ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Why Are the Wealthiest So Wealthy? New Longitudinal Empirical Evidence and Implications for Theories of Wealth Inequality: Supplemental Material

Authors	Elin Halvorsen, Joachim Hubmer, Serdar Ozkan, and Sergio Salgado
Working Paper Number	2025-033A
Creation Date	December 2025
Citable Link	https://doi.org/10.20955/wp.2025.033
Suggested Citation	Halvorsen, E., Hubmer, J., Ozkan, S., Salgado, S., 2025; Why Are the Wealthiest So Wealthy? New Longitudinal Empirical Evidence and Implications for Theories of Wealth Inequality: Supplemental Material, Federal Reserve Bank of St. Louis Working Paper 2025-033. URL https://doi.org/10.20955/wp.2025.033

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

Why Are the Wealthiest So Wealthy? New Longitudinal Empirical Evidence and Implications for Theories of Wealth Inequality: Supplemental Material[1]

Joachim Hubmer[2] ⓡ Elin Halvorsen[3] ⓡ Sergio Salgado[4] ⓡ Serdar Ozkan[5]

December 23, 2025

[1]The "ⓡ" symbol indicates certified random order for authors' names (Ray and Robson, 2018).

[2]University of Pennsylvania; jhubmer@sas.upenn.edu
[3]Statistics Norway; elin.halvorsen@ssb.no
[4]The Wharton School-University of Pennsylvania; ssalgado@wharton.upenn.edu
[5]St. Louis Fed, University of Toronto; serdar.ozkan@gmail.com; www.serdarozkan.me

Table of Contents

SM.1 Additional Tables

Table SM.1.1 – Basic Sample Statistics

Panel A: Population Shares							
		1995	2000	2005	2010	2015	2019
Age 18/24		15.66	14.14	14.17	13.35	14.87	13.93
Age 25/44		36.49	36.59	34.91	33.63	32.13	32.64
Age 45/64		25.12	27.92	30.18	31.06	30.43	30.25
Age 65+		22.74	21.35	20.74	21.97	22.57	23.18
Male Head		63.20%	62.60%	62.50%	62.60%	62.10%	
Panel B: Descriptive Statistics (US$ of 2018)							
	Mean	Std. Dev.	P10	P50	P90	P99	P99.9
Safe Assets	44,000	164,614	711	16,199	102,511	379,068	1,344,367
Public Equity	8,321	137,860	0	0	13,369	124,931	609,948
Private Equity	32,194	1,518,483	0	0	921	327,148	3,949,190
Housing	295,056	303,067	2,151	219,579	654,347	1,415,022	2,230,372
Debt	89,056	102,144	80	59,599	219,409	427,636	652,016
Net wealth	290,514	1,724,325	-10,736	176,690	632,432	1,677,903	6,132,153
Household Observations: 74.4 Million							

Notes: Table SM.1.1 show cross-sectional statistics of the population of households in Norway. Panel A shows, population shares for head of household. Panel B shows household-level wealth statistics in real US$ of 2018 (1 USD=8.14 NOK). To obtain these statistics, we first calculate cross sectional moments at the annual level and then we average the statistics across all years in the sample (1993 to 2019).

Table SM.1.2 – Income and Wealth Concentration

	Bottom 50	Top 10%	Top 5%	Top 1%	Top 0.1%	Top 0.01%
Labor Earnings	5.84	35.35	21.09	6.29	1.23	0.28
Total Income	20.83	31.31	21.47	10.94	5.71	3.07
Safe Assets	3.55	61.51	45.85	21.80	7.85	2.75
Public Equity	-0.00	92.33	82.26	56.06	28.86	15.04
Private Equity	-0.82	100.77	100.34	91.33	63.14	35.57
Housing	7.37	38.05	23.83	7.28	1.01	0.47
Debt	3.21	42.74	25.96	8.44	0.97	0.15
Net wealth	0.37	50.95	36.17	17.58	8.69	4.52

Notes: Table SM.1.2 show cross sectional concentration statistics at the household level. To calculate these statistics, we first calculate cross sectional moments at the annual level and then we average across all years in the sample (1993 to 2019). The concentration of net wealth deviates slightly from official statistics due to our use of alternative housing values.

Table SM.1.3 – Wealth Returns

	N 000s	Mean	SD.	Skew.	Kurt.	P1	P5	P10	P50	P90	P95	P99
Panel A: Individual-level returns: FGMP (2005/2015)												
All	27,318	0.034	0.203	0.296	23.974	-0.622	-0.236	-0.103	0.022	0.185	0.292	0.735
Equity	10,072	0.083	0.363	2.463	24.412	-0.793	-0.350	-0.157	0.035	0.374	0.605	1.476
Housing	21,333	0.045	0.201	2.668	30.762	-0.534	-0.229	-0.088	0.025	0.183	0.281	0.811
Safe	27,374	-0.012	0.034	4.787	50.634	-0.061	-0.043	-0.042	-0.017	0.020	0.030	0.104
Panel B: Individual-level returns: This paper (2005/2019)												
All	26,424	0.032	0.131	0.823	13.396	-0.390	-0.138	-0.067	0.019	0.149	0.222	0.490
Equity	10,888	0.071	0.230	3.750	31.728	-0.390	-0.150	-0.100	0.032	0.263	0.416	1.017
Housing	26,315	0.039	0.135	0.490	12.311	-0.449	-0.140	-0.047	0.022	0.153	0.227	0.514
Safe	26,219	0.001	0.024	3.553	37.369	-0.034	-0.027	-0.022	-0.002	0.026	0.034	0.072
Panel C: Household-level returns: FGMP (2005/2015)												
All	19,356	0.032	0.183	0.506	16.993	-0.576	-0.228	-0.100	0.022	0.177	0.276	0.655
Equity	7,977	0.084	0.381	3.018	30.796	-0.837	-0.352	-0.157	0.034	0.374	0.608	1.537
Housing	14,550	0.044	0.187	2.246	26.583	-0.504	-0.215	-0.084	0.028	0.179	0.275	0.738
Safe	19,431	-0.013	0.030	3.387	31.369	-0.057	-0.042	-0.042	-0.017	0.020	0.029	0.083
Panel D: Household-level returns: This paper (2005/2019)												
All	17,807	0.030	0.123	0.701	15.120	-0.376	-0.129	-0.060	0.018	0.141	0.205	0.431
Equity	8,508	0.070	0.239	4.007	37.861	-0.407	-0.154	-0.100	0.031	0.263	0.415	1.030
Housing	17,723	0.040	0.131	0.804	16.300	-0.405	-0.134	-0.046	0.024	0.152	0.219	0.481
Safe	17,675	0.001	0.022	2.687	25.537	-0.032	-0.027	-0.022	-0.001	0.026	0.034	0.068

Notes: Table SM.1.3 shows cross sectional statistics of the returns distribution for different asset classes based on a pooled sample of households between 2005 and 2019. We calculate returns following Fagereng et al (2020). Equity corresponds to the returns on private and publicly traded firms. The Kelley skewness is defined as $\mathcal{S}_{\mathcal{K}} = \frac{P90-P50}{P90-P10} - \frac{P50-P10}{P90-P10}$.

Table SM.1.4 – Sample Statistics: US SCF Data

Descriptive Statistics (US$ of 2018)							
	Mean	SD	P10	P50	P90	P99	P99.9
Safe Assets	125,615	602,358	85	16,521	281,479	1,620,551	5,924,482
Public Equity	84,644	1,109,028	0	0	76,413	1,569,328	9,102,842
Private Equity	91,180	1,825,445	0	0	7,301	1,574,025	12,985,575
Housing	237,051	1,477,831	0	98,010	457,038	2,389,650	10,598,920
Gross Wealth	538,491	3,293,036	382	143,885	938,809	7,116,825	31,126,536
Debt	78,513	532,779	-2	12,596	194,056	694,872	2,637,272
Net wealth	459,978	3,113,103	-1,741	78,847	801,826	6,685,830	27,845,214

Notes: Table SM.1.4 show cross sectional statistics of the population of households in the United States using data from SCF in real US$ of 2018. To obtain these statistics, we first calculate cross sectional moments at the annual level and then we average the statistics across all years in the sample after 1989.

Table SM.1.5 – Income and Wealth Concentration: US SCF Data

	Bottom 50	Top 10%	Top 5%	Top 1%	Top 0.1%	Top 0.01%
Income	9.41	49.92	38.58	21.42	8.32	3.03
Safe Assets	1.60	70.35	55.24	28.46	9.16	2.89
Public Equity	-0.04	95.77	87.27	59.96	25.47	8.90
Private Equity	-0.01	99.95	97.48	77.97	36.47	13.67
Housing	4.77	59.39	47.08	26.87	11.17	4.87
Gross Wealth	3.86	68.31	56.79	33.12	12.06	3.88
Debt	-0.08	58.84	43.93	23.31	10.95	5.59
Net wealth	1.78	73.37	61.55	36.24	13.28	4.33

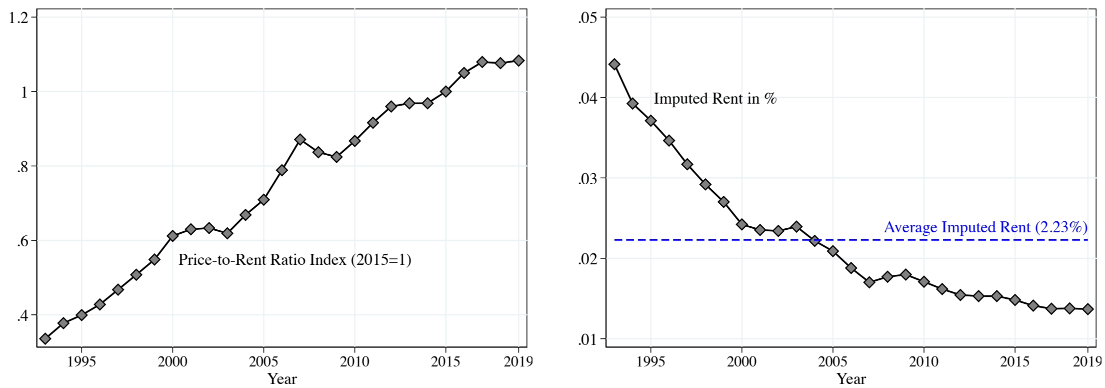
Notes: Table SM.1.5 show cross sectional statistics of the population of households in the United States using data from SCF in real US$ of 2018. To obtain these statistics, we first calculate cross sectional moments at the annual level and then we average the statistics across all years in the sample after 1989.

SM.2 Additional Figures

SM.2.1 Aggregate Time Series and Cross Sectional Moments

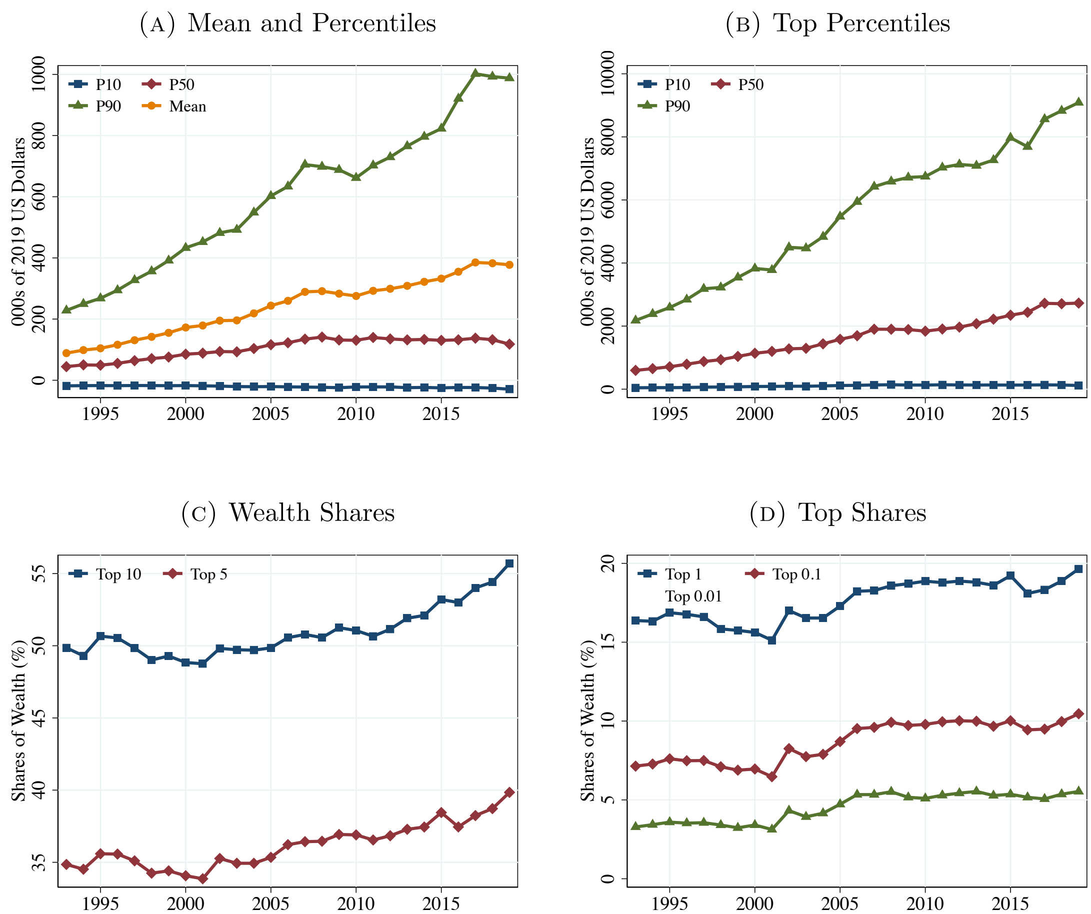
Figure SM.2.1 – House Prices and Rental Equivalence in Norway

(a) Evolution of House-to-Price Ratio in Norway (b) Rental Equivalence for Homeowners



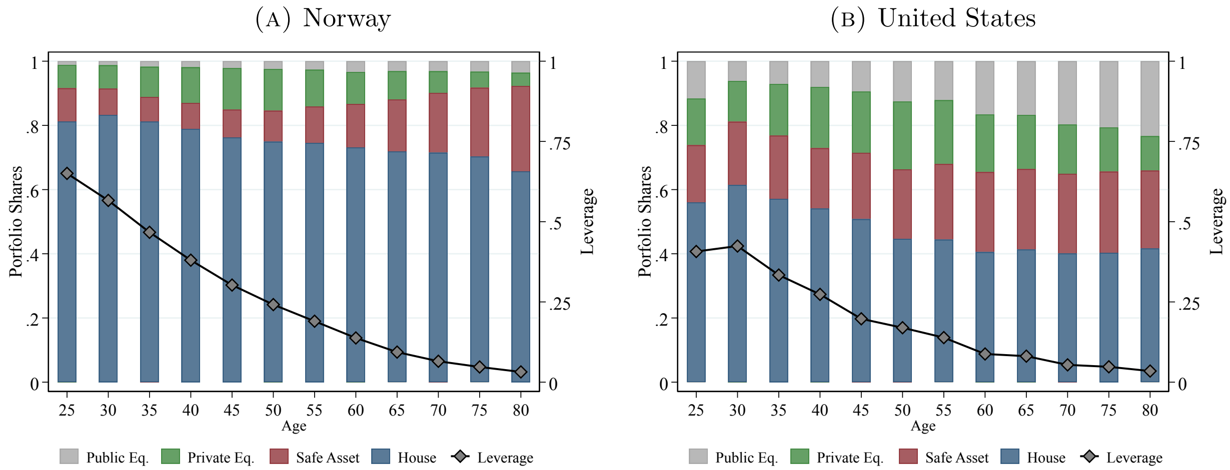
Notes: Imputed rental equivalence for homeowners, based on Eika et al. (2020) and time-varying price-to-rent ratio from the Analytical Price Indicators of the OECD available here. Rental income is equal to House Value$_{i,t}$ $\times$ (rent$_t$/price$_t$).

Figure SM.2.2 – Time Series of Wealth and Concentration



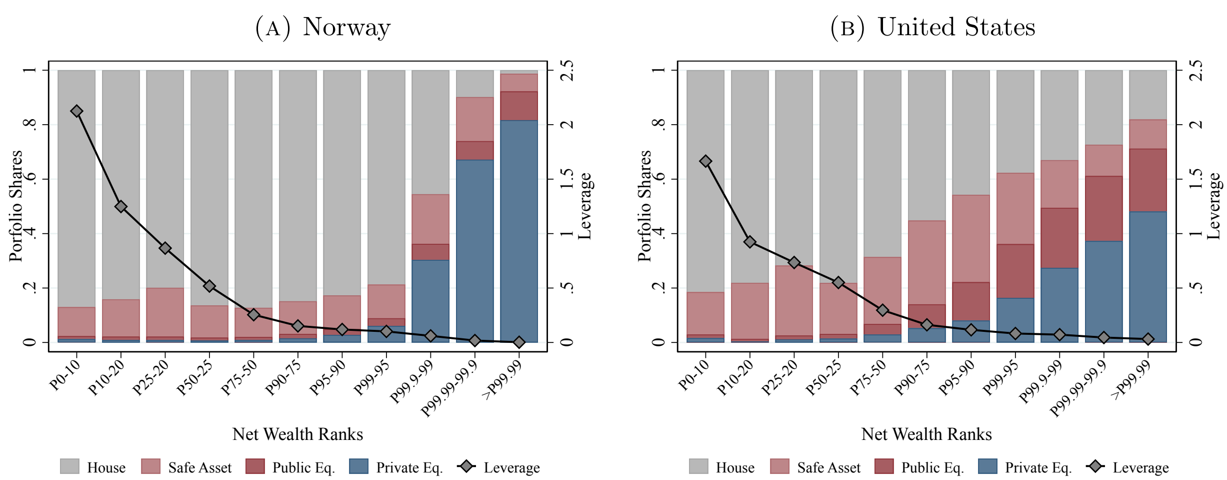
Notes: Figure SM.2.2 shows time series of different moments of the wealth distribution in Norway.

