

Decory Edwards*

2026-02-11

*blank
0blank

3 Data

For the statistical analysis in this paper, I use the RAND Longitudinal version of the Household Retirement Survey (HRS) public data. I focus on the years 2002-2022; as this time horizon was filled with many economically relevant events (GFC, Covid), we can be creative regarding how to assess the accuracy of the data measured here. This is useful because, although there are many sources of income data to compare results to, there are less counterparts like this for return measures. The variables of interest fall into the categories: i) income, ii) wealth and portfolio composition, iii) returns, iv) trust, v) demographics, and vi) other controls.

3.1 Income

First, I use two measures of income: labor and total. Labor income is a narrow measure only capturing earnings and unemployment income. Total income is a more broad measure including retirement and capital income. First, it is clear that mean incomes are relatively flat over the period.

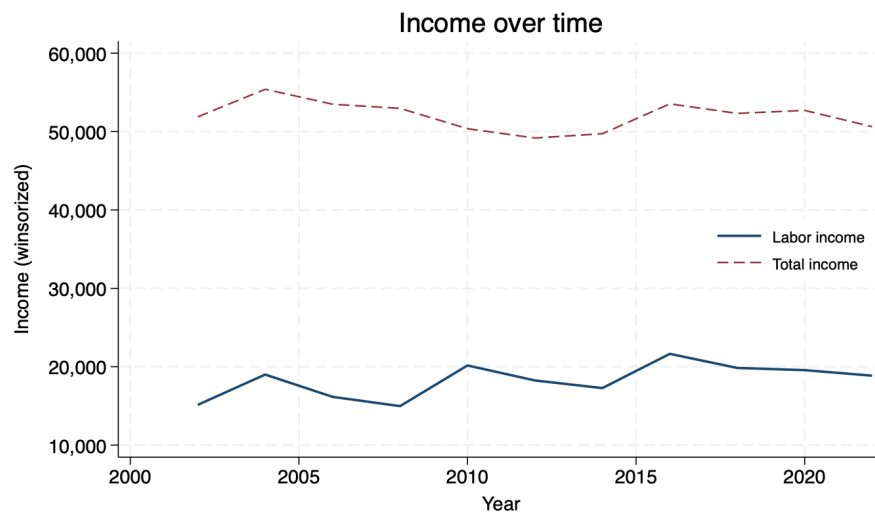


Figure 1: Income over time

In particular, I pooled the observations of the survey together to get a sense of the distribution of incomes measured in the survey. This is captured by the table

To further capture this point, I consider income growth as well by considering the log difference in income across waves in table .

Table 1: Mean income by year (real, winsorized)

Year	Labor income (mean \$)	Total income (mean \$)	Obs
2002	15,136	51,879	18,165
2004	19,003	55,398	20,129
2006	16,159	53,490	18,469
2008	14,976	52,958	17,217
2010	20,165	50,369	22,034
2012	18,250	49,178	20,554
2014	17,278	49,727	18,747
2016	21,644	53,539	20,912
2018	19,847	52,320	17,146
2020	19,568	52,702	15,723
2022	18,860	50,624	15,856

Real USD; winsorized at 1st and 99th percentile.

Table 2: Income (real, winsorized): summary statistics

Variable	Obs	Mean	SD	P50	P95	Min	Max
Labor income (real, winsorized)	204,952	18,329	37,031	0	97,922	0	215,257
Total income (real, winsorized)	204,952	52,003	68,034	29,103	177,494	0	457,538

Real USD, winsorized; summary over person-years.

