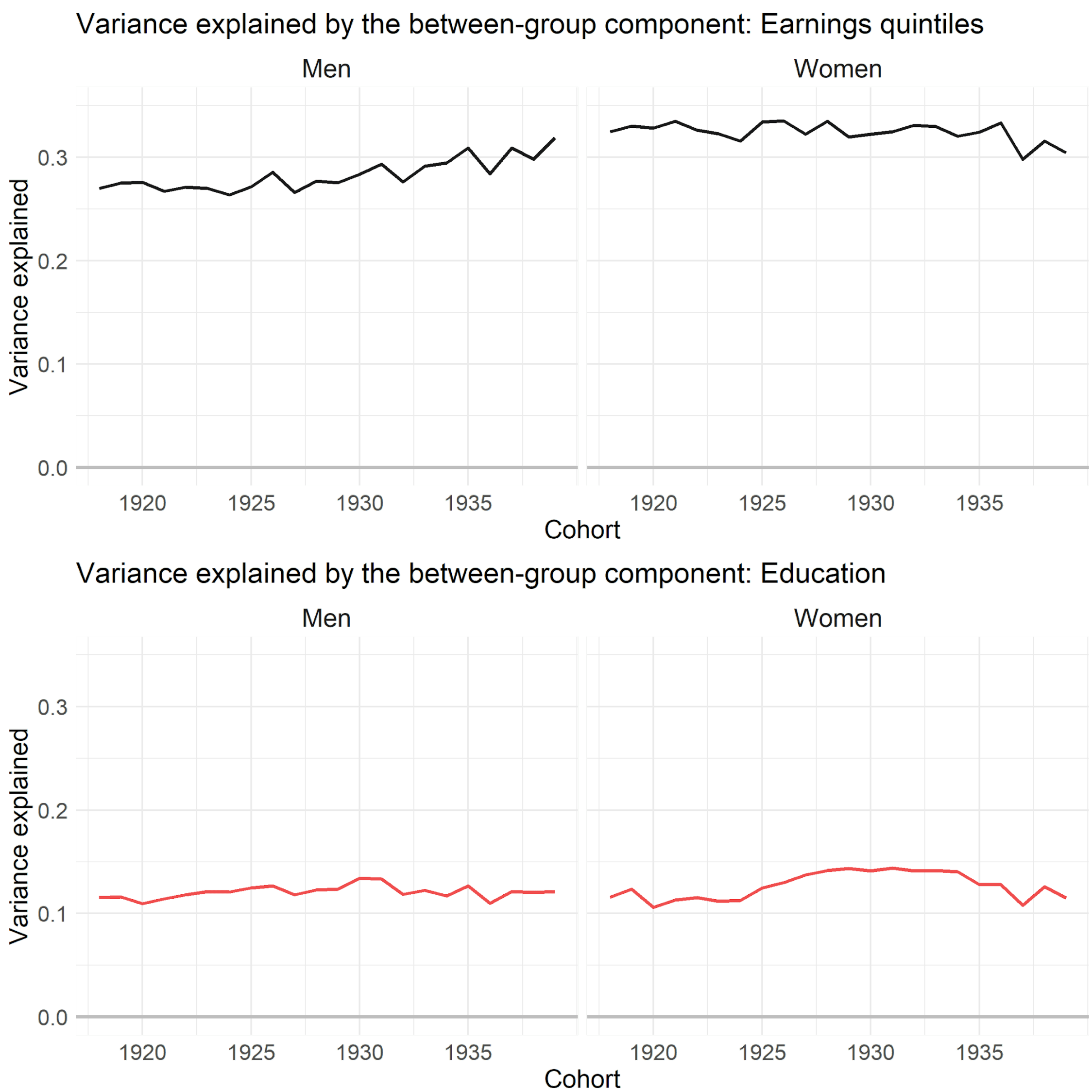

Figure A4. Total variance in lifetime pension explained by earnings quintiles (upper panels) and education (lower panels). *Source*: Authors' calculations based on Swedish register data. *Notes*: In the variance decomposition in the upper panels, we divide individuals into five equally-sized quintile groups based on average earnings between ages 50 and 59, separately by gender. In the analysis in the lower panels, we drop individuals with unknown educational levels, and have three levels of education in the decomposition: primary, secondary, and tertiary education.