Figure SM.2.3 – Portfolio Composition over the Life Cycle



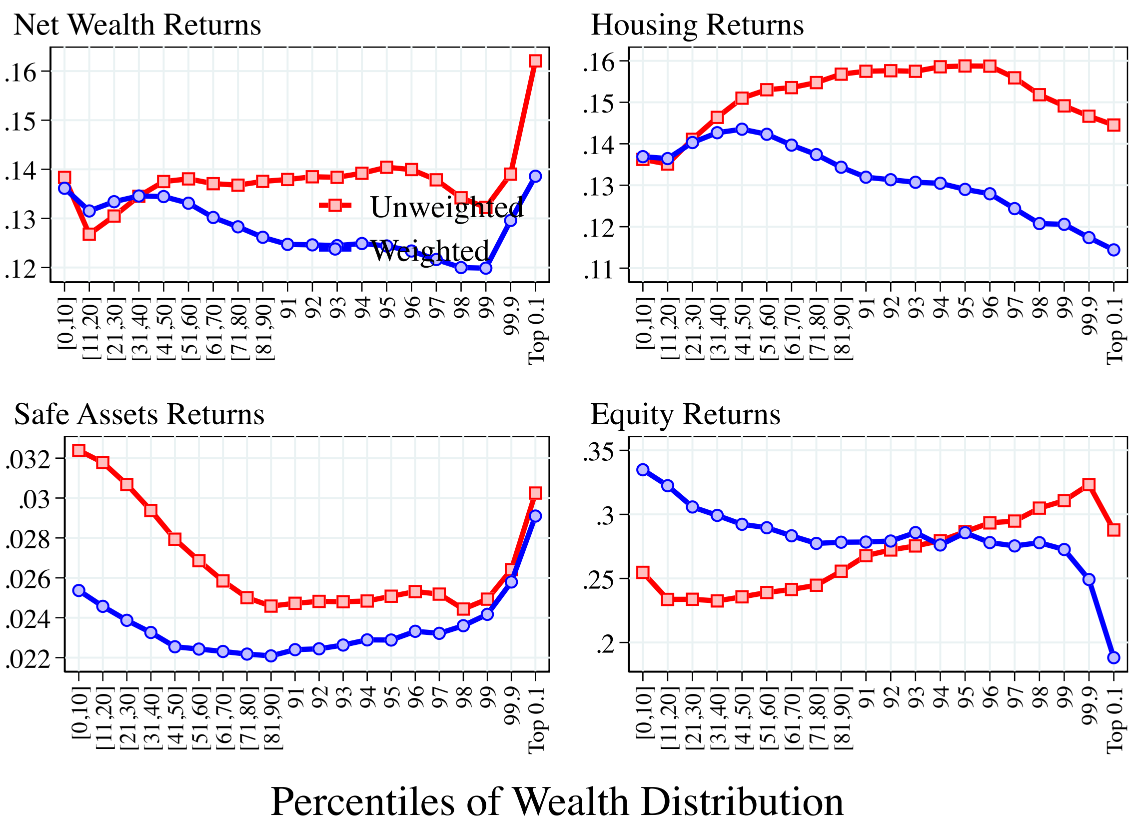
Notes: Figure SM.2.3 shows the portfolio shares and leverage within five-year age groups labeled by their starting age (25-29, 30-34, and so on) for Norway and the United States. For comparability, private equity is the net value of private businesses. Leverage does not include firm debt. Portfolio shares are calculated as the ratio between the value of all assets in a particular category (e.g. total value of safe assets) over the total value of gross wealth (i.e. sum of wealth in housing, safe assets, public equity, and private equity) within an age group. Similarly, within-group leverage, is the ratio between the sum all debt (e.g. mortgages, student debt, credit card debt) within a wealth rank and age group and the sum of all total assets within the same group.

Figure SM.2.4 – Portfolio Composition over the Wealth Distribution



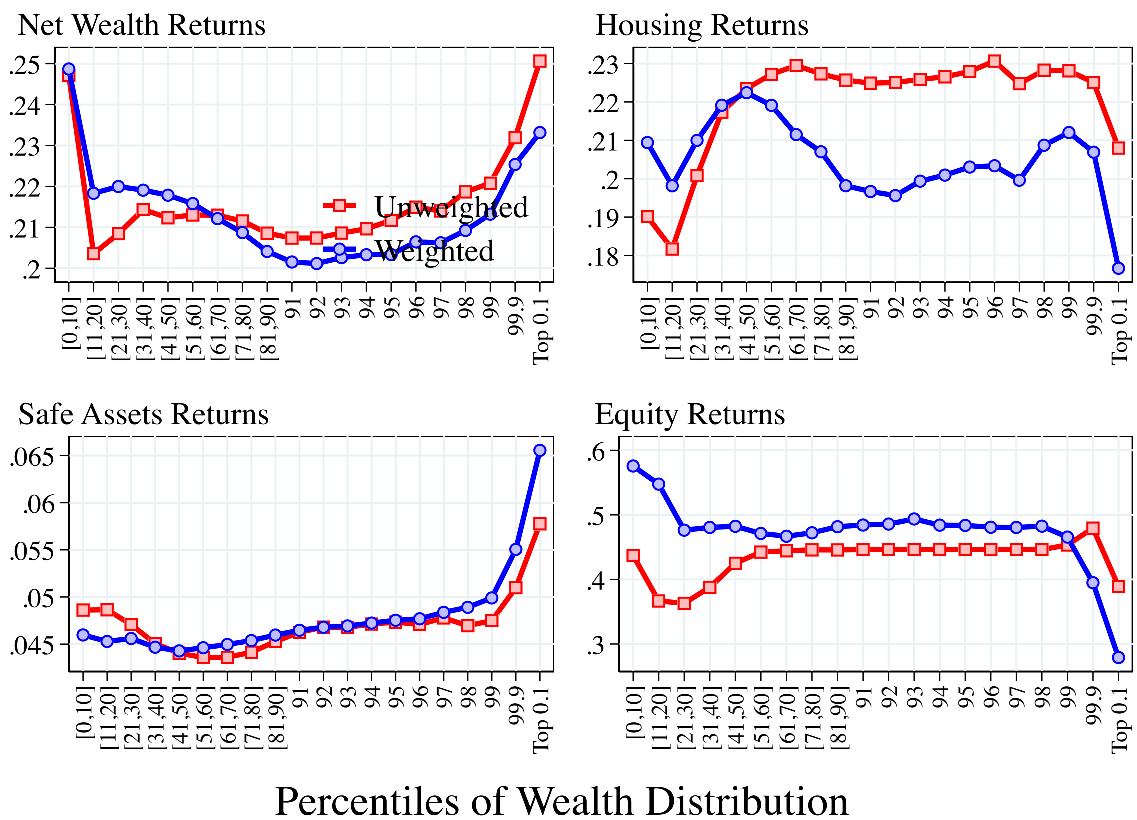
Notes: Figure SM.2.4 shows the portfolio shares and leverage within wealth percentiles for Norway and the United States. For comparability, private equity is the net value of private businesses. Leverage does not include firm debt. Portfolio shares are calculated as the ratio between the value of all assets in a particular category (e.g. total value of safe assets) over the total value of gross wealth (i.e. sum of wealth in housing, safe assets, public equity, and private equity) within an wealth group. Similarly, within-group leverage, is the ratio between the sum all debt (e.g. mortgages, student debt, credit card debt) within a wealth group and the sum of all total assets within the same group.

Figure SM.2.5 – Cross-Sectional Standard Deviation of Returns



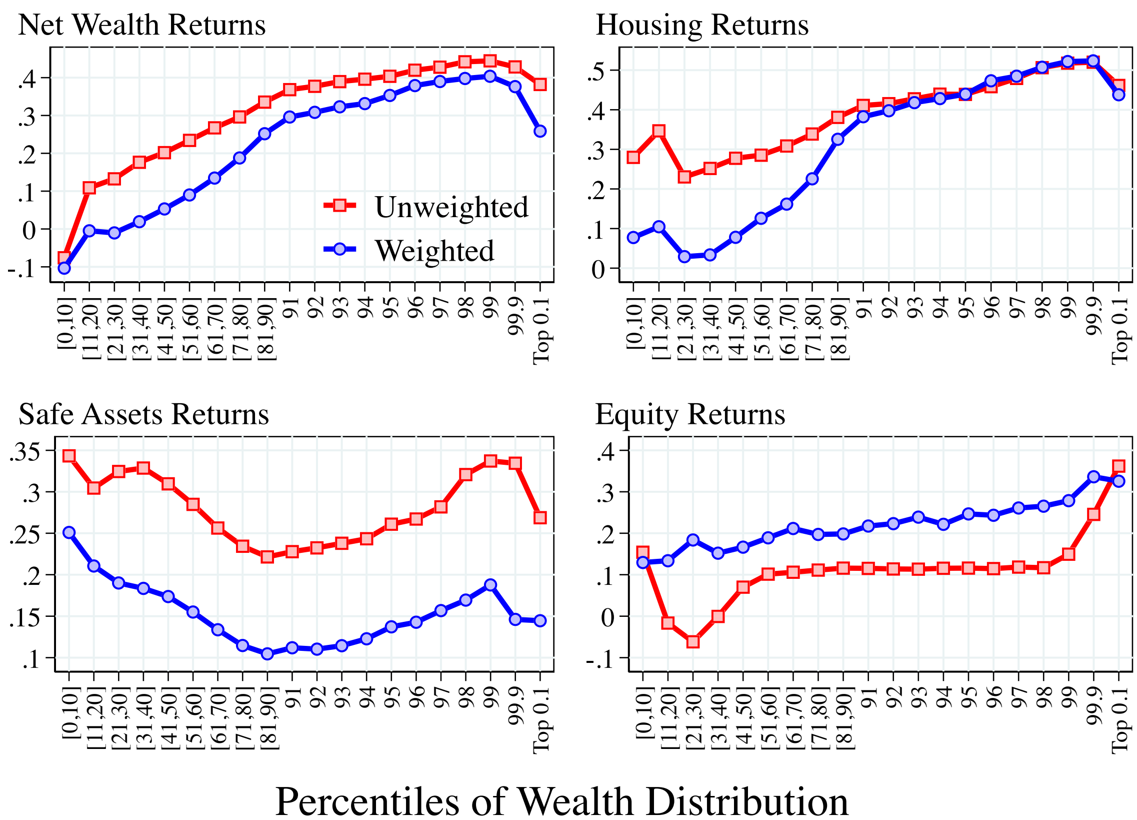
Notes: Figure SM.2.5 shows the standard deviation returns within different quantiles of the households net worth distribution. To construct this figure, we pool household observations between 2005 and 2019. Weighted averages are weighted using the value of the corresponding asset. Negative or missing asset values are assigned a weight of 0.

Figure SM.2.6 – Cross Sectional P90-P10 of Returns



Notes: Figure SM.2.6 shows the P90-P10 of returns within different quantiles of the households net worth distribution. To construct this figure, we pool household observations between 2005 and 2019. Weighted averages are weighted using the value of the corresponding asset. Negative or missing asset values are assigned a weight of 0.

Figure SM.2.7 – Cross Sectional Kelley Skewness of Returns

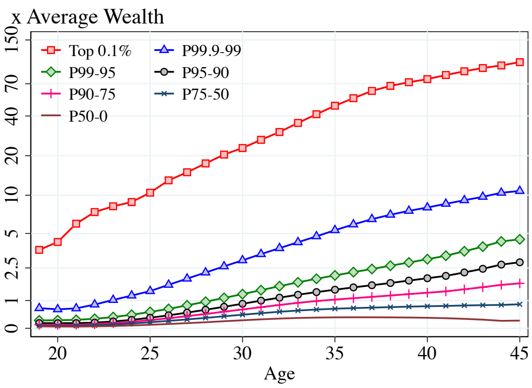


Notes: Figure SM.2.7 shows the Kelley Skewness returns within different quantiles of the households net worth distribution. To construct this figure, we pool household observations between 2005 and 2019. Weighted averages are weighted using the value of the corresponding asset. Negative or missing asset values are assigned a weight of 0. Kelley Skewness is defined as $\mathcal{S}_{\mathcal{K}} = \frac{P90-P50}{P90-P10} - \frac{P50-P10}{P90-P10}$.

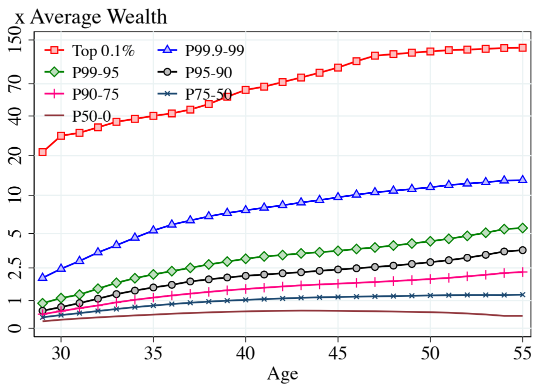
SM.2.2 Backward-Looking Results

Figure SM.2.8 – Backward-Looking Wealth Profiles: Age Groups

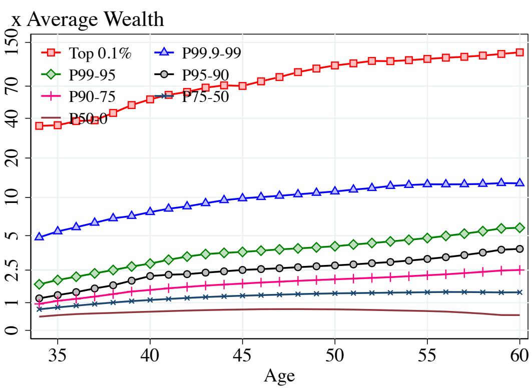
(a) Age [45–49]



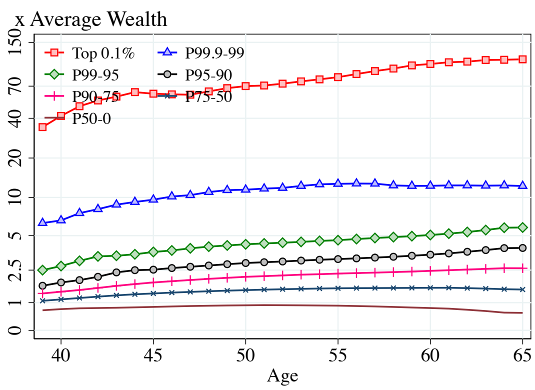
(b) Age [55–59]



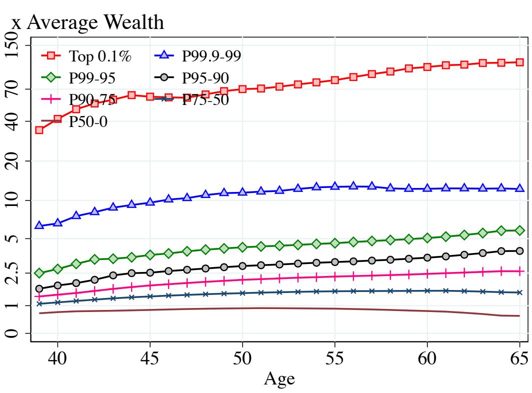
(c) Age [60–64]



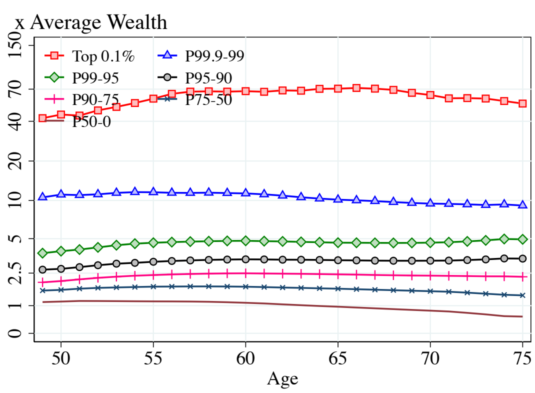
(d) Age [65–69]



(e) Age [70–74]



(f) Age [75–79]



Notes: Figure SM.2.8 shows the evolution of average wealth for different wealth groups conditional on their wealth at the end of the sample period sorted by BW_j^h.