Table 3: Income growth: summary statistics

Variable	Obs	Mean	SD	P50	P95	Min	Max
Log labor income growth (2004)	4,053	-0.1429	0.9658	-0.0490	1.2037	-5.3748	4.8562
Log labor income growth (2006)	5,334	-0.1031	0.9112	-0.0256	1.1879	-5.7080	5.2030
Log labor income growth (2008)	4,468	-0.0873	0.8806	-0.0232	1.1282	-5.1897	5.2897
Log labor income growth (2010)	3,769	-0.1265	0.9786	-0.0130	1.1501	-5.5375	4.9488
Log labor income growth (2012)	6,475	-0.0926	0.9230	-0.0514	1.2295	-5.4439	6.3863
Log labor income growth (2014)	5,439	-0.0719	0.9303	-0.0306	1.2454	-5.5520	5.3963
Log labor income growth (2016)	4,319	-0.0431	0.9628	0.0201	1.2798	-5.1496	5.5314
Log labor income growth (2018)	5,269	-0.0510	0.9891	-0.0228	1.3566	-6.2785	6.3304
Log labor income growth (2020)	4,337	-0.0637	1.0101	-0.0051	1.3046	-5.7117	6.3082
Log labor income growth (2022)	3,337	-0.1797	1.0885	-0.0875	1.3148	-5.9752	5.4186
Log total income growth (2004)	15,707	-0.0342	0.9076	-0.0276	1.3372	-6.1775	6.6288
Log total income growth (2006)	17,275	-0.0296	0.9022	-0.0220	1.3168	-6.4724	6.9622
Log total income growth (2008)	15,997	-0.0225	0.8646	-0.0152	1.2958	-5.6983	5.9585
Log total income growth (2010)	14,507	-0.0821	0.9196	-0.0055	1.2191	-7.0460	6.1571
Log total income growth (2012)	18,706	-0.0480	0.9596	-0.0325	1.4187	-6.7117	6.9345
Log total income growth (2014)	17,128	0.0090	0.9139	-0.0001	1.3945	-6.4459	6.2165
Log total income growth (2016)	15,023	-0.0109	0.9538	0.0037	1.3725	-6.4116	6.9414
Log total income growth (2018)	15,468	-0.0403	1.0045	-0.0362	1.4816	-7.0159	6.7023
Log total income growth (2020)	13,308	-0.0011	0.9655	0.0012	1.4729	-5.9648	7.2582
Log total income growth (2022)	11,444	-0.0932	0.9818	-0.0658	1.3958	-7.2233	5.9135

Two-year log difference by end year; $\ln(\text{income})$, zero income dropped (N reflects). Labor = earnings+unemployment; total = all components.

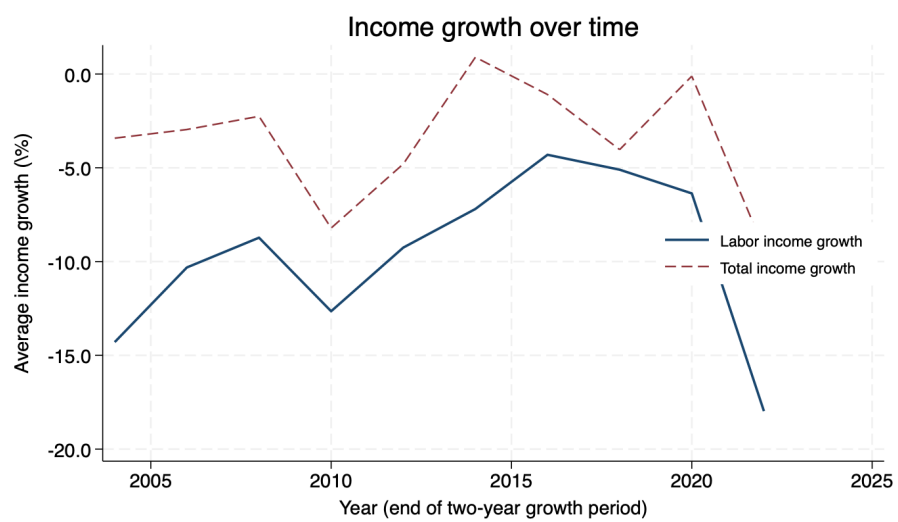


Figure 2: Income over time

An important finding in the literature on earnings in the U.S. is that it is generally hump-shaped over the lifecycle. It is a good sign then that the following tables show

Table 4: Mean income by age group (2022)

Age (midpoint)	Labor income (mean \$)	Total income (mean \$)	Obs
25	0	11,355	2
30	32,109	88,033	8
35	19,994	60,831	28
40	25,892	47,445	95
45	36,851	63,805	233
50	43,922	69,037	1,286
55	38,021	62,013	2,285
60	28,185	55,428	2,924
65	14,749	46,993	2,641
70	6,732	43,629	2,204
75	3,487	43,555	1,401
80	1,464	39,317	1,421
85	528	37,902	871
90	519	39,475	361
95	0	46,538	79
100	0	14,517	16
105	0	11,112	1
.	.	.	0

Five-year age bins (e.g., 50 = 50–54). Real USD, winsorized.