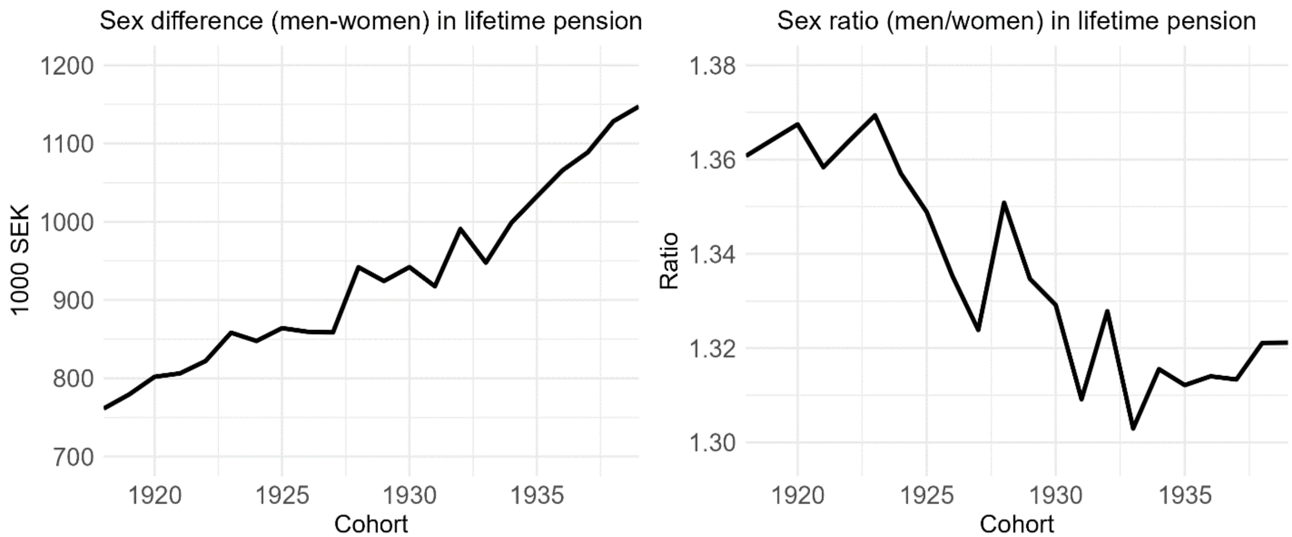

Figure A5. Cohort trends of sex differences in lifetime pension. Left: Sex difference (men – women). Right: Sex ratio (men/women). *Source*: Authors' calculations based on Swedish register data.

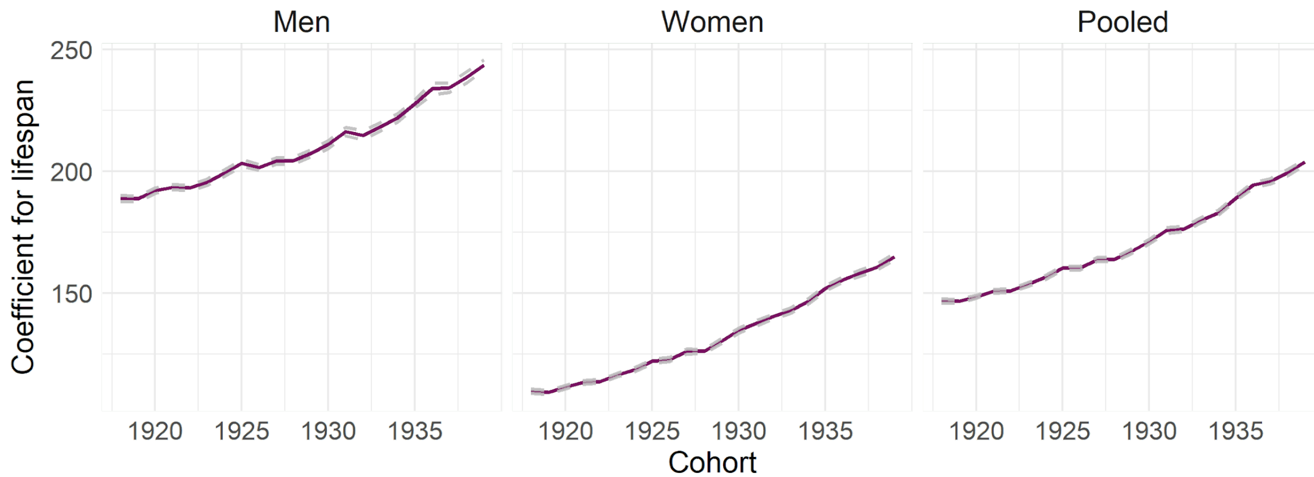

Figure A6. Coefficients for lifespan in the full models predicting lifetime pension. Left: Men. Middle: Women. Right: Pooled (men and women). *Source*: Authors' calculations based on Swedish register data. *Notes:* The coefficients are for lifespan for cohort-specific full models. Grey dashed lines denote 95% confidence intervals. Other predicting variables include earnings, education, civil status, occupation (EGP), and metropolitan county.