Figure SM.2.9 – Long-Term Transition Matrix: Age Groups

(a) Age [45–49]

End-of-Period Wealh Rank, BW^h_j \ Initial Average Wealth Rank	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	62.7	22.2	10.3	2.8	1.7	0.3	0.0
(50-75]	46.5	28.1	16.5	4.9	3.4	0.5	0.0
(75-90]	36.4	28.1	21.0	7.4	6.0	1.2	0.0
(90-95]	27.4	27.1	23.8	9.7	9.4	2.5	0.1
(95-99]	19.4	24.5	23.6	12.4	14.7	4.9	0.5
(99-99.9]	12.1	19.4	22.2	13.2	19.3	10.3	3.5
Top 0.1%	4.8	9.9	15.1	8.5	19.1	20.5	22.2

(b) Age [55–59]

End-of-Period Wealh Rank, BW^h_j \ Initial Average Wealth Rank	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	67.6	20.7	8.5	2.0	1.1	0.1	0.0
(50-75]	42.9	31.4	17.6	4.8	2.9	0.3	0.0
(75-90]	29.7	30.0	23.9	8.8	6.6	1.0	0.0
(90-95]	21.9	24.5	26.4	12.3	12.2	2.7	0.1
(95-99]	15.1	19.3	24.7	14.8	19.2	6.5	0.4
(99-99.9]	10.4	12.7	18.8	13.2	23.9	17.1	4.0
Top 0.1%	4.1	5.0	9.5	6.4	18.3	29.5	27.1

(c) Age [60–64]

End-of-Period Wealh Rank, BW^h_j \ Initial Average Wealth Rank	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	69.4	20.1	7.8	1.7	0.8	0.1	0.0
(50-75]	41.7	32.7	17.9	4.8	2.6	0.3	0.0
(75-90]	27.5	30.5	25.3	9.2	6.6	0.9	0.0
(90-95]	19.3	23.7	27.6	13.6	13.2	2.5	0.1
(95-99]	12.5	17.5	25.1	15.8	21.8	6.9	0.4
(99-99.9]	8.0	10.7	15.8	12.2	27.5	21.8	4.0
Top 0.1%	2.5	6.0	9.0	5.1	16.0	31.8	29.7

(d) Age [65–69]

End-of-Period Wealh Rank, BW^h_j \ Initial Average Wealth Rank	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	70.1	20.0	7.5	1.6	0.7	0.1	0.0
(50-75]	41.5	33.3	18.0	4.6	2.4	0.2	0.0
(75-90]	26.4	31.1	26.0	9.4	6.4	0.7	0.0
(90-95]	17.3	23.3	29.1	14.3	13.6	2.3	0.1
(95-99]	11.1	15.5	24.7	17.1	24.2	7.1	0.4
(99-99.9]	6.2	7.9	13.7	11.5	30.4	26.3	4.1
Top 0.1%	2.0	5.3	7.2	4.3	13.5	34.7	33.0

(e) Age [70-74]

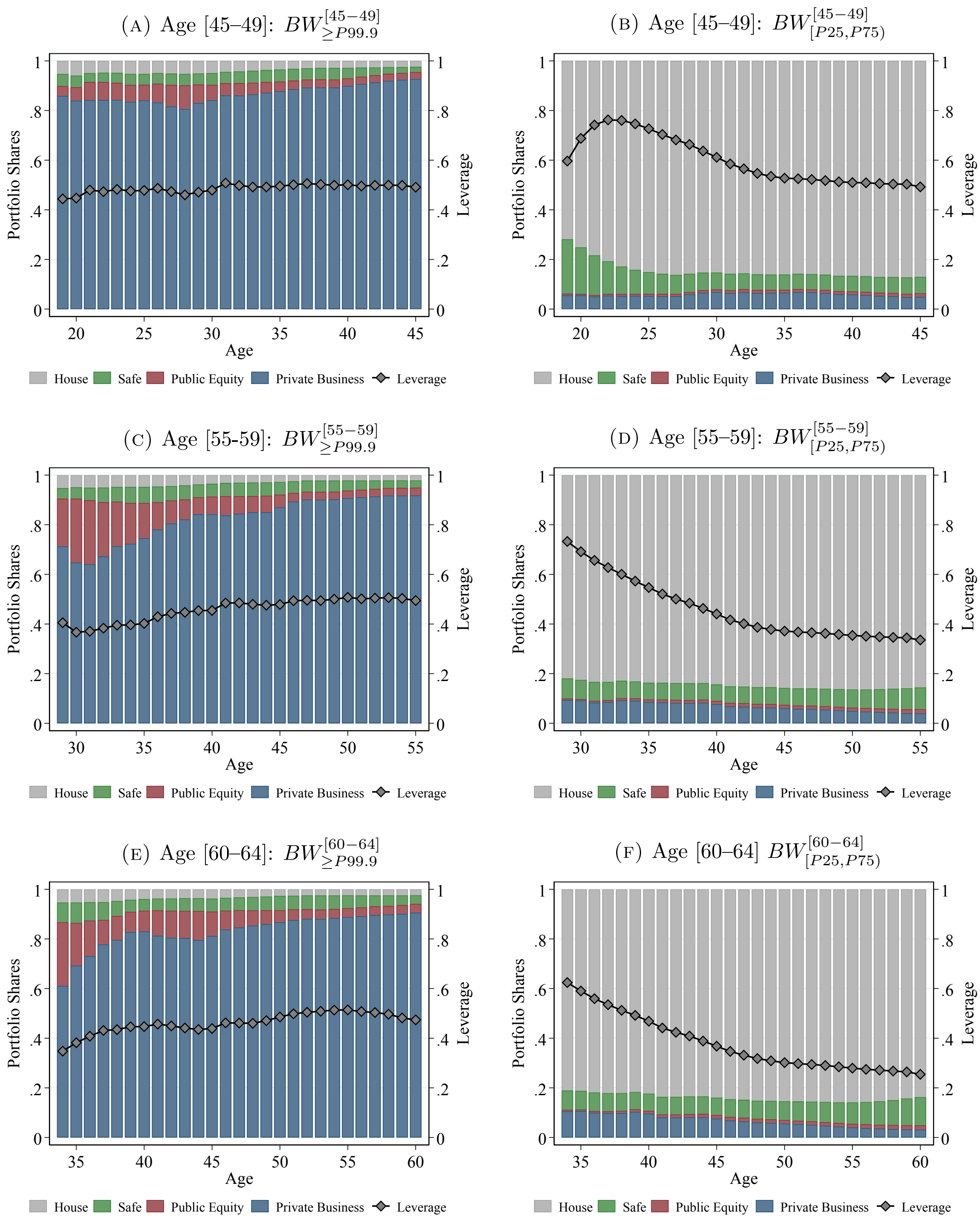
End-of-Period Wealh Rank, BW^h_j \ Initial Average Wealth Rank	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	70.4	19.7	7.6	1.6	0.7	0.1	0.0
(50-75]	41.9	33.6	17.5	4.4	2.3	0.2	0.0
(75-90]	26.1	31.8	26.2	9.1	6.2	0.7	0.0
(90-95]	16.2	23.5	30.1	14.8	13.3	2.0	0.1
(95-99]	9.5	14.4	25.1	18.3	25.6	6.8	0.3
(99-99.9]	4.7	5.5	11.4	12.1	32.9	29.1	4.3
Top 0.1%	1.4	3.3	5.1	5.2	14.1	39.1	31.8

(f) Age [75–79]

End-of-Period Wealh Rank, BW^h_j \ Initial Average Wealth Rank	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	70.6	19.3	7.6	1.7	0.7	0.1	0.0
(50-75]	42.6	33.5	17.0	4.4	2.3	0.2	0.0
(75-90]	25.9	32.7	26.2	8.6	5.7	0.7	0.0
(90-95]	15.6	24.2	31.2	14.3	12.7	2.0	0.1
(95-99]	9.6	13.5	25.3	19.1	25.9	6.4	0.3
(99-99.9]	3.3	4.6	9.9	11.9	36.6	29.3	4.4
Top 0.1%	0.4	3.2	3.2	5.0	16.7	37.4	34.1

Notes: Figure SM.2.9 shows the intragenerational persistence of net wealth. Figure SM.2.9 shows the results by first sorting household whose head is in different age groups in the conditioning year and then again by $\overline{W}_{i,1993}$. Each cell represent the fraction of household in different percentiles of the wealth distribution in $\overline{W}_{i,1993}$ (columns), conditional on their percentile of the wealth distribution in the conditioning year, BW_j^h (rows).

Figure SM.2.10 – Backward-Looking Portfolio Shares: Age Groups



(a) Age [45–49]: $BW^{[45-49]}_{\geq P99.9}$

(b) Age [45–49]: $BW^{[45-49]}_{[P25,P75)}$

(c) Age [55-59]: $BW^{[55-59]}_{\geq P99.9}$

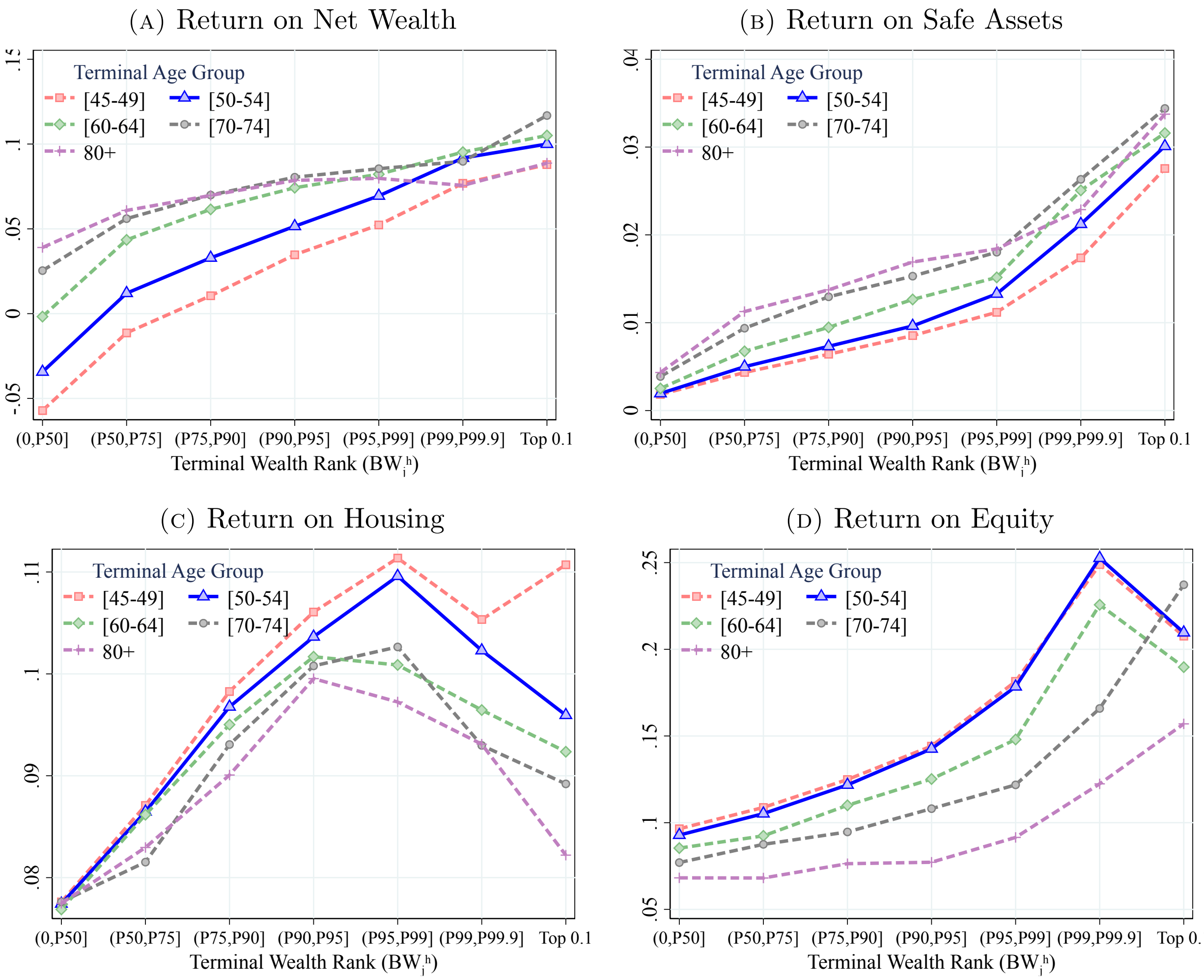
(d) Age [55–59]: $BW^{[55-59]}_{[P25,P75)}$

(e) Age [60–64]: $BW^{[60-64]}_{\geq P99.9}$

(f) Age [60–64] $BW^{[60-64]}_{[P25,P75)}$

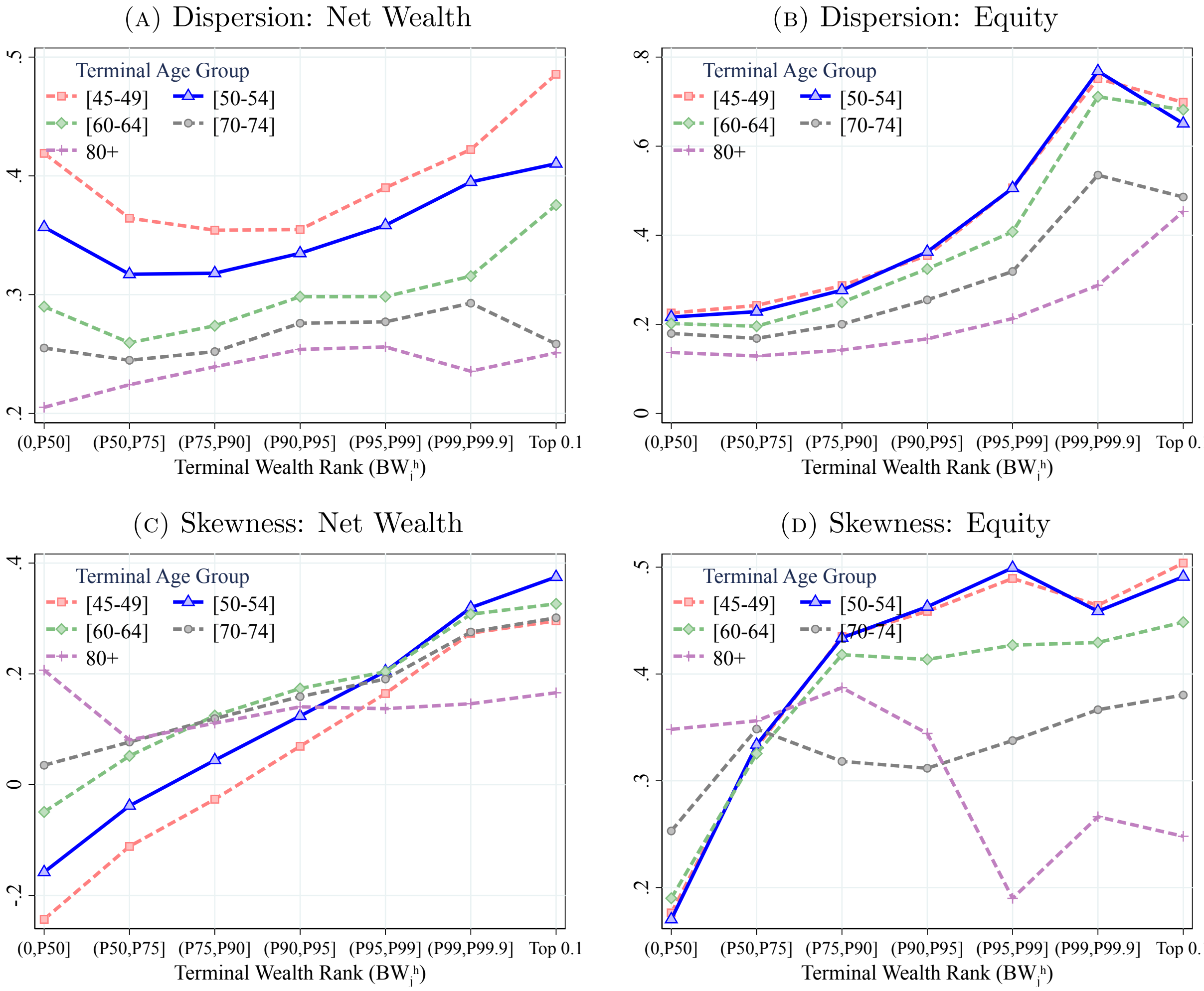
Notes: Figure SM.2.10 shows the evolution of the portfolio shares (left y-axis) and leverage (right y-axis).

Figure SM.2.11 – Returns on Assets Across the Wealth Distribution-Unweighted



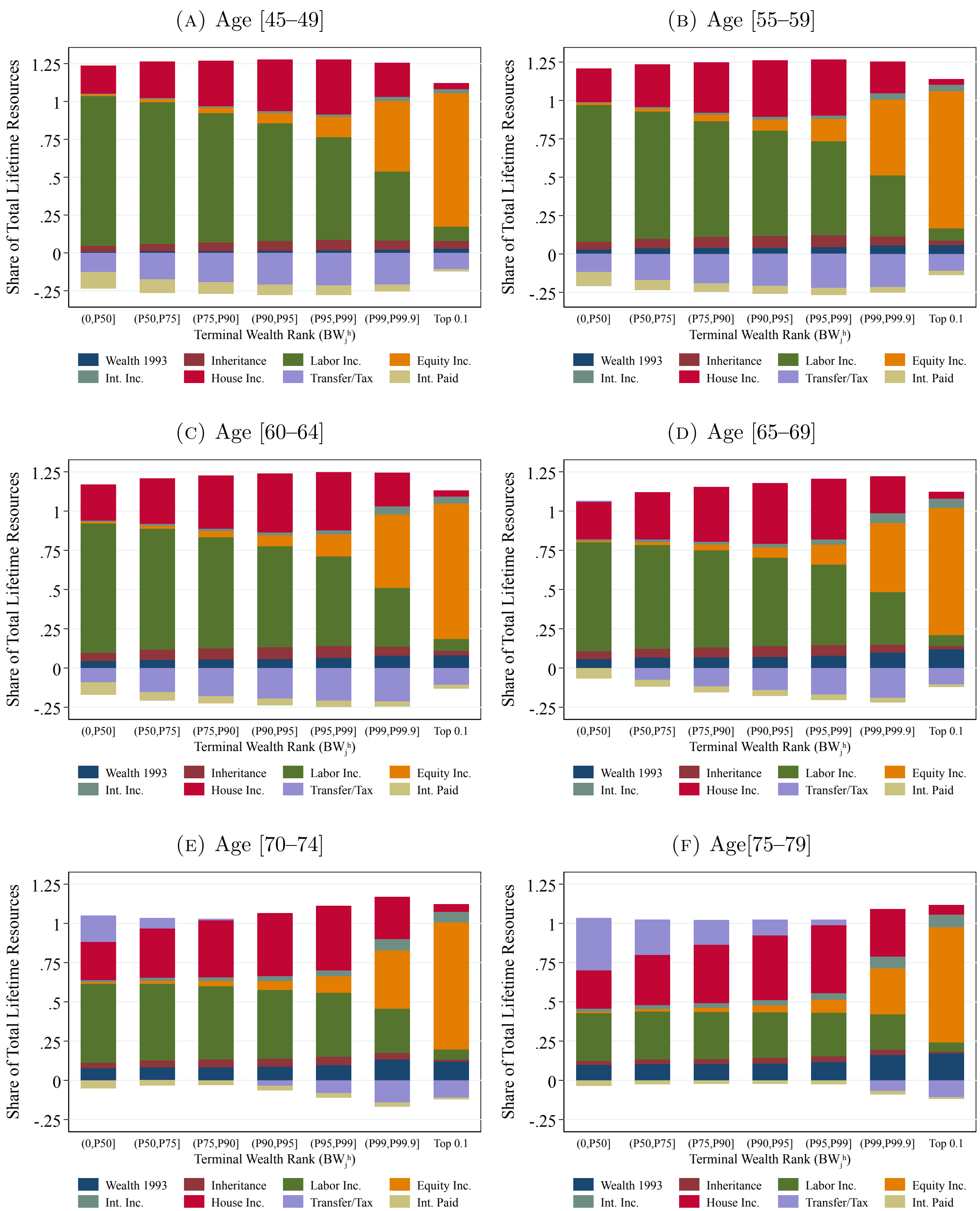
Notes: Figure SM.2.11 shows the 26-years mean of the value-weighted average gross annual returns within age and wealth groups across different conditioning years for different asset classes.