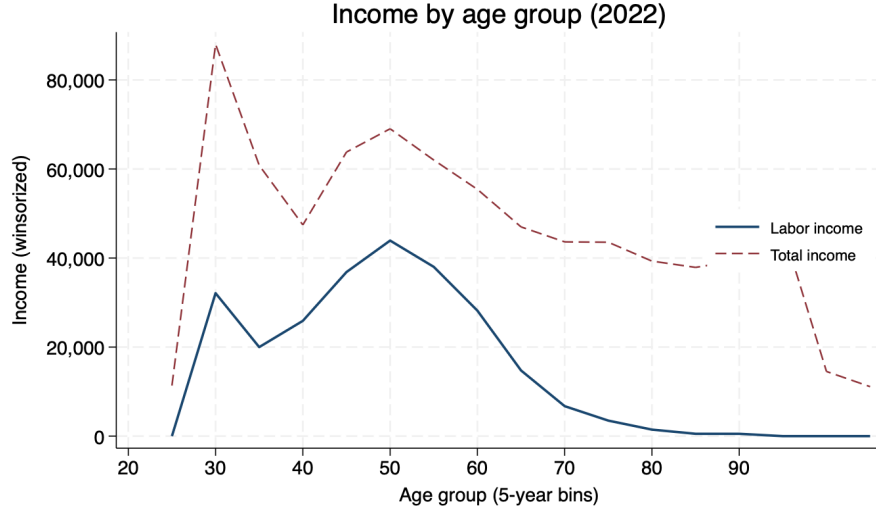


Figure 3: Income by age group (2022)

Another empirical finding in the literature on earnings is that on average individuals with more education earn higher income. The trend capture in the table suggests that the earnings data is in line with one would respect regarding earnings for a representative survey in the U.S.¹

Table 5: Mean income by education group (real, winsorized)

Education	Labor income (mean \$)	Total income (mean \$)	Obs
no hs	6,075	23,987	44,654
hs	12,590	39,600	63,826
some college	20,297	53,482	47,390
4yr degree	32,778	82,576	24,820
grad	37,444	104,064	23,091

Real USD, winsorized. no hs = <12y; hs = 12y; some college = 13–15y; 4yr = 16y; grad = 17+y.

¹Although the HRS oversamples older households, I use the provided the respondent-level weights for this interpretation of the summary statistics of the data.

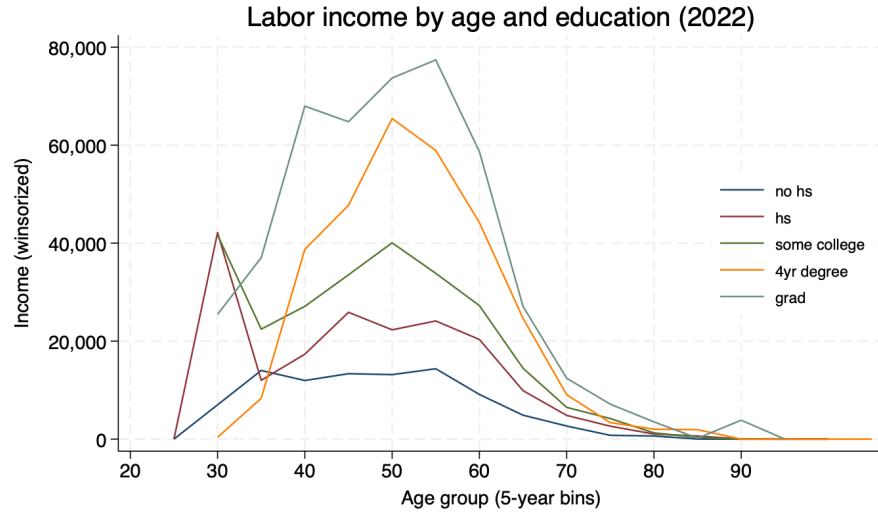


Figure 4: Labor income by age and education (2022)