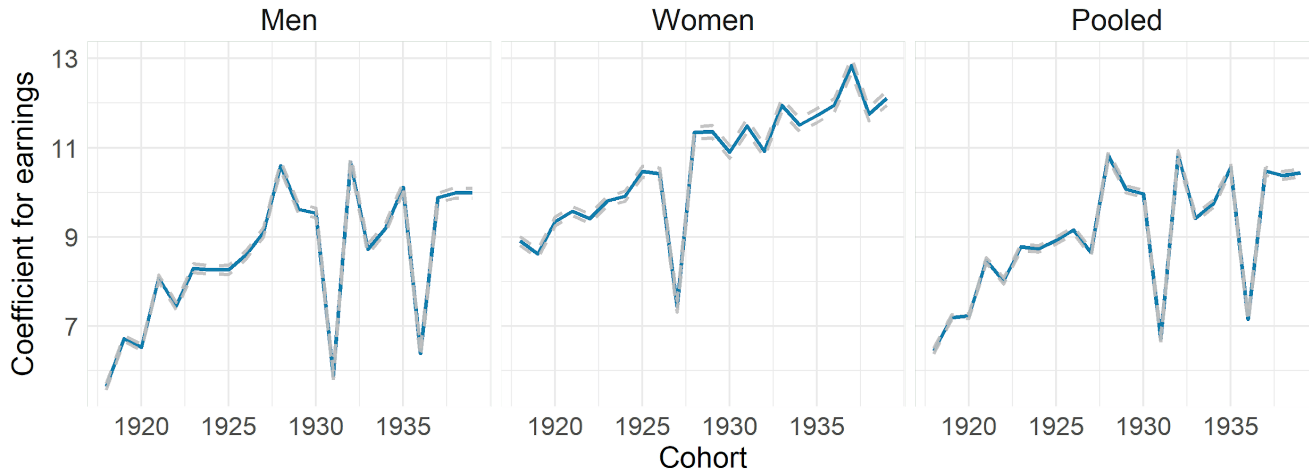

Figure A7. Coefficients for earnings in the full models predicting lifetime pension. Left: Men. Middle: Women. Right: Pooled (men and women). *Source*: Authors' calculations based on Swedish register data. *Notes:* The coefficients are for lifespan for cohort-specific full models. Grey dashed lines denote 95% confidence intervals. Other predicting variables include lifespan, education, civil status, occupation (EGP), and metropolitan county.