Figure SM.2.12 – Dispersion and Skewness of Rates of Log-term Returns-Unweighted



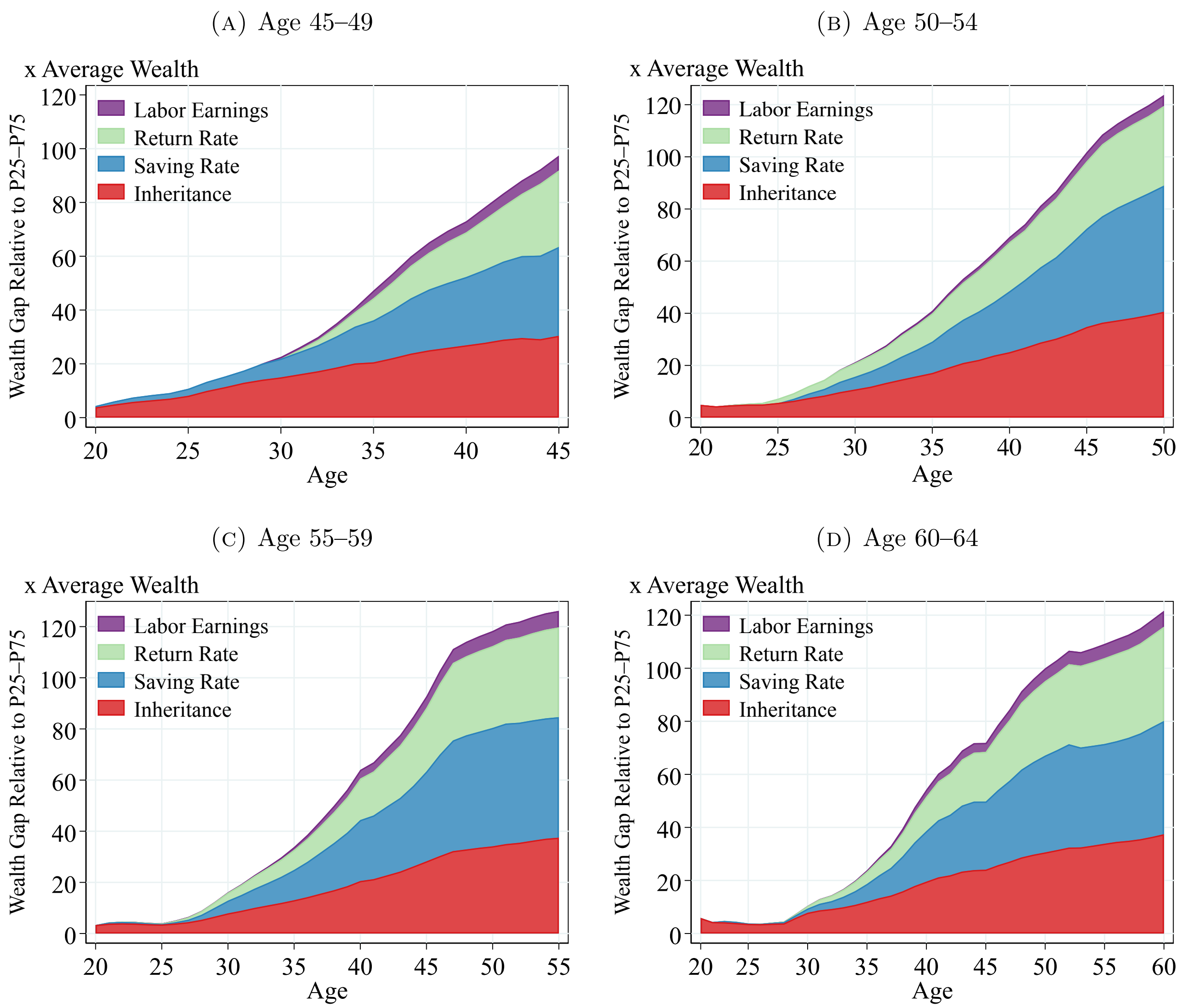
Notes: Figure SM.2.12 shows the 26-years mean of the value-weighted cross-sectional moments of the gross annual returns within age and wealth groups across different conditioning years for different asset classes.

Figure SM.2.13 – Backward Decomposition of Life Time Resources: Age Groups



Notes: Figure SM.2.13 shows the shares of lifetime income for a sample of households in a given conditioning year for different age groups conditional on BW_j^h. Lifetime resources refers to the sum of initial wealth (net worth in 1993) and all income sources between 1994 and the conditioning year. We average these shares across conditioning years.

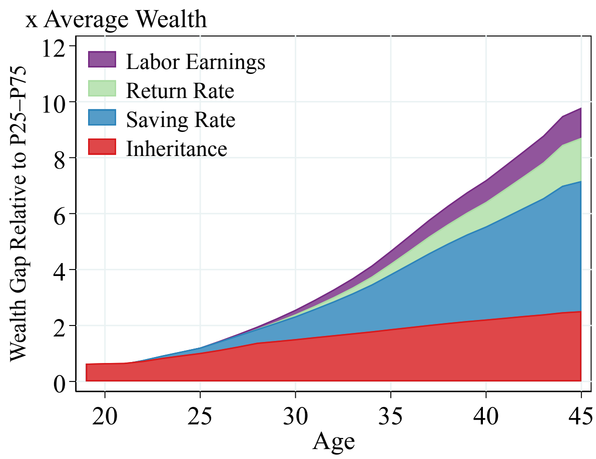
FIGURE SM.2.14 – DETERMINANTS OF TOP 0.1% WEALTH: AGE GROUPS



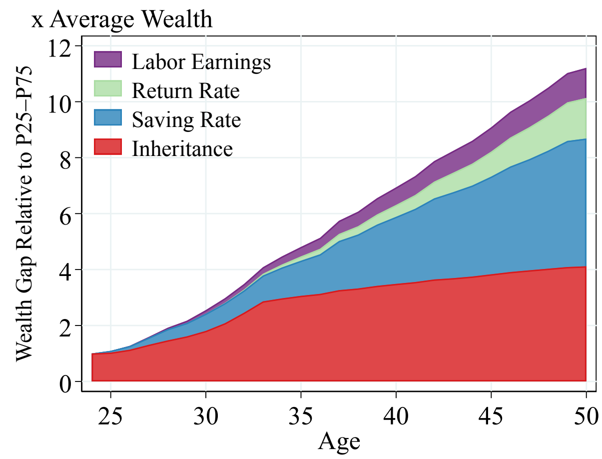
Notes: Figure SM.2.14 shows the S-O decomposition for households in $BW^h_{\geq P99.9}$ for different age groups.

Figure SM.2.15 – Determinants of Wealth Accumulation [50–54]

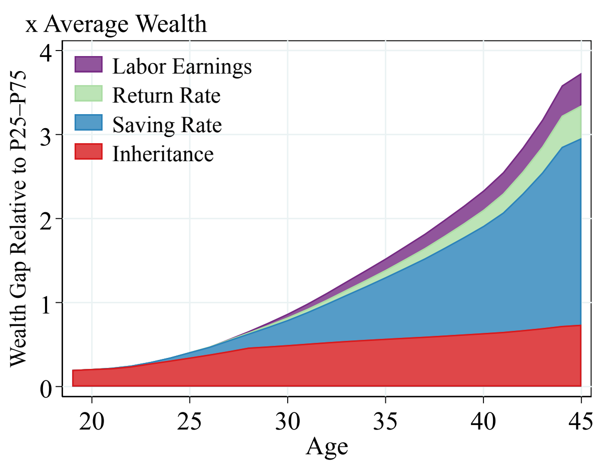
(a) $BW^{[45-49]}_{[99,99.9\%)}$



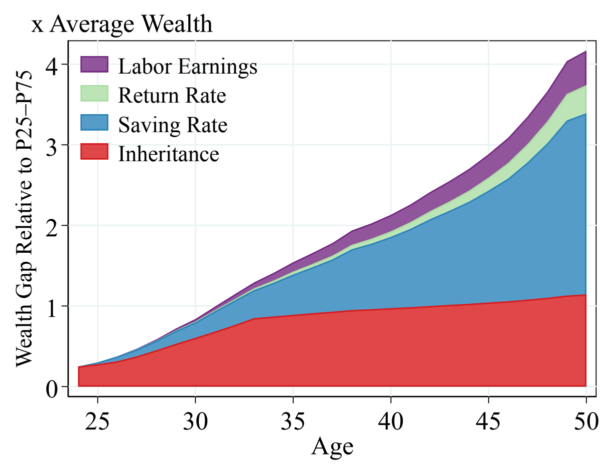
(b) $BW^{[50-54]}_{[99,99.9\%)}$



(c) $BW^{[45-49]}_{[95,99\%)}$

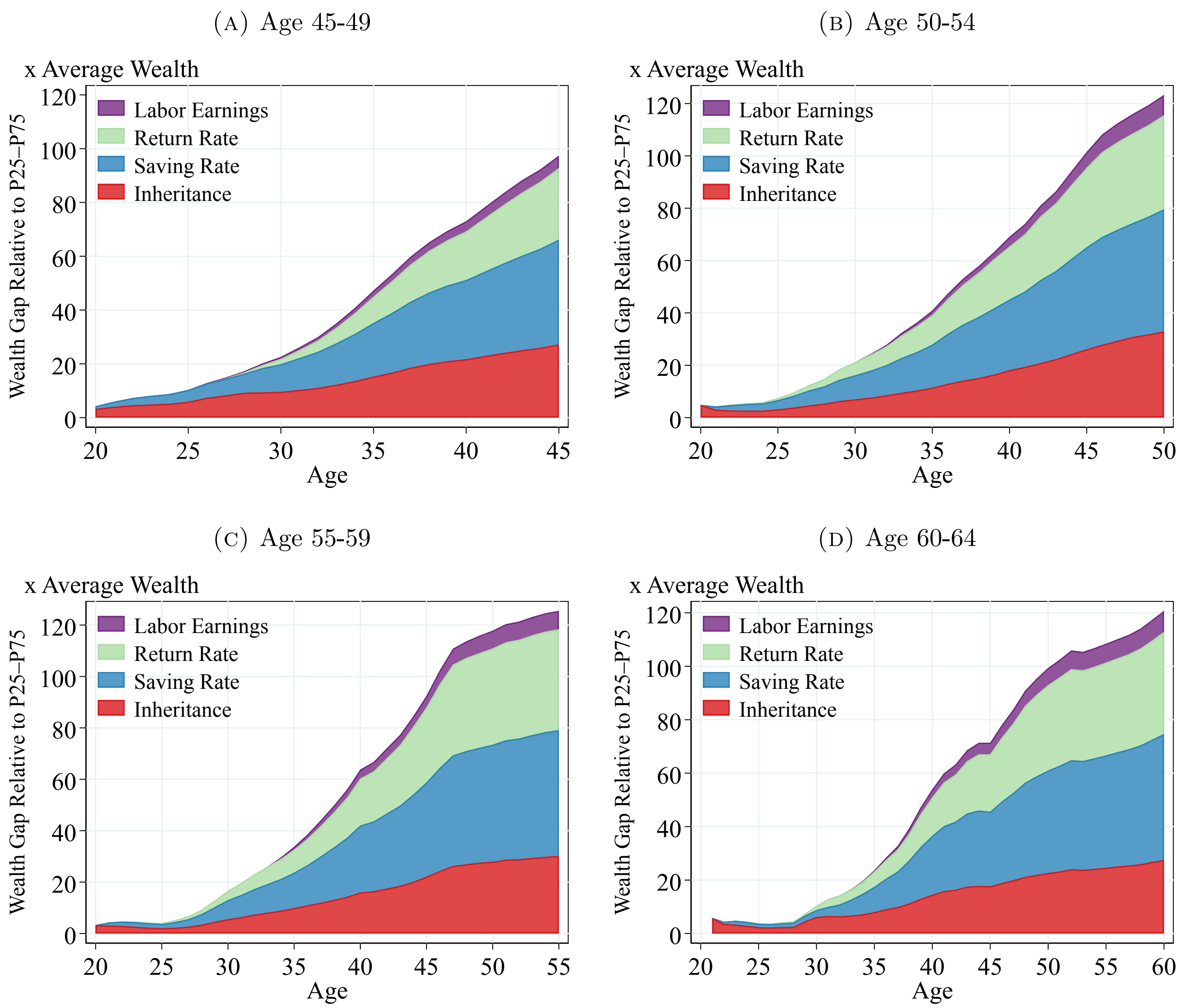


(d) $BW^{[50-54]}_{[95,99\%)}$



Notes: Figure SM.2.15 shows the S-O decomposition for households in $BW^{h}_{[P99,P99.9)}$ for different age groups.

Figure SM.2.16 – S-O Decomposition: Cash-on-Hand Saving Rate



Notes: Figure SM.2.16 shows the S-O decomposition for households in $BW^{j}_{\geq P99.9}$ for different age groups. The saving rate is defined as $\tilde{S}_{it} = W_{i,t} / \left(W_{i,t-1} + \tilde{L}_{i,t} + \tilde{H}_{i,t} + \tilde{R}_{i,t} W_{i,t-1} \right)$.

Figure SM.2.17 – Intergenerational Transition Matrix: Age Groups

(a) Age [45–49]

Parents Life Time Wealth Rank (columns); End-of-Period Wealh Rank (BW^h_j) (rows)

End-of-Period Wealh Rank (BW^h_j)	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	47.6	36.8	12.7	2.1	0.7	0.1	0.0
(50-75]	34.7	41.1	18.8	3.8	1.5	0.1	0.0
(75-90]	27.0	39.1	24.5	6.1	3.0	0.3	0.0
(90-95]	21.9	35.6	27.1	8.7	5.8	0.9	0.0
(95-99]	19.5	30.6	26.5	11.1	9.8	2.3	0.1
(99-99.9]	14.0	22.9	23.1	10.0	17.3	11.5	1.3
Top 0.1%	9.2	16.6	18.4	6.1	8.6	18.4	22.7

(b) Age [55–59]

Parents Life Time Wealth Rank (columns); End-of-Period Wealh Rank (BW^h_j) (rows)

End-of-Period Wealh Rank (BW^h_j)	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	47.2	37.6	12.5	2.0	0.7	0.0	0.0
(50-75]	33.9	41.7	18.7	4.0	1.6	0.1	0.0
(75-90]	26.4	40.1	23.7	6.3	3.2	0.3	0.0
(90-95]	21.8	35.0	27.0	9.2	6.4	0.6	0.0
(95-99]	16.3	32.0	27.2	12.2	10.3	2.0	0.1
(99-99.9]	14.0	28.1	21.3	10.9	15.5	8.6	1.6
Top 0.1%	7.0	22.8	15.8	11.4	13.9	17.7	11.4

(c) Age [60–64]

Parents Life Time Wealth Rank (columns); End-of-Period Wealh Rank (BW^h_j) (rows)

End-of-Period Wealh Rank (BW^h_j)	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	48.8	36.7	11.8	2.0	0.7	0.0	0.0
(50-75]	36.4	40.6	17.5	3.7	1.7	0.1	0.0
(75-90]	28.8	39.7	21.8	6.1	3.4	0.3	0.0
(90-95]	22.5	36.0	25.2	8.9	6.4	0.9	0.0
(95-99]	18.3	31.2	27.0	11.3	9.8	2.2	0.1
(99-99.9]	16.2	25.6	22.0	10.5	15.9	8.5	1.2
Top 0.1%	13.7	17.6	19.8	11.5	9.9	16.8	10.7

(d) Age [65–69]