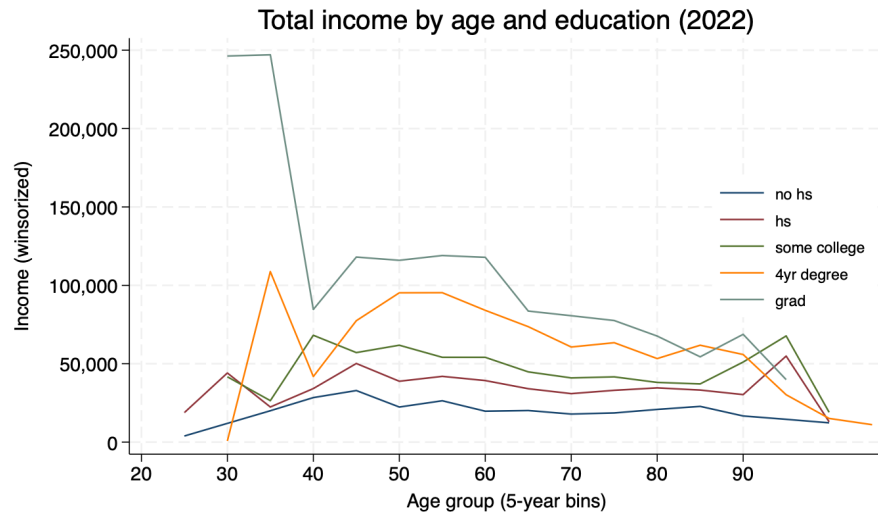


Figure 5: Total income by age and education (2022)

3.2 Wealth and asset class definitions

The HRS asks a number of questions aimed at measuring household wealth in the sample. I collect data on **interest income and dividends**, **capital gains**, **net investment flows**, and **previous period wealth holdings**, all of which are necessary to define the desired measure of household

returns. Since population-level administration data on these objects for individual taxpayers is not available in the U.S., it is important to understand each component needed in the return calculation. This will help to determine whether or not our measured distribution of returns in this dataset is sensible.

3.2.1 Capital gains

The possible asset classes for which capital gains can be computed in the HRS are i) primary residences, ii) secondary residences, iii) other real estate, iv) private business, v) IRA/Keogh (or “retirement”), vi) stocks/mutual funds, vii) bonds, viii) checking/savings/money market, ix) cds/t-bills, x) vehicles, xi) other assets. The possible liabilities are i) mortgages on primary residence, ii) mortgage on secondary residence, iii) other home loans, and iv) total other debt.

Capital gains for each asset class are computed as the estimated change in valuation across survey waves. The general trend is that, these changes in valuation tend to move around alot over the period. That said, capital gains are generally higher at the end of the period than at the start of the period for most asset classes. This can be seen for capital gains to busines ownership, in figure

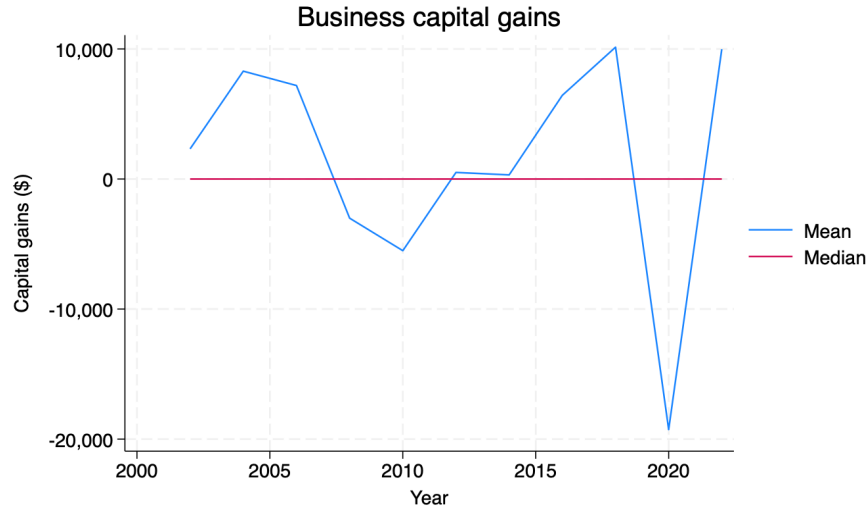


Figure 6: Capital gains: business, by year

The opposite is true for the bonds asset class: capital gains have generally fallen on average over the time period. This can be seen in the figure