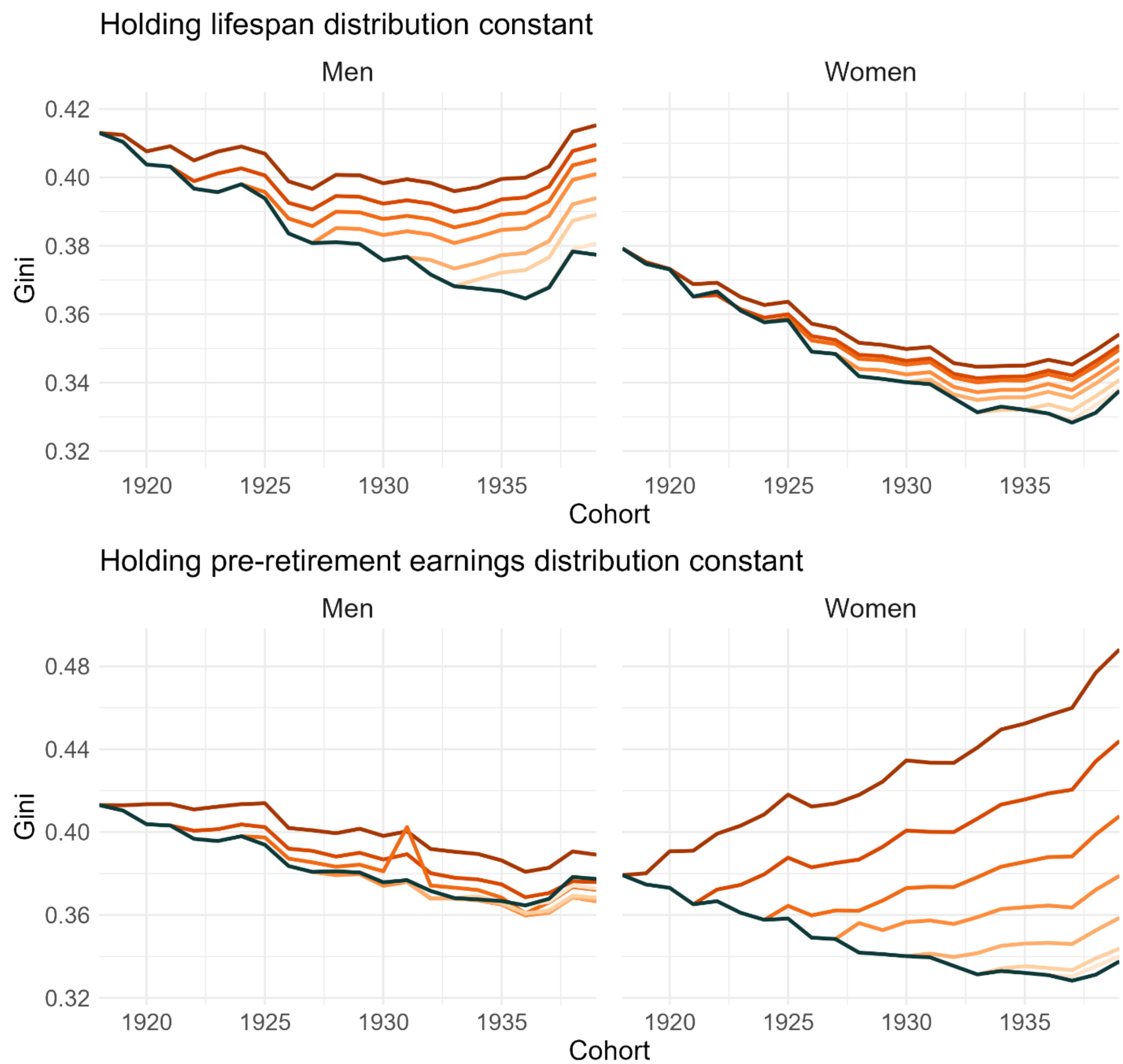

Figure A8. Observed and counterfactual Gini coefficients for lifetime pensions (1) holding lifespan distributions constant (upper panels) and (2) holding pre-retirement earnings distributions constant (lower panels). *Source*: Authors' calculations based on Swedish register data. *Notes*: The green line shows the observed Gini trend, and the orange lines show the counterfactual Gini trends. The benchmark years are 1918, 1921, 1924, 1927, 1930, 1933, and 1936. The lighter colors of the counterfactual trend lines denote more recent years.

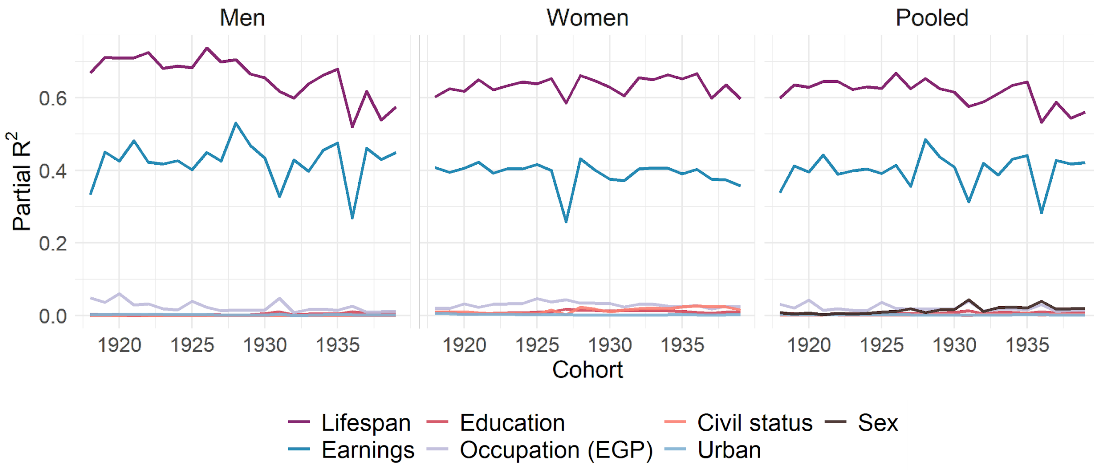

Figure A9. Partial R^2 for predicting variables. *Source*: Authors' calculations based on Swedish register data.

Figure A10. Relative importance of predicting variables and residuals. *Source*: Authors' calculations based on Swedish register data. *Notes:* The sum of the non-grey parts is equivalent to the R^2 of the regression models. The decomposition of R^2 uses the method proposed by Lindemann, Merenda, and Gold (1980): R^2 partitioned by averaging over orders.