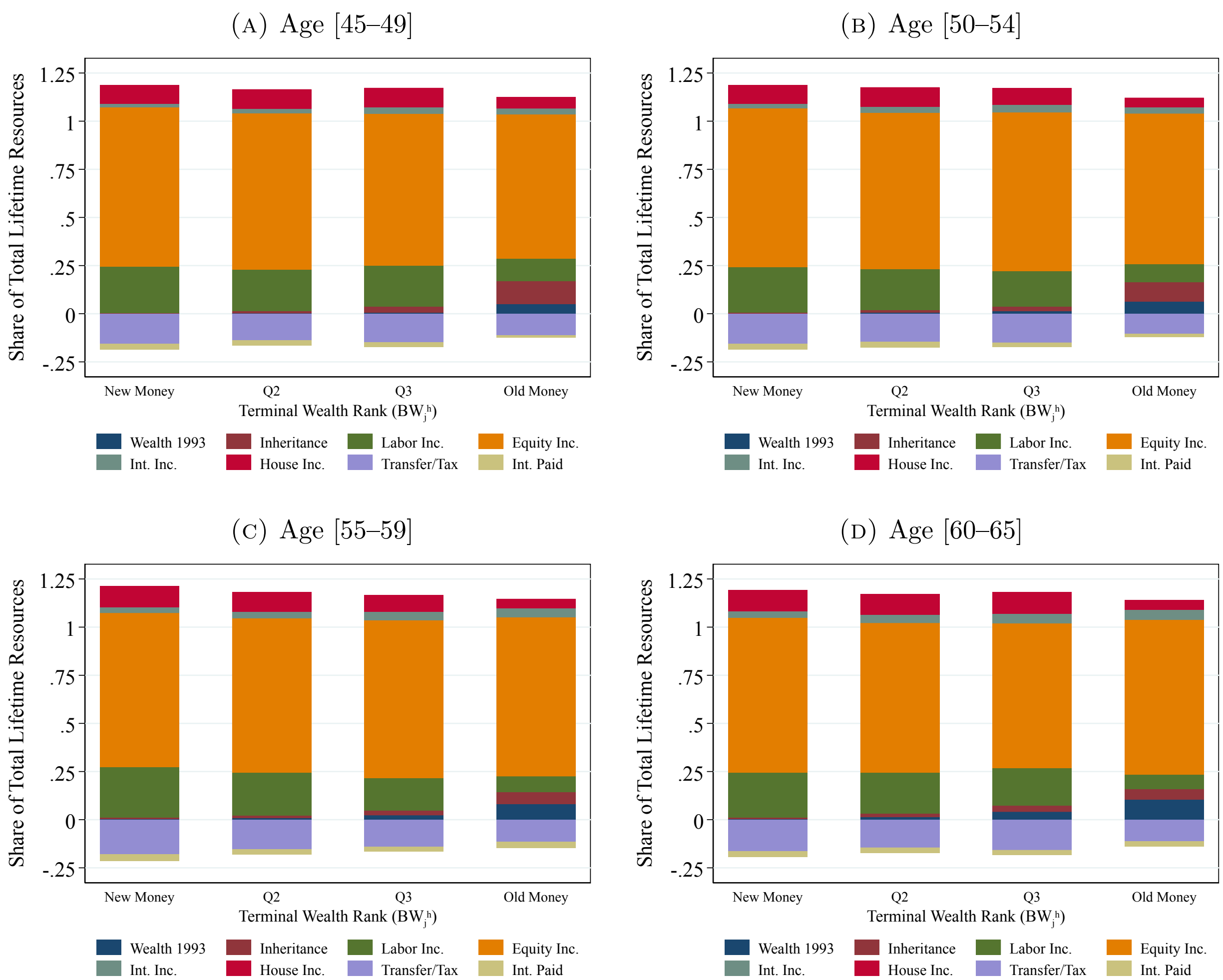
Parents Life Time Wealth Rank (columns); End-of-Period Wealh Rank (BW^h_j) (rows)

End-of-Period Wealh Rank (BW^h_j)	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	51.9	34.4	10.9	2.0	0.8	0.0	0.0
(50-75]	38.9	38.4	17.2	3.7	1.7	0.1	0.0
(75-90]	31.8	36.6	21.8	5.9	3.6	0.3	0.0
(90-95]	25.0	33.3	25.7	8.3	6.8	1.0	0.0
(95-99]	20.4	29.7	25.0	11.8	10.7	2.3	0.1
(99-99.9]	16.1	25.9	22.3	10.9	16.3	8.1	0.3
Top 0.1%	7.1	14.3	24.3	10.0	18.6	20.0	5.7

Notes: Figure SM.2.17 shows the intergenerational persistence of net wealth. It shows the results by first sorting household within age groups by the lifetime wealth (until 2015) of their parents. Each cell represent the fraction of household in different percentiles of the parents wealth distribution (columns), conditional on their percentile of the wealth distribution in the conditioning year (until 2015), BW_j^h (rows). Each row sums to 100. The Parents Life Time Wealth Rank is calculate as the rank of the average wealth adjusted for an age and year specific mean.

SM.2.3 New Money and Old Money: Additional Figures

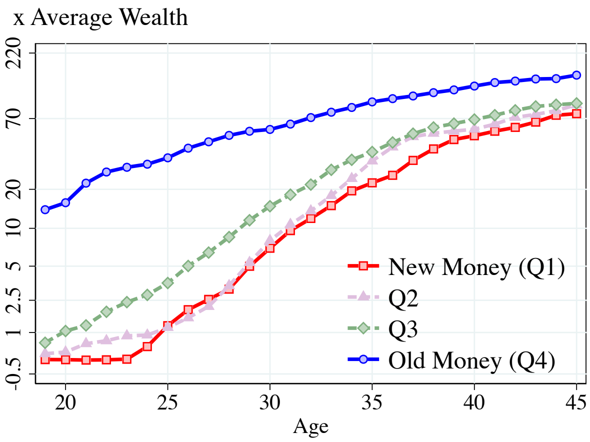
FIGURE SM.2.18 – INCOME SOURCES FOR NEW- AND OLD-MONEY HOUSEHOLDS



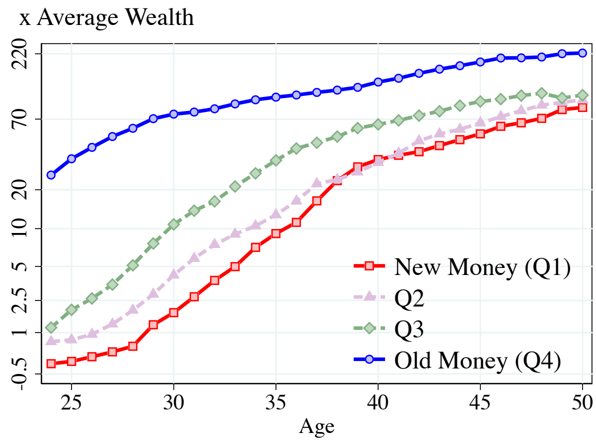
Notes: Figure SM.2.18 shows the shares of lifetime income for a sample of households in a given conditioning year for different age groups conditional on $BW^h_{\geq P99.9}$ and were in different quartiles of the initial average wealth distribution $(\overline{W}_{i,1994})$. Lifetime income refers to the sum of initial wealth (net worth in 1993) and all income sources between 1994 and the conditioning year. We average these shares across conditioning years.

FIGURE SM.2.19 – AVERAGE WEALTH PROFILE: OLD MONEY AND NEW MONEY

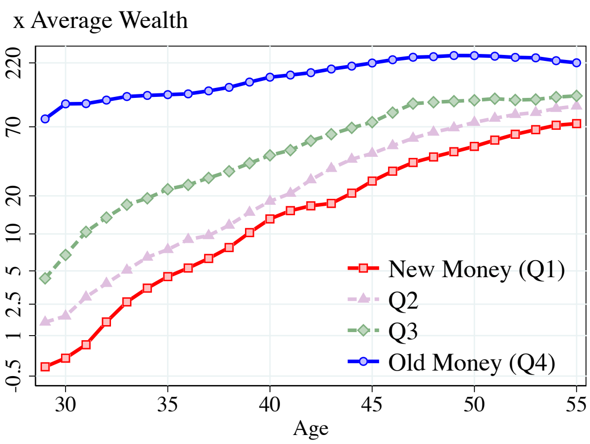
(A) Age [45–49]



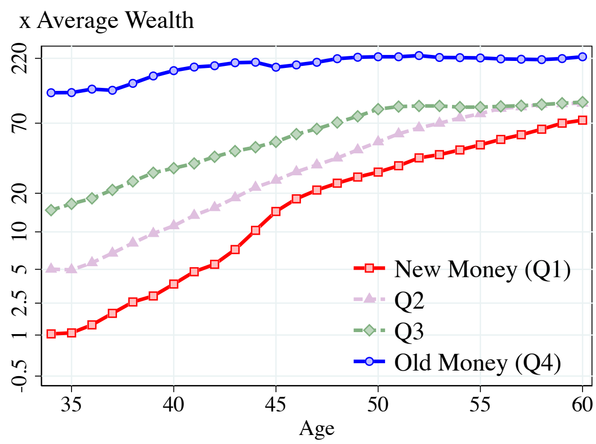
(B) Age [50–54]



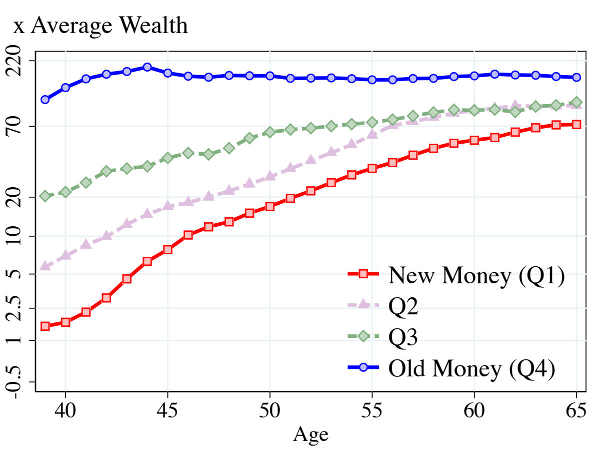
(C) Age [55–59]



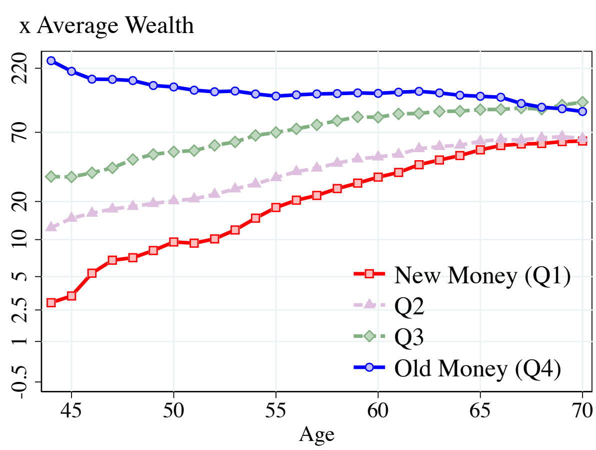
(D) Age [60–64]



(E) Age [65–69]



(F) Age [70–74]



Notes: Figure SM.2.19 shows the average wealth profile for household whose head is in different wealth age and belong to the top 0.1% of the wealth distribution at the end of the sample $\left(BW^{h}_{\geq P99.9}\right)$ and were in different quartiles of the initial average wealth distribution $(\overline{W}_{i,1994})$.

Figure SM.2.20 – Intergenerational Transition Matrix: Age Groups

(A) Age [45-49]

Parents Life Time Wealth Rank (columns); Children Wealth Rank in 1993 (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
Q1	16.3	24.8	28.9	9.5	12.9	5.3	2.4
Q2	16.9	24.6	20.6	9.5	17.2	8.7	2.5
Q3	12.8	23.6	23.3	11.3	13.0	12.5	3.5
Q4	7.6	15.4	17.0	8.1	23.0	23.2	5.7

(B) Age [50-54]

Parents Life Time Wealth Rank (columns); Children Wealth Rank in 1993 (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
Q1	20.8	22.8	27.2	9.6	12.7	5.7	1.1
Q2	17.2	24.5	28.3	8.4	14.6	5.5	1.5
Q3	12.1	28.0	23.6	10.4	15.7	8.4	1.8
Q4	6.4	15.7	18.7	11.2	21.6	20.3	6.2

(C) Age [55-59]

Parents Life Time Wealth Rank (columns); Children Wealth Rank in 1993 (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
Q1	18.4	37.9	19.2	12.4	7.4	3.7	1.1
Q2	15.7	30.6	27.5	9.2	12.6	3.9	0.5
Q3	11.3	24.7	19.2	13.6	17.8	10.8	2.6
Q4	7.8	17.0	17.2	8.6	23.5	19.6	6.3

(D) Age [60-64]

Parents Life Time Wealth Rank (columns); Children Wealth Rank in 1993 (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
Q1	21.7	31.6	24.0	8.9	11.2	2.3	0.3
Q2	19.6	26.0	23.2	14.1	11.3	4.8	1.0
Q3	12.9	26.5	22.6	10.6	17.4	8.7	1.3
Q4	9.7	15.2	17.4	8.7	21.3	21.6	6.1

(E) Age [65-69]

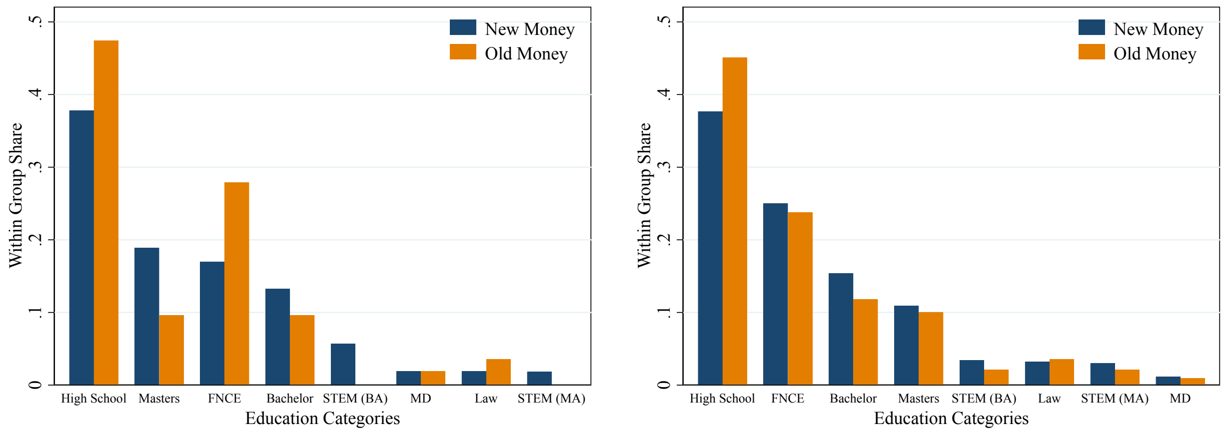
Parents Life Time Wealth Rank (columns); Children Wealth Rank in 1993 (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
Q1	22.6	30.2	23.1	7.5	12.6	3.5	0.5
Q2	14.8	26.7	24.3	12.4	16.7	5.2	0.0
Q3	13.0	22.3	24.9	11.9	18.7	8.8	0.5
Q4	10.9	20.2	17.6	11.4	18.1	19.7	2.1

Notes: Figure SM.2.20 shows a intergenerational transition matrix between households wealth in 2015 and their parental household wealth for households in different age groups. Each cell represent the fraction of household in different percentiles of the parents wealth distribution (columns), conditional on their percentile of the wealth distribution in the conditioning year, BW_j^h (rows). Each row sums to 100. The Parents Life Time Wealth Rank is calculate as the rank of the average wealth adjusted for an age and year specific mean.

Figure SM.2.21 – Education Shares for New and Old Money Households

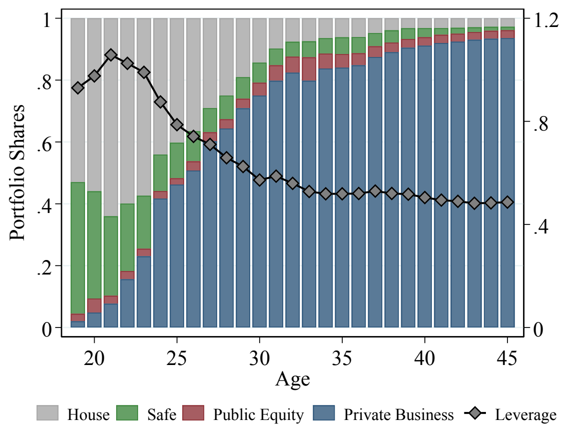
(a) Top 0.1% (b) Top 1%



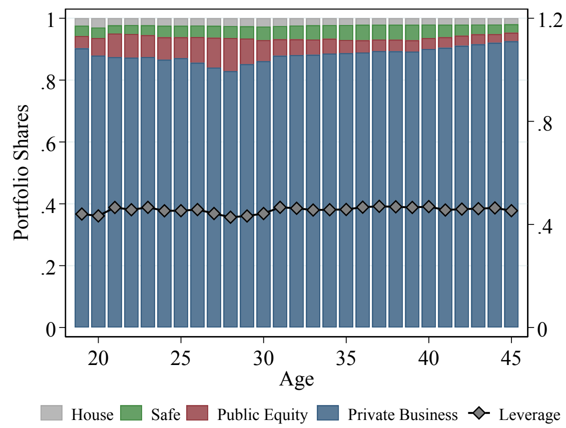
Notes: Figure SM.2.21 the share of different education groups households (highest degree of the head of the household) for households at the top 0.1% and top 1% among 50 to 54 year old households ($BW^{50-54}_{\geq P99.9}$and $BW^{50-54}_{\geq P99}$ respectively) divided in New Money (first quartile in the initial average wealth, $\overline{W}_{i,1994}$) and Old Money (forth quartile in the initial average wealth, $\overline{W}_{i,1994}$). HS is High-school or less, FNCE BA/MA is Bachelor or MBA on a finance or business administration major, BA and MA are other bachelor degrees or master degrees, MD is Medical Doctor or Dentist, H-STEM is BA or MA on a health related degree (except for Medical Doctor or Dentist) and STEM major. The data in this Figure covers the period 1993-2015.