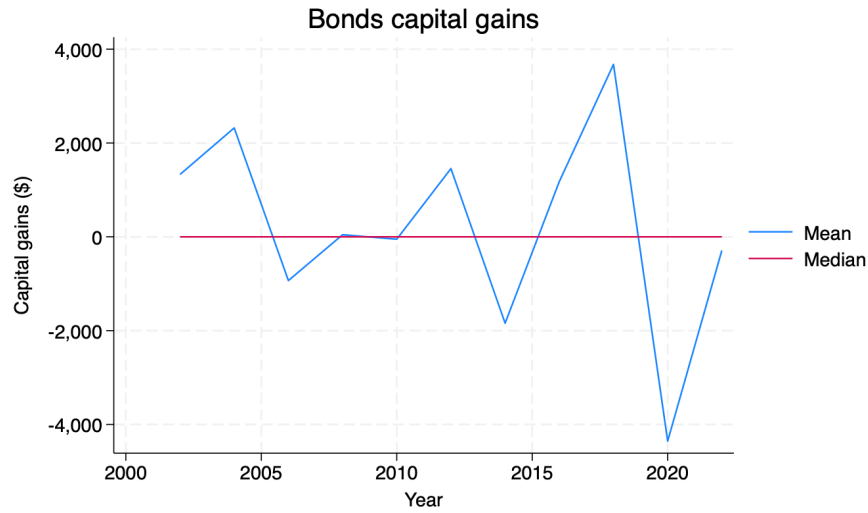


Figure 7: Capital gains: business, by year

Another feature that is apparent by taking a look at capital gains is that, most respondent hold no assets in a particular class. This can be seen by the vertical red line capturing a median of 0 for each wave of the survey. Residential assets generally are the bulk of individual portfolios, and yet the median is still 0 here in almost every wave. This is another sign that the data is sensible so far – large amount of non-participation despite evidence of returns is consistent with the empirically documented equity premium puzzle.

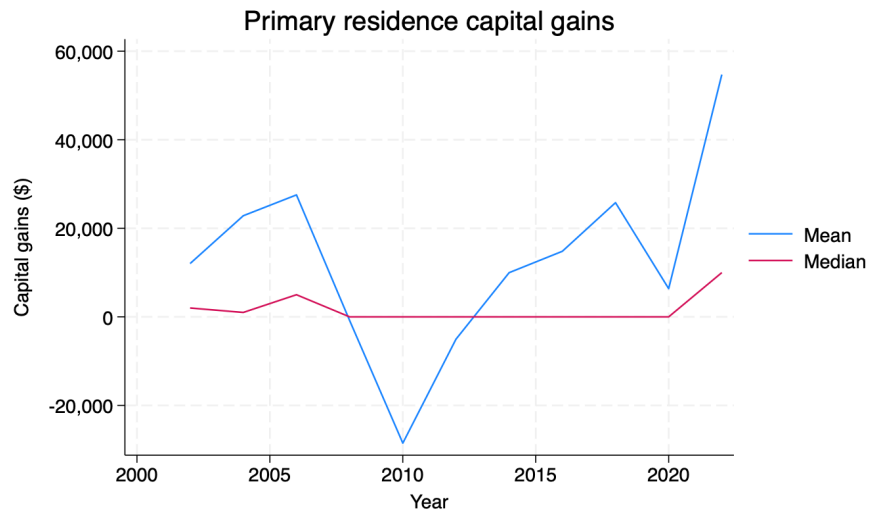


Figure 8: Capital gains: residential, by year

Notably, the dip in capital gain leading up to 2010 for primary residences, secondary residences, and real estate is also a good sign regarding how realibly measured the data is.