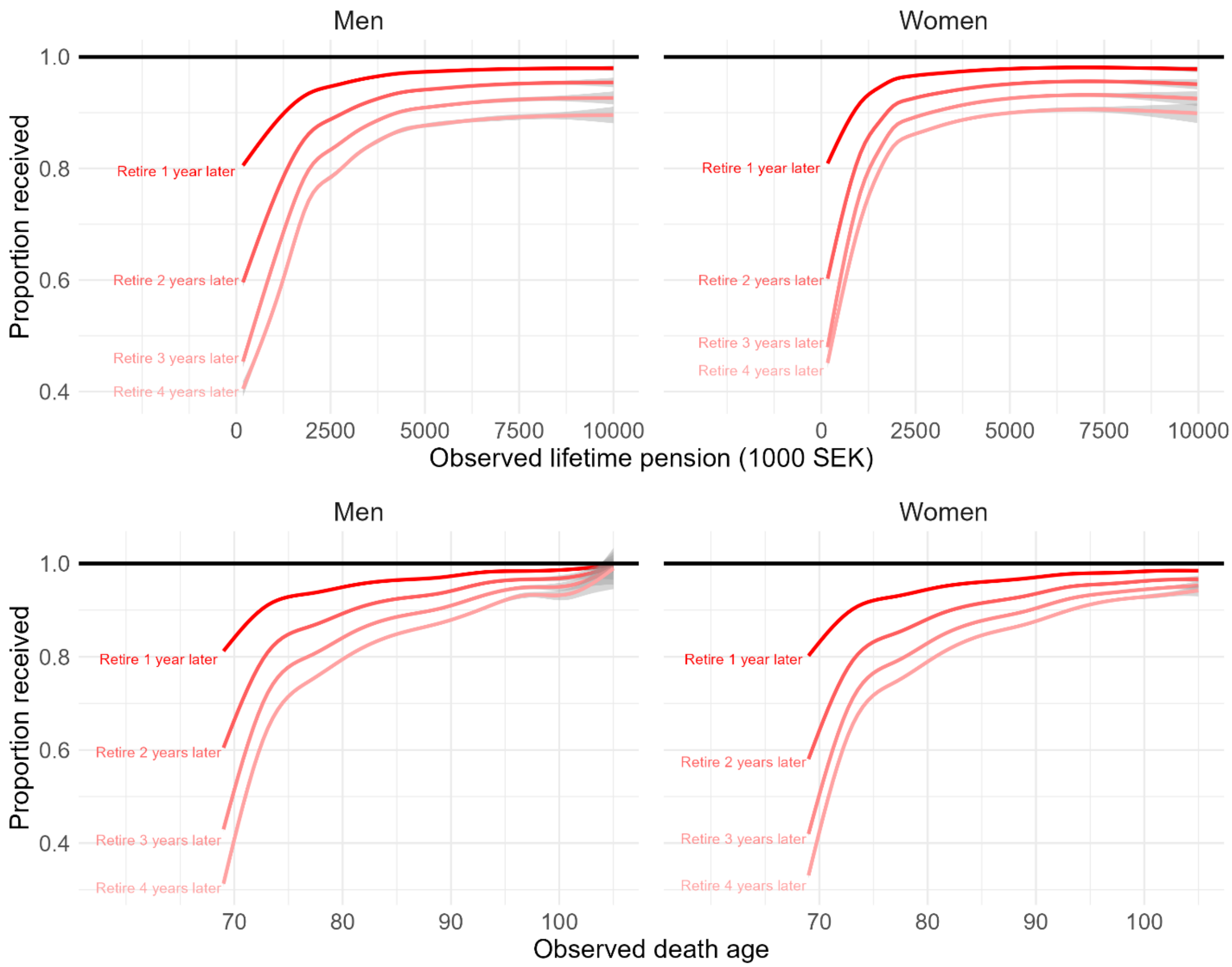

Figure A11. Proportion received across observed lifetime pension (upper panels) and observed death age (lower panels) in the scenarios of increasing retirement ages. *Source*: Authors' calculations based on Swedish register data.