Figure SM.2.22 – Portfolio Shares: Old Money and New Money and Age Groups

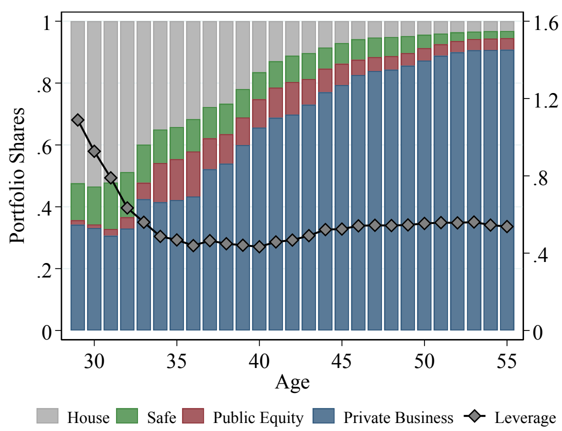
(A) Age [45–49]: New Money



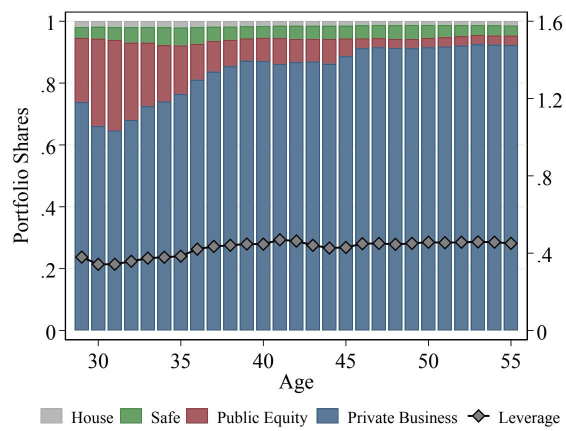
(B) Age [45–49]: Old Money



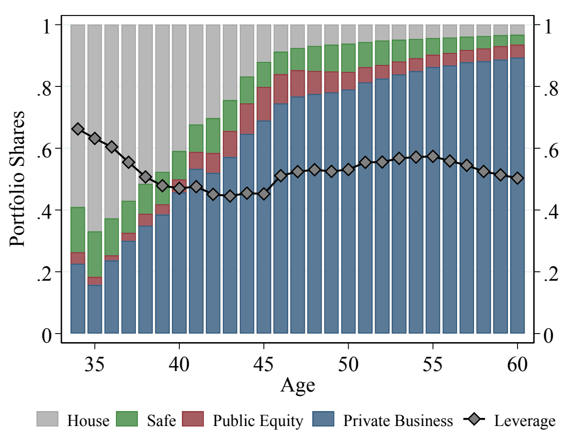
(C) Age [55–59]: New Money



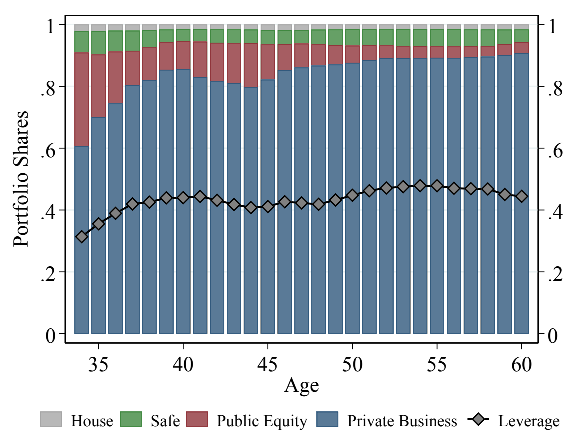
(D) Age 55–59: Old Money



(E) Age 60–64: New Money



(F) Age 60–64: Old Money

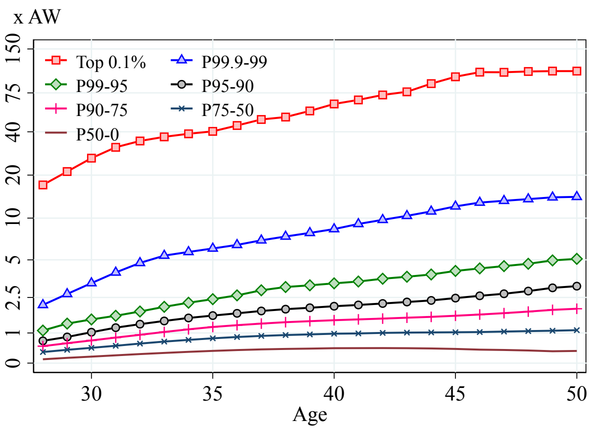


Notes: Figure SM.2.22 shows the portfolio composition and leverage for households that belong to the top 1% in a conditioning year τ for the New and Old Money households.

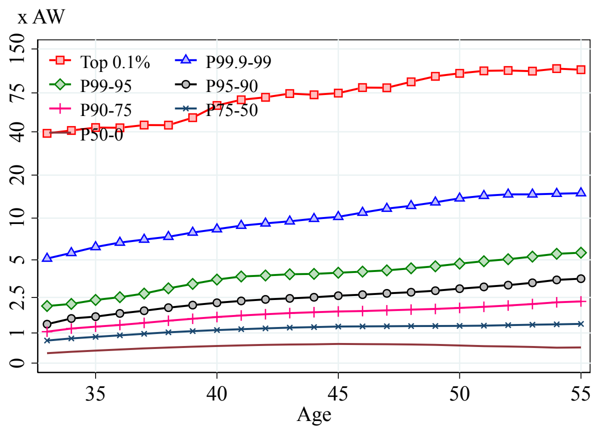
SM.2.4 Balanced Sample

Figure SM.2.23 – Balanced Backward-Looking Wealth Profiles

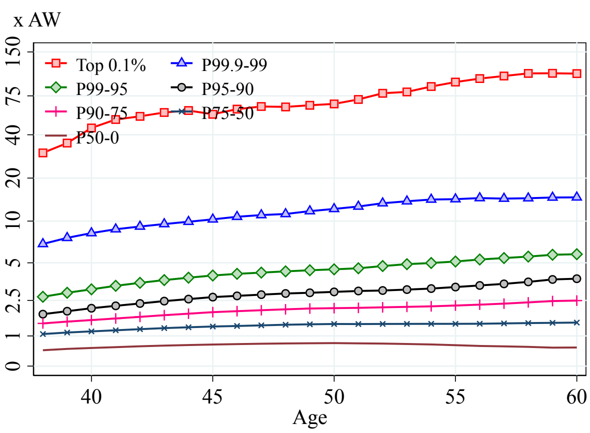
(a) Age [50–54]



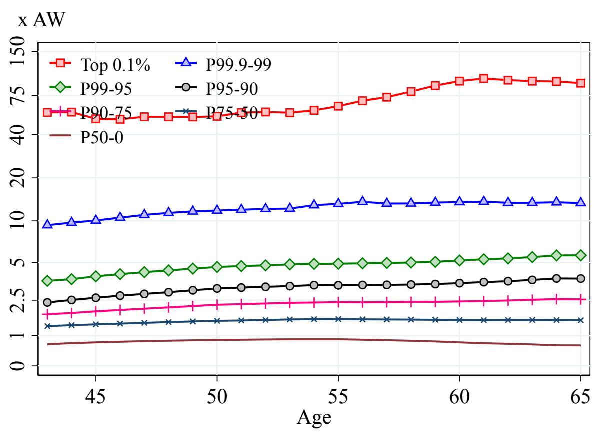
(b) Age [55–59]



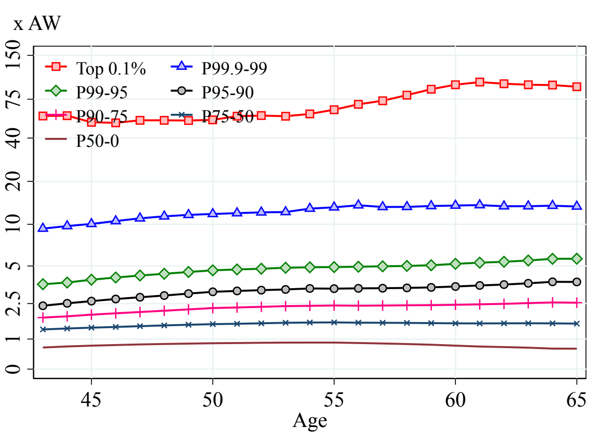
(c) Age [60–64]



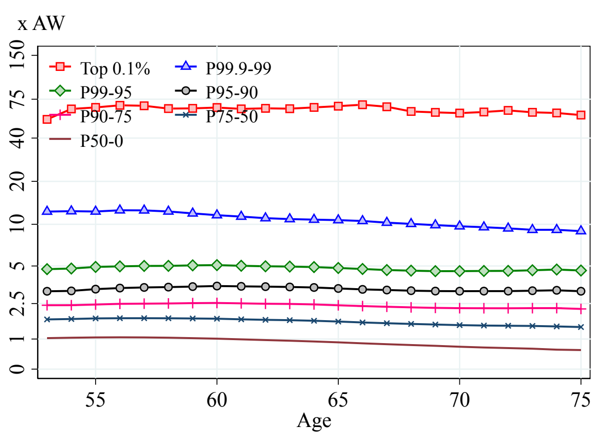
(d) Age [65–69]



(e) Age [70–74]



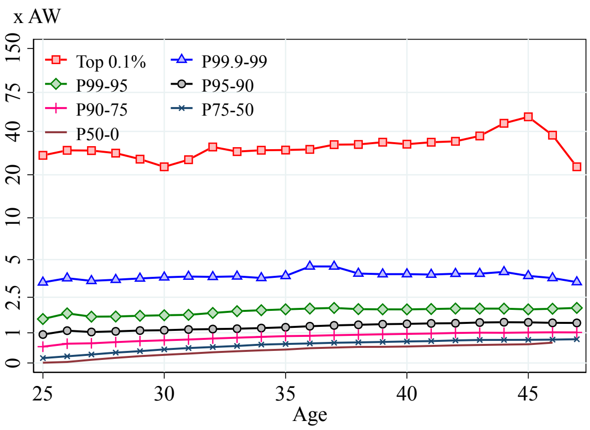
(f) Age [75–79]



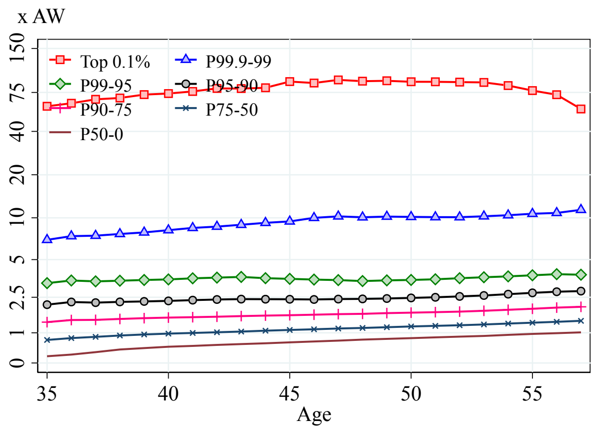
Notes: Figure SM.2.23 shows average wealth for different BW_j^h groups considering households that have been stable for at least ten years. The data in this Figure covers the period 1993-2015.

Figure SM.2.24 – Balanced Forward-Looking Wealth Profiles

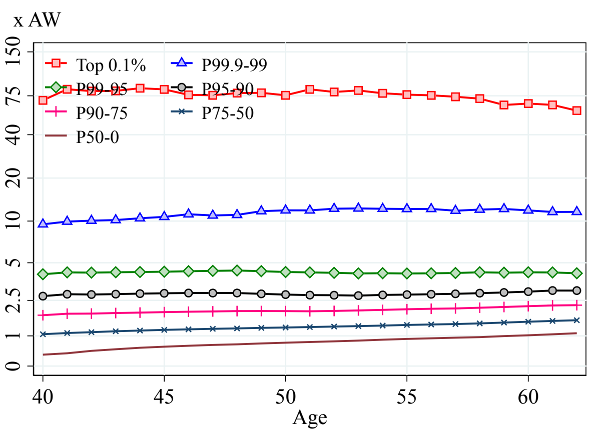
(a) Age [25–29]



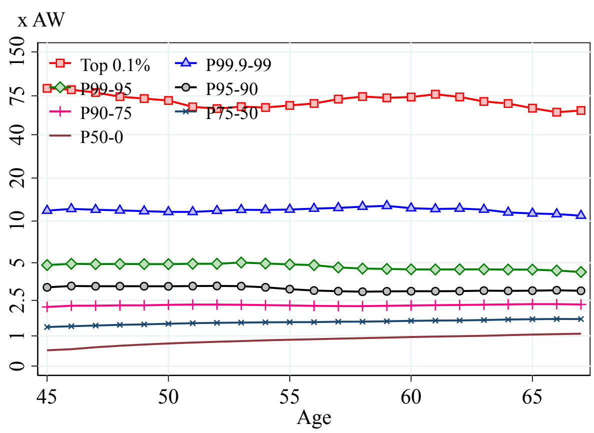
(b) Age [35–39]



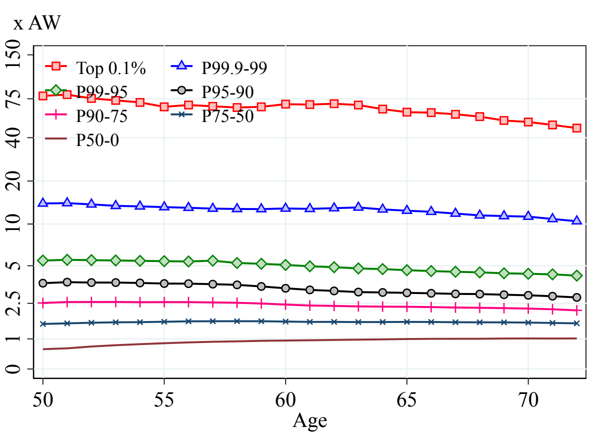
(c) Age [40–44]



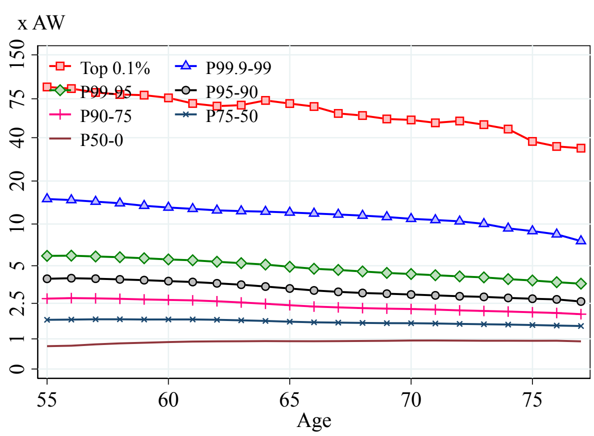
(d) Age [45–49]



(e) Age [50–54]



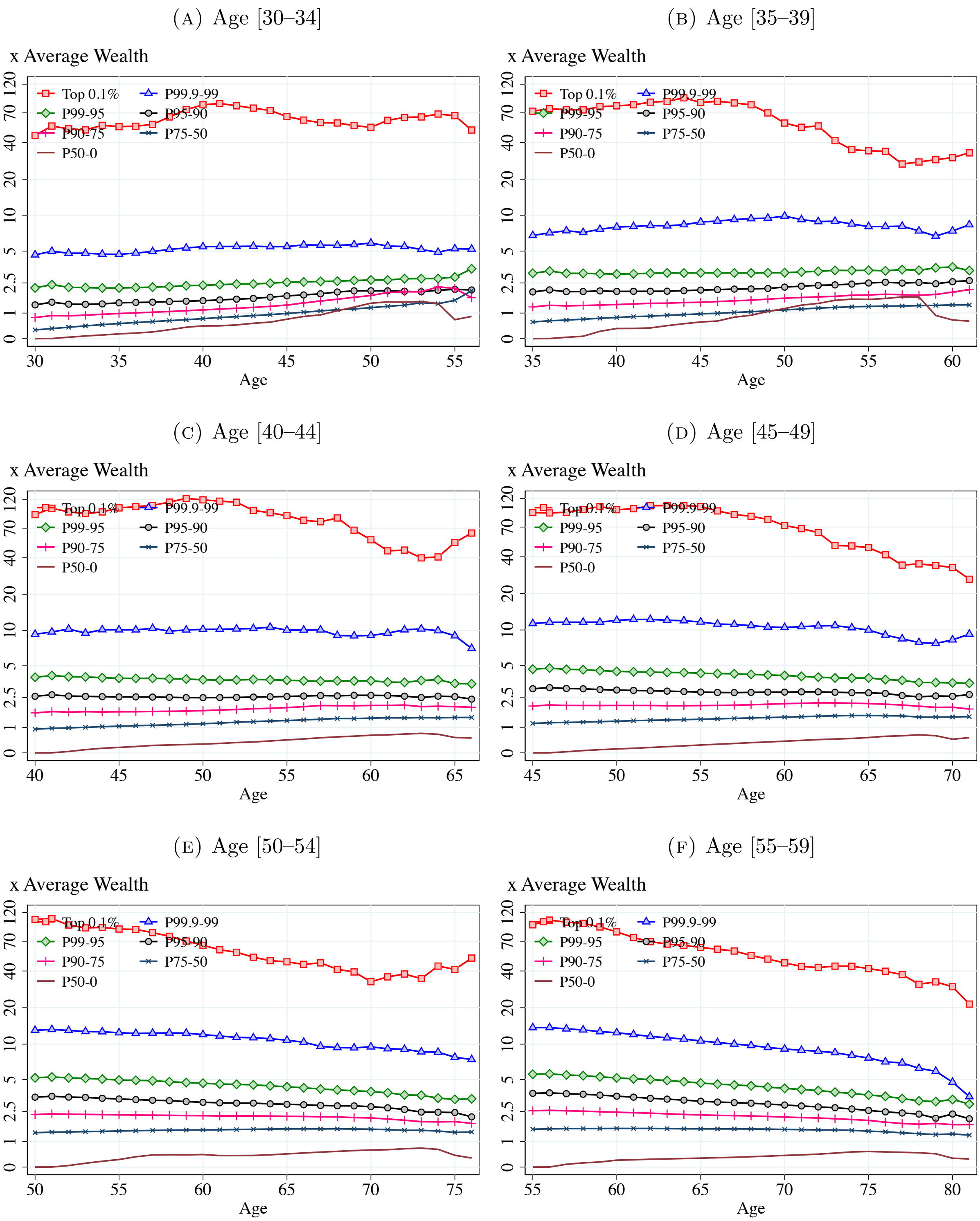
(f) Age [55–59]



Notes: Figure SM.2.24 shows average wealth for different FW_j^h groups considering households that have been stable for at least ten years. The data in this Figure covers the period 1993-2015.

SM.2.5 Forward-Looking Results

Figure SM.2.25 – Forward-Looking Wealth Profiles: Age Groups



(A) Age [30–34]

(B) Age [35–39]

(C) Age [40–44]

(D) Age [45–49]

(E) Age [50–54]

(F) Age [55–59]

Notes: Figure SM.2.25 shows the evolution of average household for households in different FW_j^h groups.