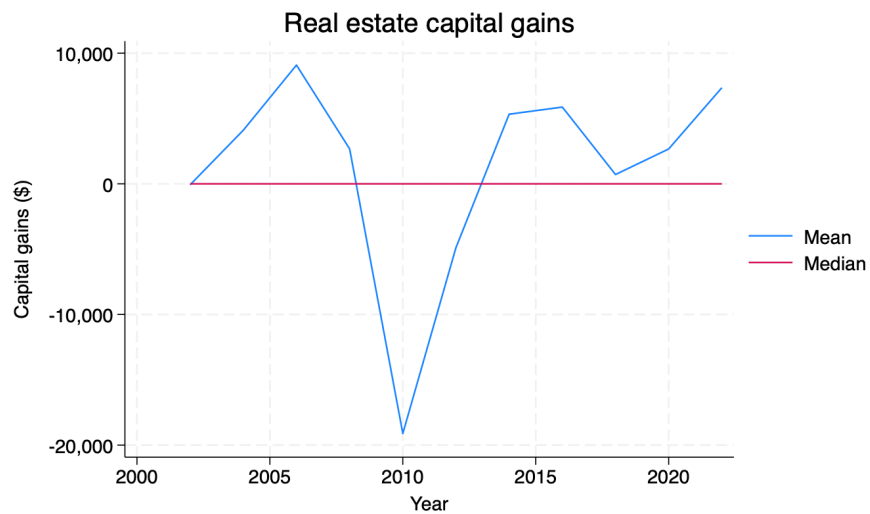


Figure 9: Capital gains: residential, by year

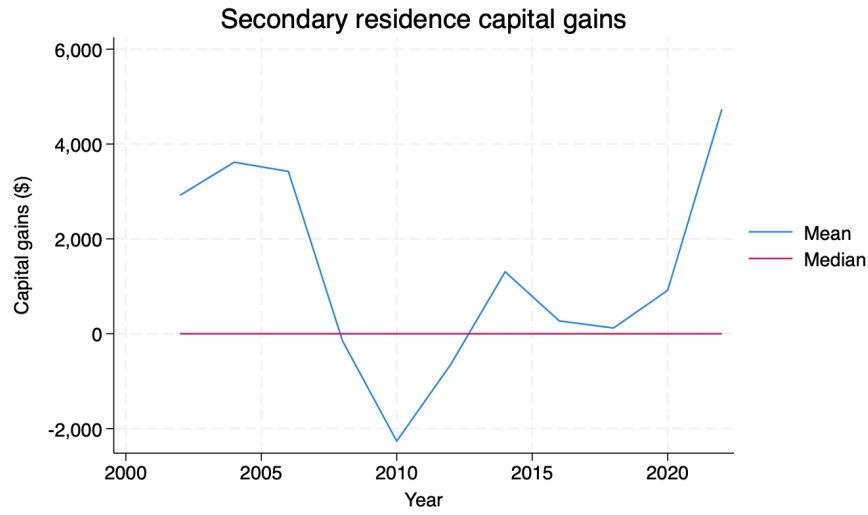


Figure 10: Capital gains: residential, by year

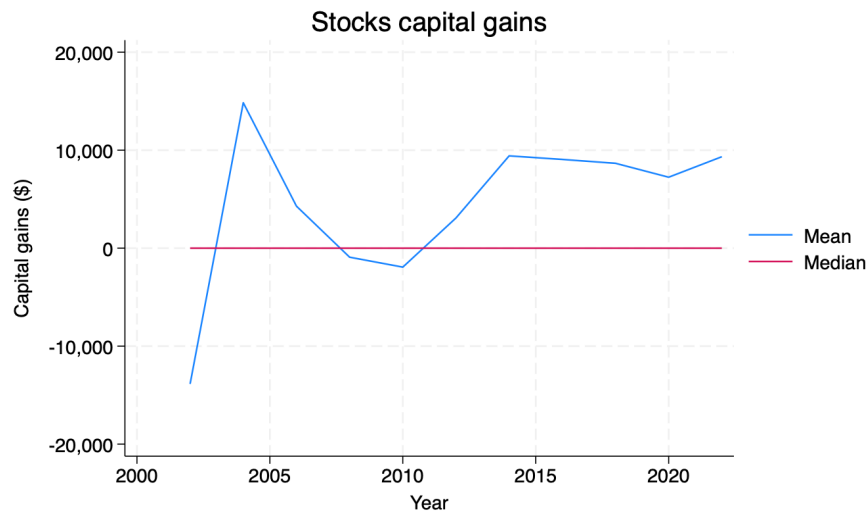


Figure 11: Capital gains: stocks, by year

3.2.2 Interest income and dividends

The HRS RAND longitudinal file has a distinctive way in which interest income and dividends received on these assets is measured. In particular, there is a variable which measures “household capital income received over the last calendar year, including business or farm income self-employment

earnings, gross rent, dividend and interest income, trust funds or royalties, and other asset income”. This particular feature of the data is the driving factor for the key modeling assumptions of this paper regarding measuring returns: asset classes are defined as narrowly as they can be given observation of interest income and dividends on those assets. For example, although we see capital gains for stocks, business, bonds, and real estate, we do not observe interest income or dividends on these assets individually. Thus, the narrowest asset class defined in this paper will be called “core” and will be comprised of these assets.

The RAND version of the data offers a number of variables measuring pension and annuity income as well as other forms of retirement income. I use this to construct a measure of returns to retirement assets and to define a broader notion of portfolio returns by considering returns to core and retirement assets. A key assumption is that there are no interest income or dividends earned on residential assets². With this assumption, we can consider