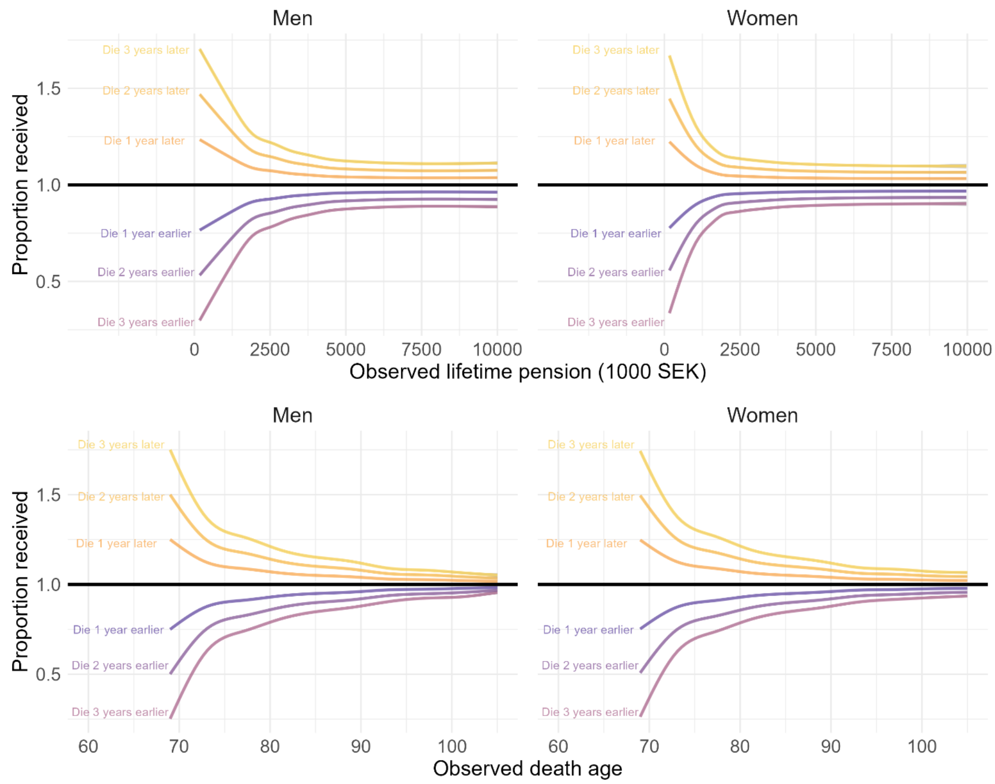

Figure A12. Proportion received across observed lifetime pension (upper panels) and observed death age (lower panels) in the scenarios of changing lifespans. *Source*: Authors' calculations based on Swedish register data.

Appendix 3. Tables

Table A1. Observed yearly pension trajectory from age 80, 1925 cohort

	Mean pension at age 80 (1000 SEK)		Average % change in pension income					
			1 year later		5 years later		10 years later	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Men								
Bottom 20%	125.49	25.13	-0.84	10.67	0.60	6.58	1.52	6.28
Second 20%	170.52	6.99	0.40	1.74	0.49	5.12	0.35	5.93
Third 20%	195.05	7.55	0.33	1.75	0.28	3.99	0.10	5.50
Fourth 20%	228.69	12.31	0.06	2.26	-0.24	5.00	-0.46	6.74
Top 20%	338.06	145.77	-0.57	4.50	-1.25	7.65	-1.56	9.46
Total	211.67	97.87	-0.12	5.40	-0.06	5.88	-0.15	7.18
Women								
Bottom 20%	73.35	10.63	-0.56	68.85	-0.90	12.20	0.72	20.90
Second 20%	93.80	4.18	-0.55	6.68	-1.97	27.09	-1.99	33.98
Third 20%	112.12	6.98	-0.06	8.95	-0.26	16.92	0.04	14.86
Fourth 20%	142.55	10.83	0.06	4.64	0.13	11.20	0.99	16.93
Top 20%	209.30	61.53	-0.31	4.26	-0.73	19.35	-0.46	11.20
Total	126.55	55.11	-0.28	30.11	-0.76	18.56	-0.20	21.23

Source: Authors' calculations based on Swedish register data. *Notes:* Since we only imputed pension data for ages 80 and above, here we only show the trajectories from age 80 for cohorts where we have observed data. In our data, individual yearly pension was stable from around age 70. Individuals were grouped into 20% groups based on their pension income at age 80. Changes relative to pension income at age 80 at 1, 5, 10 years later correspond to pension income at ages 81, 85, and 90. For men the changes are very minor. Changes are larger for women as many benefited from changes making the guarantee pension more generous, as well as occasionally the deaths of their husbands.

Table A2. Descriptive Statistics

Variable	Mean	SD	Min	Max
Women	0.51	0.50	0	1
Cohort				
Cohort 1918~1924	0.36	0.48	0	1
Cohort 1925~1929	0.22	0.42	0	1
Cohort 1930~1934	0.20	0.40	0	1
Cohort 1935~1939	0.22	0.41	0	1
Lifetime pension (1000 SEK)	3190.85	2483.34	3.00	230126.77
Lifespan at age 65 (year)	19.48	8.86	0.00	41
Pre-retirement earnings (1000 SEK)	208.98	147.29	3.00	21498.63
Yearly pension age age 70 (1000 SEK)	172.51	102.98	3.00	7842.53
Occupation (EGP)				
I (higher grade professionals)	0.07	0.26	0	1
II (lower grade professionals)	0.13	0.33	0	1
IIIa (higher grade non-manual employees)	0.08	0.27	0	1
IIIb (lower grade non-manual employees)	0.07	0.25	0	1
IVa+b (Small proprietors, artisans, etc.)	0.06	0.23	0	1
IVc (farmers and self-employed workers)	0.04	0.20	0	1
V+VI (skilled workers)	0.11	0.31	0	1
VIIa+b (non-skilled workers)	0.27	0.45	0	1
NA (including those not employed)	0.18	0.38	0	1
Education				
Primary school	0.65	0.48	0	1
Secondary school	0.25	0.43	0	1
Any college and above	0.08	0.28	0	1
Education missing	0.03	0.16	0	1
Years of education	8.88	2.56	7	19
Civil status				
Married	0.77	0.42	0	1
Divorced/separated	0.10	0.30	0	1
Widowed	0.03	0.17	0	1
Never married	0.10	0.29	0	1
Metropolitan county	0.34	0.47	0	1
N		1694133		