Figure SM.2.26 – Forward-Looking Transition Matrix: Age Groups

(a) Age [30–34]

Ending Average Wealth Rank (columns); Start-of-Period Wealh Rank, FW^h_j (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	59.2	21.9	12.1	3.7	2.7	0.5	0.0
(50-75]	50.9	26.8	14.3	4.3	3.2	0.5	0.0
(75-90]	37.5	31.1	19.3	6.3	4.9	0.9	0.1
(90-95]	29.1	28.9	23.7	9.1	7.5	1.5	0.1
(95-99]	25.1	23.8	23.8	11.6	12.0	3.4	0.3
(99-99.9]	14.8	15.2	20.0	13.3	21.9	13.1	1.7
Top 0.1%	2.6	3.5	4.3	4.6	14.8	43.3	26.9

(b) Age [35–39]

Ending Average Wealth Rank (columns); Start-of-Period Wealh Rank, FW^h_j (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	64.0	20.5	10.2	3.0	2.0	0.3	0.0
(50-75]	47.5	29.6	15.2	4.3	3.0	0.4	0.0
(75-90]	32.1	32.1	22.4	7.3	5.3	0.8	0.0
(90-95]	24.9	26.7	25.9	11.1	9.4	1.9	0.1
(95-99]	21.7	20.7	23.8	13.4	15.7	4.5	0.2
(99-99.9]	7.9	10.8	17.6	13.4	26.3	21.0	2.9
Top 0.1%	2.6	3.2	4.3	3.5	10.2	42.1	34.1

(c) Age [40–44]

Ending Average Wealth Rank (columns); Start-of-Period Wealh Rank, FW^h_j (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	66.7	19.9	9.1	2.6	1.5	0.2	0.0
(50-75]	45.6	31.4	15.8	4.2	2.7	0.3	0.0
(75-90]	29.4	32.1	24.3	7.9	5.6	0.7	0.0
(90-95]	23.3	25.0	26.6	12.3	10.9	1.8	0.1
(95-99]	18.2	18.8	24.5	14.7	18.4	5.2	0.3
(99-99.9]	4.8	8.1	16.0	12.6	28.7	26.3	3.5
Top 0.1%	2.0	1.4	6.2	2.9	11.5	39.9	36.0

(d) Age [45–49]

Ending Average Wealth Rank (columns); Start-of-Period Wealh Rank, FW^h_j (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	68.3	19.5	8.5	2.3	1.3	0.1	0.0
(50-75]	44.7	32.3	16.1	4.2	2.5	0.3	0.0
(75-90]	28.8	31.7	24.9	8.3	5.6	0.7	0.0
(90-95]	22.5	25.0	26.8	12.6	11.3	1.6	0.1
(95-99]	14.3	18.3	25.8	15.3	20.3	5.8	0.2
(99-99.9]	3.6	6.3	15.3	12.9	30.1	28.1	3.9
Top 0.1%	1.6	1.1	3.9	3.7	14.5	38.7	36.6

(e) Age [50-54]

Ending Average Wealth Rank (columns); Start-of-Period Wealh Rank, FW^h_j (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	69.3	19.1	8.1	2.1	1.2	0.1	0.0
(50-75]	44.7	32.7	16.1	4.0	2.2	0.2	0.0
(75-90]	29.0	31.8	24.8	8.3	5.5	0.6	0.0
(90-95]	21.3	25.5	27.5	12.9	11.3	1.6	0.0
(95-99]	12.2	17.4	26.1	16.2	22.1	5.7	0.2
(99-99.9]	3.4	6.4	14.4	12.3	30.9	28.7	3.9
Top 0.1%	2.4	1.1	3.6	6.1	10.5	36.7	39.5

(f) Age [55-59]

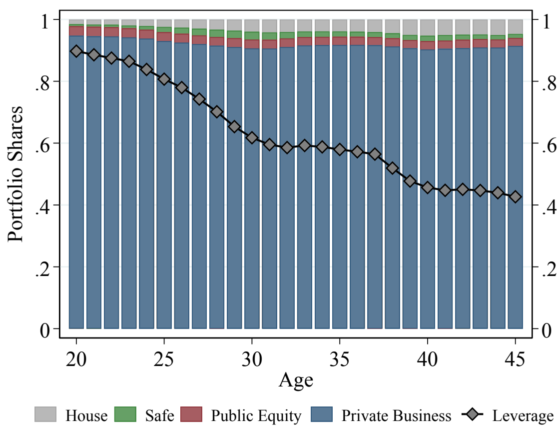
Ending Average Wealth Rank (columns); Start-of-Period Wealh Rank, FW^h_j (rows)

	[0,50]	(50-75]	(75-90]	(90-95]	(95-99]	(99-99.9]	Top 0.1%
[0,50]	70.7	18.7	7.6	1.8	1.0	0.1	0.0
(50-75]	44.6	32.7	16.5	4.0	2.1	0.2	0.0
(75-90]	29.5	32.0	24.5	8.3	5.1	0.6	0.0
(90-95]	20.8	25.5	27.5	13.1	11.7	1.3	0.0
(95-99]	11.5	17.6	26.1	16.6	22.6	5.4	0.2
(99-99.9]	3.8	6.5	14.6	12.7	30.8	28.1	3.6
Top 0.1%	1.8	0.5	2.7	5.5	12.0	38.5	38.9

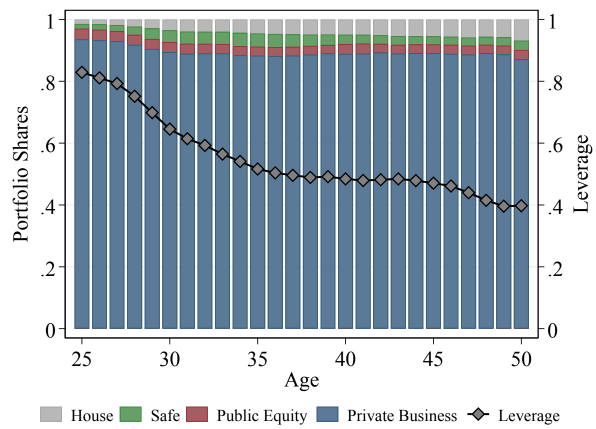
Notes: Figure SM.2.26 shows the intragenerational persistence of net wealth. Figure SM.2.26 shows the results by first sorting household whose head is in different age groups in the conditioning year and then again by $\overline{W}_{i,2019}$. Each cell represent the fraction of household in different percentiles of the wealth distribution in $\overline{W}_{i,2019}$ (columns), conditional on their percentile of the wealth distribution in the conditioning year, FW^h_j (rows).

Figure SM.2.27 – Forward-Looking Portfolio Shares for Top 0.1%: Age Groups

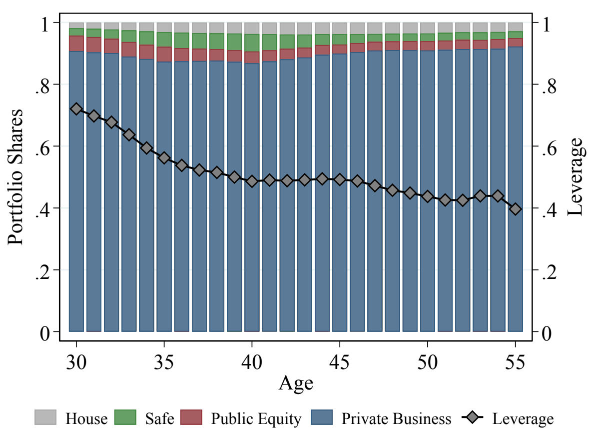
(a) Age [20-25]



(b) Age [25-29]

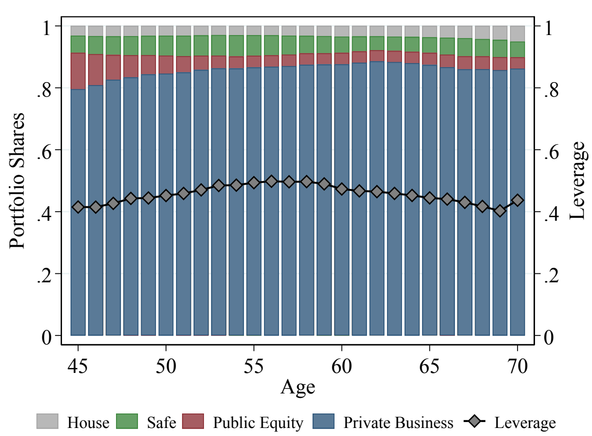


(c) Age [30-34]

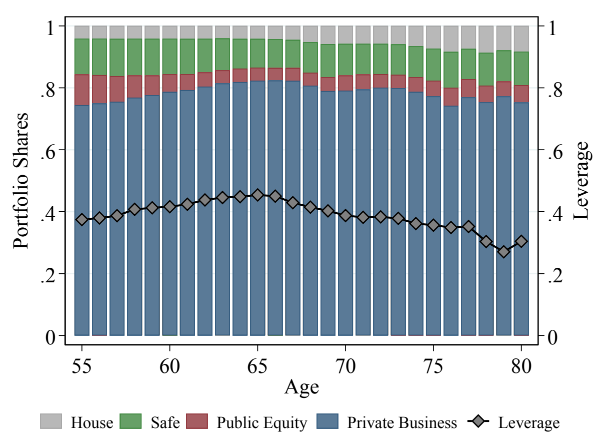


(d) Age 35-39

(e) Age [45-49]

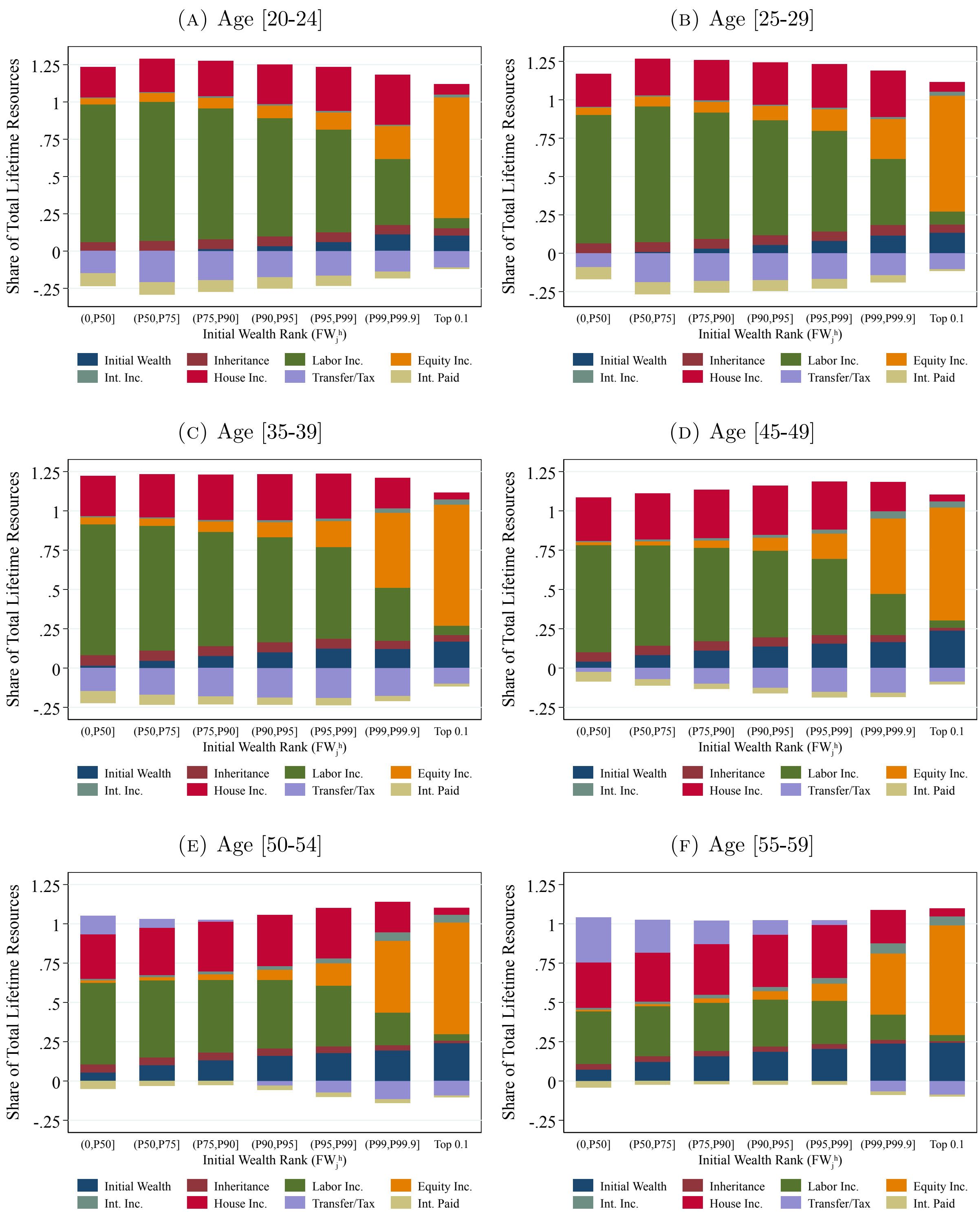


(f) Age [55-59]



Notes: Figure SM.2.27 shows the evolution of the portfolio shares (left y-axis) and leverage (right y-axis) for households.

Figure SM.2.28 – Lifetime Resources Decomposition: Age Groups



Notes: Figure SM.2.28 shows lifetime resources shares for households in different age groups and wealth rank, FW_j^h. Lifetime resources refers to the sum of initial wealth (net worth in conditioning year) and all income sources between the conditioning year and 2019. We average these shares across conditioning years.

SM.3 Methodology for Imputing Missing Data

While our administrative wealth data is very rich and spans a long period of time 1993 to 2019, the data on returns on equity is available only after 2005. To address this limitation, we employ machine learning techniques to predict these values for the years they are missing. This appendix describes the methodology used to impute missing data for Norwegian households.

SM.3.1 Model Selection: Gradient Boosting Regression

We employ **Gradient Boosting Regression** (Friedman (2001)) from Pedregosa *et al.* (2011) library as our machine learning model of choice. This technique constructs an ensemble of decision trees sequentially, with each new tree specifically targeting and correcting the residual errors from previous iterations. This adaptive approach is particularly valuable for our microdata analysis, where complex, hierarchical relationships often exist beneath the surface.

Before selecting Gradient Boosting Regression as our preferred method, we evaluated several alternative machine learning approaches for imputing financial data. These alternatives included Random Forest Regression, which builds multiple decision trees independently and averages their predictions; AdaBoost, an earlier boosting algorithm that uses a different weighting scheme; standard Decision Tree Regression; and parametric approaches such as Lasso and Ridge regression with polynomial features to capture non-linearities.

Gradient Boosting Regression consistently outperformed these alternatives in our context. A primary advantage of gradient boosting is its capacity to capture intricate non-linear relationships without explicit feature engineering. In our context, for example, we hypothesize that wealth and age significantly influence returns, but the precise functional form of these relationships remains uncertain and likely highly non-linear. Gradient boosting excels at discovering these complex patterns autonomously, enabling us to model relationships that would be difficult to specify parametrically. Additionally, the method provides valuable insights through feature importance metrics, illuminating which variables most strongly predict inheritance amounts or returns on equity.

The technique demonstrates remarkable robustness to outliers—a critical advantage given the heavily skewed distributions of wealth, inheritance, and investment returns in our population. This robustness stems from the algorithm's approach of predicting target variables by averaging values within decision tree leaves, which naturally moderates the influence of extreme observations.

SM.3.2 Hyperparameter Optimization and Validation Framework

We follow a rigorous statistical learning framework designed to ensure both optimal performance and reliable generalization to out-of-sample data. We implement a comprehensive two-stage validation strategy:

1. **Initial Train-Test Partition.** We begin by partitioning our full dataset into strictly separated training (80%) and testing (20%) samples using a fixed random seed to ensure reproducibility. This initial partition is established before any model development

begins, guaranteeing that the test set remains completely untouched during all stages of model optimization. The test set serves as our final evaluation benchmark, providing an unbiased assessment of model performance on data not used in any aspect of model development.

2. **Hyperparameter Optimization via Randomized Search.** Within the training set, we implement an efficient hyperparameter optimization procedure using RandomizedSearchCV from scikit-learn. Rather than exhaustively searching all possible hyperparameter combinations (which would be computationally prohibitive), we use a randomized search that samples from parameter distributions.
During the optimization process, each candidate parameter configuration undergoes k-fold cross-validation within the training set. This creates multiple training-validation splits, ensuring that parameters generalize well across different subsets of our data rather than overfitting to idiosyncratic patterns in any single subset.
The hyperparameter space we explore includes:

   - **learning_rate** is drawn from a uniform distribution between 0.05 to 0.15. This parameter controls how much each new tree contributes to the ensemble. Lower values require more trees but often lead to better generalization.
   - **max_depth** is drawn from random integers between 3 and 15. It determines the maximum depth of each decision tree Deeper trees can capture more complex patterns but risk overfitting.
   - **n_estimators** is a uniform integer variable between 100 and 200. It defines the number of sequential trees to build. More trees generally improve performance until diminishing returns set in.

   By defining continuous distributions rather than discrete points for parameters like learning rate (e.g., uniformly distributed between 0.05 and 0.15), we can discover optimal settings that might be missed by coarser grid-based approaches. We also utilize parallel processing (n_jobs=5) to speed up the tuning process, allowing for more thorough exploration of the parameter space.

SM.3.3 Training Model Using Post-2005 Data

We obtain both dividends and retained earnings for equity using the link between individuals' equity ownership and companies' balance sheets given by the shareholder registry, which is available only from 2005 onward. Below, we explain how we prepare the data to be used in our imputation. We also discuss our model's performance in fitting the data.