interest income and dividends on the entire portfolio as the same as interest income and dividends on the portfolio with just core and retirement assets. Thus, to compute a measure of returns to net wealth, I need to add in the remaining available capital gains per asset class (for those that receive no interest income).

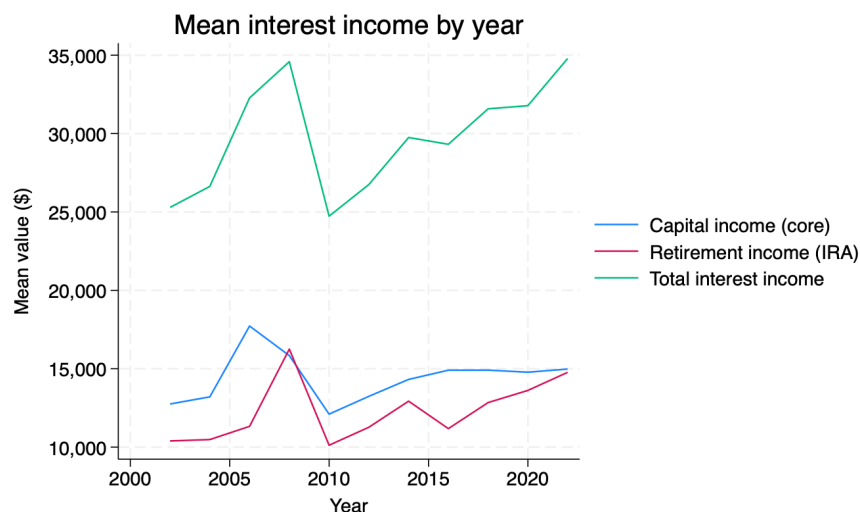


Figure 12: Interest income mean by year

3.2.3 Net investment flows

Net investment flows are vital in computing returns accurately because capital gains and interest income can miss other flows of value into a given asset class. In the HRS dataset, net investment flows are the only variable relevant to the returns calculation that RAND did not clean and process. For this reason, there are substantially less observation for this variable per asset class than the others. To work around this, I do two things. First, I process the net investment flow per asset class myself by using the associated flag variable (asking yes or no if an individual has “bought or

²Or on any of the asset classes for which capital gains are available for.

sold since the previous wave”) for each variable. Second, I assume that if an individual receives interest income on that asset class, but their flow is nonmissing, than the nonmissing flow to that asset class is treated as 0.³

With this in mind, I present a figure of net investment flows into the assets available in the dataset. As you can see, the magnitudes for flows into a given class are comparable. I use these flows to construct net investment flows into i) core assets, ii) retirement assets, iii) residential assets, iv) core+residential assets, and v) net wealth.