Source: Authors' calculations based on Swedish register data. *Notes:* We used an eight-category version of the EGP scheme. I: higher grade professionals, administrators, and officials; managers in large industrial establishments, and large proprietors. II: lower grade professionals, administrators, and officials; higher grade technicians; managers in small industrial establishments; supervisors of non-manual employees. IIIa: higher grade routine non-manual employees (administration and commerce). IIIb: lower grade routine non-manual employees (sales and services). IVa+b: small proprietors, artisans, and so on, with and without employees. IVc: farmers and small holders; self-employed workers in primary production. V+VI: skilled workers. VIIa+b: non-skilled workers and agricultural laborers.

Table A3. Pearson correlation coefficients between key variables, cohort combined.

	Years of education	Pre-retirement earnings	Pension at age 70	Lifetime pension	Lifespan
Men					
Years of education	1.00				
Pre-retirement earnings	0.44	1.00			
Pension at age 70	0.45	0.79	1.00		
Lifetime pension	0.36	0.60	0.77	1.00	
Lifespan	0.11	0.13	0.14	0.70	1.00
Women					
Years of education	1.00				
Pre-retirement earnings	0.42	1.00			
Pension at age 70	0.44	0.80	1.00		
Lifetime pension	0.39	0.63	0.75	1.00	
Lifespan	0.09	0.08	0.08	0.65	1.00

Source: Authors' calculations based on Swedish register data.

Table A4. Gini and additional inequality measures for lifetime pension by cohort, men.

Cohort	N	Mean	Gini	P90/P10	P90/P50	P50/P10	S80/S20	S90/S40
1918	40338	2871.44	0.413	11.59	2.25	5.15	13.34	2.03
1919	40203	2919.79	0.410	11.18	2.22	5.04	13.14	2.00
1920	48365	2984.38	0.404	10.77	2.19	4.92	12.90	1.91
1921	45081	3056.47	0.403	10.85	2.17	5.00	12.86	1.90
1922	41183	3080.86	0.397	11.25	2.16	5.21	12.47	1.82
1923	40594	3181.84	0.396	11.65	2.13	5.48	12.40	1.82
1924	39336	3222.02	0.398	10.81	2.16	5.02	12.48	1.84
1925	38583	3340.86	0.394	9.99	2.11	4.72	12.14	1.80
1926	36883	3423.02	0.384	9.36	2.06	4.54	10.93	1.69
1927	35641	3510.06	0.381	9.29	2.03	4.58	10.77	1.67
1928	35880	3626.93	0.381	9.39	2.03	4.63	10.74	1.67
1929	34305	3685.83	0.381	9.06	2.01	4.51	10.63	1.67
1930	34662	3804.43	0.376	8.70	1.99	4.36	10.21	1.62
1931	34169	3885.81	0.377	8.91	2.02	4.42	10.39	1.63
1932	33803	4013.30	0.372	8.31	1.99	4.18	9.87	1.58
1933	32401	4076.43	0.368	8.19	1.98	4.14	9.74	1.54
1934	32434	4163.64	0.367	8.05	1.98	4.06	9.56	1.53
1935	33492	4339.94	0.367	8.09	2.00	4.05	9.50	1.53
1936	34569	4458.97	0.365	7.75	1.99	3.90	9.26	1.51
1937	35468	4563.65	0.368	7.74	2.00	3.86	9.33	1.54
1938	37010	4642.66	0.378	8.20	2.06	3.98	9.98	1.64
1939	38544	4719.75	0.377	8.08	2.07	3.90	9.86	1.64

Source: Authors' calculations based on Swedish register data. *Notes*: P90/P10 refers to the ratio between the 90th and 10th percentiles. P90/P50 refers to the ratio between the 90th and 50th percentiles. P50/P10 refers to the ratio between the 50th and 10th percentiles. S80/S20 refers to the share ratio of lifetime pension between the top 20% and bottom 20%. S90/S40 refers to the share ratio of lifetime pension between the top 90% and bottom 40%.

Table A5. Gini and additional inequality measures for lifetime pension by cohort, women.