SM.3.3.1 Data Preparation

We define total wealth as the sum of private equity, public equity, housing, and safe assets minus liabilities. Equity holdings are calculated as the sum of private and public equity components. Following established practice in the literature (Fagereng *et al.* (2020)), we exclude individuals with financial wealth below NOK 9,000 (approximately USD 1,000) to ensure reliable return calculations and avoid extreme values caused by small denominators. Furthermore, to mitigate the influence of extreme return values, we winsorize the equity return variable at the 0.5th and 99.5th percentiles, replacing extreme values with these percentile thresholds. We compute equity returns by summing capital gains on private equity, capital gains on stocks, and dividend income, then dividing by the average equity holdings between the current and previous year (à la Dietz (1968)).

We normalize wealth and equity variables by the average wealth and average equity holdings in each year, while income variables are normalized by average earnings, ensuring comparability across time. The full list of predictor variables includes:

- **Individual Characteristics.**

1. Individual's age
2. Gender indicator (1 for male, 0 for female)
3. Labor income, normalized by average earnings
4. Self-employment income, normalized by average earnings
5. Sum of inheritance and inter-vivos transfers received, normalized by average wealth

- **Wealth Measures.**

1. Total wealth, normalized by average wealth
2. Arc-percent change in wealth from previous year
3. Individual's percentile position in the wealth distribution within year
4. Total equity holdings (private + public), normalized by average equity holdings
5. Arc-percent change in equity holdings from previous year
6. Average of current and previous year's equity holdings, normalized
7. Individual's percentile position in the equity distribution within year

- **Portfolio Composition.**

1. Share of total assets held in equity
2. Arc-percent change in equity share from previous year
3. Share of total assets held in housing
4. Arc-percent change in housing share from previous year
5. Share of total assets held in safe assets (e.g., deposits, bonds)
6. Arc-percent change in safe assets share from previous year

7. Share of equity holdings invested in publicly traded securities
8. Arc-percent change in public equity share from previous year
9. Leverage ratio (liabilities divided by total assets)

- **Returns on Other Asset Classes.**

1. Return on housing investments
2. Arc-percent change in housing returns
3. Return on safe assets
4. Arc-percent change in safe asset returns

- **Macroeconomic Indicators.**

1. Average wealth in the population for the given year
2. Arc-percent change in average wealth from previous year
3. Average equity holdings in the population for the given year
4. Arc-percent change in average equity holdings from previous year
5. Average earnings in the population for the given year
6. Arc-percent change in average earnings from previous year

This rich feature set enables the model to capture diverse factors influencing equity returns, including individual financial circumstances, portfolio decisions, position in the wealth distribution, and broader economic conditions. The inclusion of arc-percent changes for many variables allows the model to incorporate dynamic financial behaviors and trends over time.

SM.3.4 Model Performance and Results

We now discuss how good our imputation of returns on equity is. The explanatory power of our model is relatively high with an overall R-squared of about 0.68 for the training data, 0.46 for the out-of-sample test, and 0.64 for the entire data (SM.3.1).

Figures SM.3.1 and SM.3.2 show the impressive performance of our gradient boosting model in capturing the heterogeneity of equity returns across both the wealth and equity distributions. The predicted returns (red lines) track closely the actual returns (blue lines) across training, test, and full samples demonstrates the model's ability to generalize effectively.

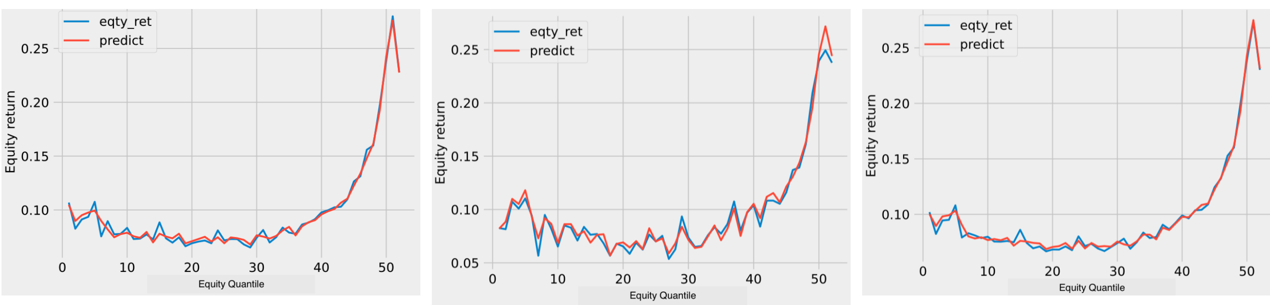
Figure SM.3.1 reveals several key patterns in equity returns across the wealth distribution. The model successfully captures the subtle U-shaped pattern in the bottom quartile of the wealth distribution, where both the poorest households (0-5th percentiles) and those around

Table SM.3.1 – R^2 for Imputing Inheritances

	(1) Train Sample	(2) Test Sample	(3) Entire Data
R^2	0.68	0.46	0.64
Obs	991,478	247,869	1,239,347

Figure SM.3.1 – Average returns over the Wealth Distribution

(A) Training sample (B) Test Sample (C) Whole sample



Notes: Panel A shows the average equity returns across the wealth distribution. Panels B and C show the same figure for test and entire samples, respectively.

the 25th percentile show slightly elevated returns compared to households in the 10-15th percentiles. From approximately the 15th to 40th percentiles, equity returns remain relatively stable around 6%, reflecting more homogeneous investment strategies among middle-wealth households. The model tracks this plateau with high precision. Beginning around the 40th percentile, we observe a steady increase in equity returns that accelerates for the wealthiest households. The model accurately captures this important non-linear relationship, showing how returns rise from approximately 6% in the middle of the distribution to over 14% at the very top.

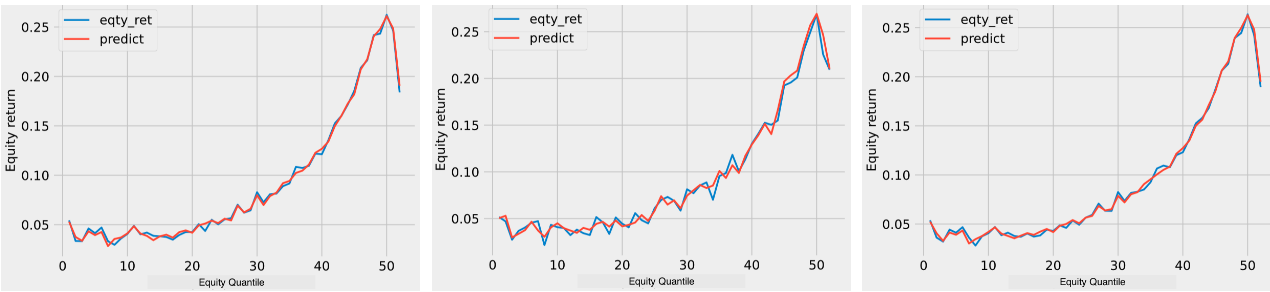
The striking agreement between actual and predicted returns across all three panels confirms that our model successfully replicates the wealth gradient in investment performance documented in previous literature. This gradient—where wealthier households achieve systematically higher returns—represents a crucial mechanism in wealth inequality dynamics that our imputation method preserves.

Figure SM.3.2 offers a complementary perspective by plotting returns against the equity distribution itself (among equity holders). A clear monotonic relationship between the amount of equity held and the returns achieved. Households with larger equity positions (higher percentiles) systematically outperform those with smaller positions, with returns rising from approximately 4% at the bottom to 14% at the top of the distribution. The model captures the subtle downturn in returns at the very top of the equity distribution (above the 45th percentile), potentially reflecting different investment constraints or strategies among those with the largest equity holdings.

Again, the remarkable similarity in fit quality between training and test samples indicates that the model is not overfitting to the training data but has learned generalizable patterns

Figure SM.3.2 – Average Returns over the Equity Distribution

(a) Training sample (b) Test Sample (c) Whole sample



Notes: Panel A shows the average equity returns across the equity distribution (among those that hold equities). Panels B and C show the same figure for test and entire samples, respectively.

in equity returns. The alignment between actual and predicted values is particularly notable given the inherent volatility and idiosyncratic nature of investment returns. By accurately modeling returns across both wealth and equity dimensions, our imputation approach preserves the crucial relationship between household financial position and investment performance—a key driver of wealth dynamics that would be lost if simplistic imputation methods (such as assigning average returns to all households) were employed.

These results provide strong validation for our machine learning approach to equity returns imputation, demonstrating that the model successfully captures both the level of returns and their complex variation across different segments of the population. The maintained heterogeneity in returns is essential for our empirical investigation of lifecycle wealth accumulation.

References

Dietz, P. O. (1968). Components of a measurement model: rate of return, risk, and timing. *The Journal of Finance*, **23** (2), 267–275. SM.3.3.1

Fagereng, A., Guiso, L., Malacrino, D. and Pistaferri, L. (2020). Heterogeneity and persistence in returns to wealth. *Econometrica*, **88** (1), 115–170. SM.3.3.1

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232. SM.3.1, SM.4

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, **12**, 2825–2830. SM.3.1

SM.4 2-Step Imputation Method for Inter vivos Gifts

After 2013, inter vivos gifts are no longer recorded in the inheritance and gift registry. We therefore construct inter vivos transfers for 2014–2019 using two sources of information. First, for households with observed private-equity transfers in the shareholder registry, we treat these transactions as inter vivos gifts and use their recorded values directly. Second, for remaining households, we impute inter vivos transfers using a two-stage estimation procedure using pre-2014 data.

In the first stage, we estimate the probability of receiving a gift as a function of recipient age, recipient wealth group in period t–1, parental wealth group, and immigrant status. In the second stage, we use machine learning, specifically gradient boosting regression (Friedman (2001)) from **?** library, to predict the conditional value of a transfer based on recipient and parental characteristics. The final imputed inter vivos gift is the product of the predicted probability of receiving a gift and the predicted conditional value.

In this appendix we explain the details of our imputation method.

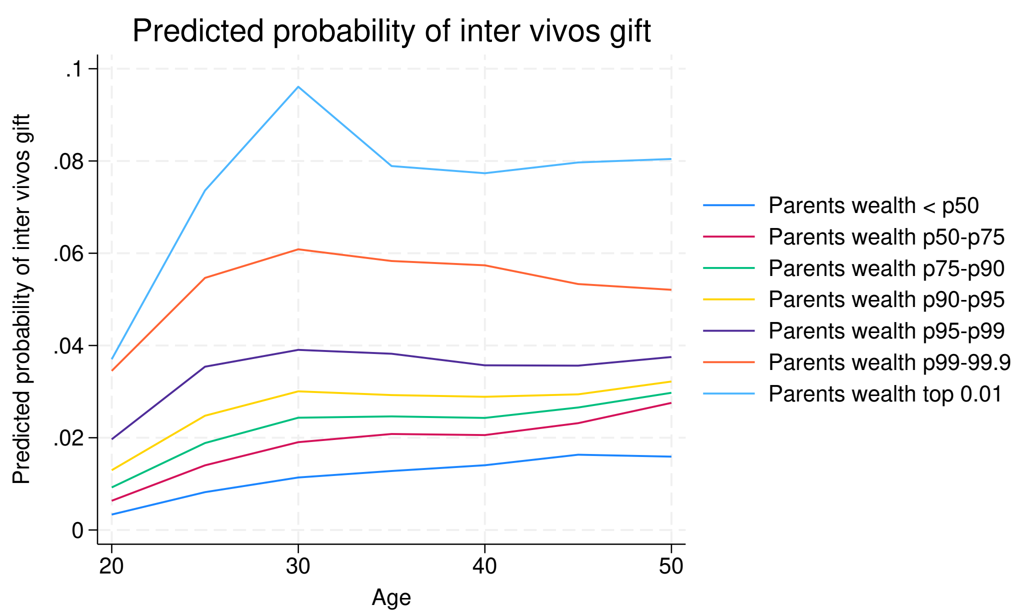


FIGURE SM.4.1 – Predicted probability of inter vivos transfers, by age and parental wealth

SM.4.1 Probability of transfer

We estimate the probability of receiving a gift as a Logit function of own age, own wealth group in period t-1, parental wealth group, and immigrant status. Figure SM.4.1 shows predicted probabilities of receiving a gift by own age and parental wealth group.

SM.4.2 Conditional value of transfer

We structure our data set in long form, where each observation represents an individual-year combination. This longitudinal format allows us to track changes in wealth, income, and family circumstances over time, capturing both static characteristics and dynamic financial behaviors that may influence gift patterns. Our machine learning model incorporates a carefully selected set of explanatory variables spanning multiple dimensions of individual and family circumstances:

- **Demographics:** male (binary), age (continuous), number of siblings (count), immigration status (binary).
- **Parental Vital Information:** mother's age (continuous), mother's survival status (binary indicator for alive at year-end), mother's mortality event (binary indicator for death during year), father's age (continuous), father's survival status (binary indicator for alive at year-end), father's mortality event (binary indicator for death during year).
- **Individual's Wealth Portfolio:** total wealth (normalized by average wealth), wealth dynamics (arc-percent change since previous year), asset allocation (equity share and housing share of total portfolio), portfolio re-balancing (arc-percent changes in equity and housing shares).
- **Parental Wealth Portfolio:** parents' combined wealth (normalized by average wealth), parents' asset allocation (equity share and housing share of portfolio)
- **Individual's Income Measures:** capital income (normalized by average wealth), labor income (normalized by average earnings), income dynamics (arc-percent changes in both capital and labor income)
- **Historical Intergenerational Transfers Indicators:** Ever received an inter vivos transfers or inheritances (binary).
- **Target Variable:** Our primary target variable is inter vivos transfers, normalized by the average wealth (AW) in the corresponding year.

We normalize wealth and income variables by the average wealth (AW) and average earnings (AE) in each year to ensure comparability across time. We keep only non-zero observations.

We now discuss how good our imputation of inter vivos transfers is. The explanatory power of our model is relatively high with an overall R-squared of about 0.94 for the training data, 0.33 for the out-of-sample test, and 0.82 for the entire data (Table SM.4.1). Figure SM.4.2 illustrates the remarkable performance of our gradient boosting model in capturing the relationship between wealth and inheritance across the wealth distribution. The figure displays the logarithm of average inheritance (normalized by average wealth) against wealth groups (same groups as in previous figures), with actual inheritance values shown in blue and model predictions in red. This logarithmic scale enables visualization of the wide range of inheritance values while highlighting relative differences.

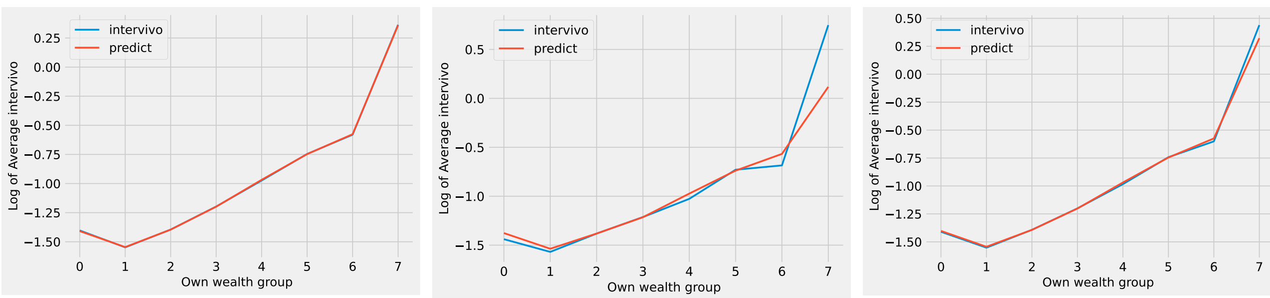
Across all three panels—training sample (A), test sample (B), and the entire sample (C)—the model accurately reproduces the distinctive non-linear pattern of gifts across the wealth distribution. Most notably, the model successfully captures several critical empirical patterns:

Table SM.4.1 – R^2 for Imputing Inter Vivos Gifts

	(1) Train Sample	(2) Test Sample	(3) Entire Data
R^2	0.94	0.33	0.82
Obs	43,058	10,765	53,823

Figure SM.4.2 – Average Inter Vivos Transfers over the Wealth Distribution

(A) Training sample (B) Test Sample (C) Whole sample



Notes: Panel A shows the average inheritance (in terms of average wealth) across the wealth distribution. Panels B and C show the same figure for test and entire samples, respectively.

1. The U-shaped relationship at the lower end of the wealth distribution, with relatively higher transfers for the very poorest households (negative net wealth), followed by a dip for households in the 0-50th percentiles.

2. The steady increase in transfers values from the 50th percentile upward, reflecting the increasing probability and magnitude of wealth transfers as household wealth rises

3. The acceleration in transfers amounts at the very top of the distribution (above the 99.9th percentile), where the curve steepens significantly, capturing the concentration of large inheritances among the wealthiest households

The model's predictions closely track actual values across both the training and test samples, demonstrating strong generalization performance rather than overfitting. This consistency between samples provides confidence that the imputation methodology will produce reliable estimates for the post-2014 period not covered by actual data. These results confirm that our machine learning approach successfully preserves the complex distributional characteristics of inter vivos patterns—a crucial requirement for subsequent analyses of wealth mobility and inequality. By accurately capturing how inter vivos transfers varies non-linearly with wealth, the model provides imputed values that maintain the important structural relationships observed in the empirical data, enabling more robust analysis of wealth dynamics over the entire 1993-2019 period.

References

DIETZ, P. O. (1968). Components of a measurement model: rate of return, risk, and timing. *The Journal of Finance*, **23** (2), 267–275. SM.3.3.1

FAGERENG, A., GUISO, L., MALACRINO, D. and PISTAFERRI, L. (2020). Heterogeneity and persistence in returns to wealth. *Econometrica*, **88** (1), 115–170. SM.3.3.1

FRIEDMAN, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232. SM.3.1, SM.4

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M. and DUCHESNAY, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, **12**, 2825–2830. SM.3.1