Cohort	N	Mean	Gini	P90/P10	P90/P50	P50/P10	S80/S20	S90/S40
1918	43065	2110.34	0.379	7.67	2.33	3.30	8.83	1.61
1919	42673	2140.39	0.375	7.49	2.28	3.29	8.79	1.57
1920	51742	2182.39	0.373	7.44	2.26	3.29	8.80	1.56
1921	48185	2250.07	0.365	7.11	2.20	3.23	8.42	1.49
1922	44362	2258.73	0.367	7.25	2.20	3.29	8.56	1.50
1923	43119	2323.59	0.361	6.99	2.15	3.25	8.36	1.46
1924	42162	2374.26	0.358	6.97	2.13	3.28	8.28	1.43
1925	41037	2476.71	0.358	7.08	2.13	3.32	8.31	1.43
1926	39300	2563.52	0.349	6.73	2.09	3.22	7.74	1.35
1927	38056	2651.37	0.348	6.58	2.10	3.14	7.58	1.35
1928	38409	2684.93	0.342	6.23	2.05	3.04	7.31	1.30
1929	36546	2761.51	0.341	6.20	2.03	3.05	7.27	1.30
1930	37123	2862.28	0.340	6.18	2.02	3.06	7.21	1.30
1931	36059	2968.17	0.340	6.02	2.01	2.99	7.16	1.29
1932	35805	3022.41	0.335	5.87	1.97	2.98	6.99	1.27
1933	33927	3128.59	0.331	5.79	1.96	2.96	6.80	1.24
1934	34223	3164.09	0.333	5.93	1.96	3.03	7.01	1.25
1935	34314	3307.48	0.332	5.92	1.94	3.06	6.98	1.24
1936	36090	3393.34	0.331	5.78	1.93	2.99	6.99	1.23
1937	36568	3474.79	0.328	5.69	1.87	3.04	6.90	1.22
1938	38523	3514.25	0.331	5.79	1.88	3.08	7.11	1.24
1939	39901	3572.41	0.338	6.23	1.88	3.31	7.61	1.29

Source: Authors' calculations based on Swedish register data. *Notes*: P90/P10 refers to the ratio between the 90th and 10th percentiles. P90/P50 refers to the ratio between the 90th and 50th percentiles. P50/P10 refers to the ratio between the 50th and 10th percentiles. S80/S20 refers to the share ratio of lifetime pension between the top 20% and bottom 20%. S90/S40 refers to the share ratio of lifetime pension between the top 90% and bottom 40%.

Table A6. Gini in lifetime pension income in hypothetical scenarios, 1928 cohort.

	Men		Women	
	Gini	Change	Gini	Change
Observed	0.381	-	0.342	-
Raising minimum pension	0.366	-3.8%	0.272	-20.3%
Adding tax	0.374	-1.9%	0.324	-5.3%
Changing retirement ages				
1 year later	0.393	3.2%	0.351	2.6%
2 years later	0.406	6.6%	0.361	5.6%
3 years later	0.417	9.5%	0.368	7.8%
4 years later	0.425	11.5%	0.373	9.0%
Changing death ages				
3 years earlier	0.394	3.4%	0.354	3.6%
2 years earlier	0.390	2.4%	0.350	2.3%
1 year earlier	0.386	1.2%	0.345	1.0%
1 year later	0.368	-3.3%	0.333	-2.5%
2 years later	0.356	-6.5%	0.326	-4.8%
3 years later	0.346	-9.2%	0.319	-6.7%

Source: Authors' calculations based on Swedish register data. *Note*: In all the calculations, inflation is adjusted to SEK in the year 2018.

References

Kim, ChangHwan, and Arthur Sakamoto. 2008. "The Rise of Intra-Occupational Wage Inequality in the United States, 1983 to 2002." *American Sociological Review* 73(1):129–157.

Lee, Ronald D. 1994. "Population Age Structure, Intergenerational Transfer, and Wealth: A New Approach, with Applications to the United States." *Journal of Human Resources* 1027–1063.

Lindeman, Richard H., Peter F Merenda, and Ruth Z Gold. 1980. *Introduction to Bivariate and Multivariate Analysis*. Glenview, IL: Scott, Foresman and Company

Meng, Xin, Kailing Shen, and Sen Xue. 2013. "Economic Reform, Education Expansion, and Earnings Inequality for Urban Males in China, 1988–2009." *Journal of Comparative Economics* 41(1):227–244.

Statistics Sweden. 2020. "The Future Population of Sweden 2020–2070." *Statistiska Meddelanden* BE 18 SM 2001.

Xie, Yu and Xiang Zhou. 2014. "Income Inequality in Today's China." *Proceedings of the National Academy of Sciences* 111(19):6928–6933.

Zhou, Xiang. 2014. "Increasing Returns to Education, Changing Labor Force Structure, and the Rise of Earnings Inequality in Urban China, 1996–2010." *Social Forces* 93(2):429–455.