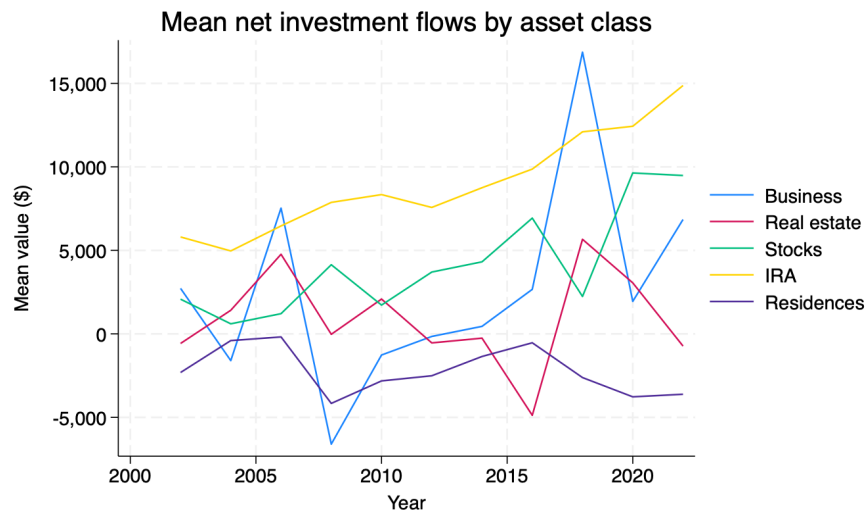


Figure 13: Flows by asset, mean by year

3.2.4 Wealth and inequality

After defining asset classes, I consider mean wealth within these narrowly defined asset classes for each year of the sample. Interestingly, retirement assets seem to perform the worst in terms of average returns. Core assets seem to offer the best performance, and all asset classes see a downward dive leading up to 2010 – historically accurate in the context of investment performance during the global financial crisis.

³Non-missing interest income (and capital gains) indicate participation within an asset class.



Figure 14: Wealth mean by year (components)

I turn attention to measures of wealth in the sample based on the portfolio definitions: i) net wealth in core assets, ii) net wealth in core and retirement assets, iii) total net wealth.

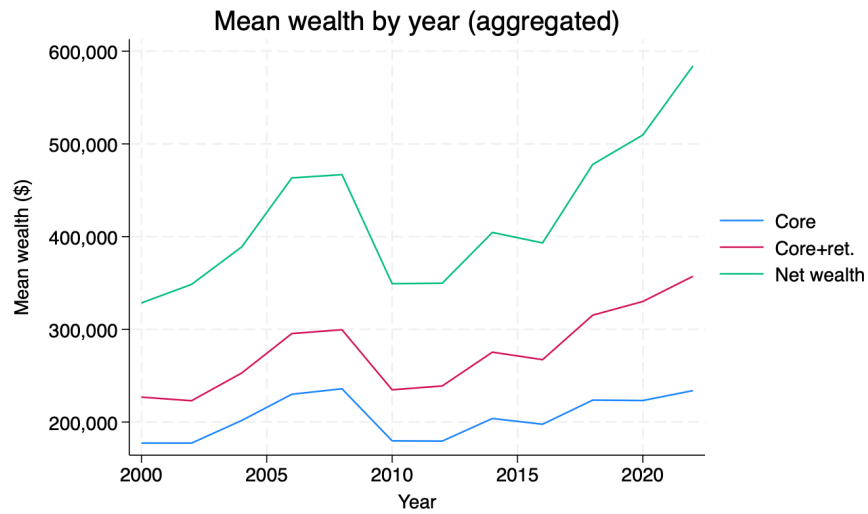


Figure 15: Wealth mean by year (aggregated)

I consider the distribution of these measures of income and wealth across respondents of the survey in the following figures. For income, the measure of labor income is significantly more unequally distributed than the measure of total income in the sample. This is likely an artifact

of the sampling method of the HRS (older likely overrepresented and are towards the end of their working years).

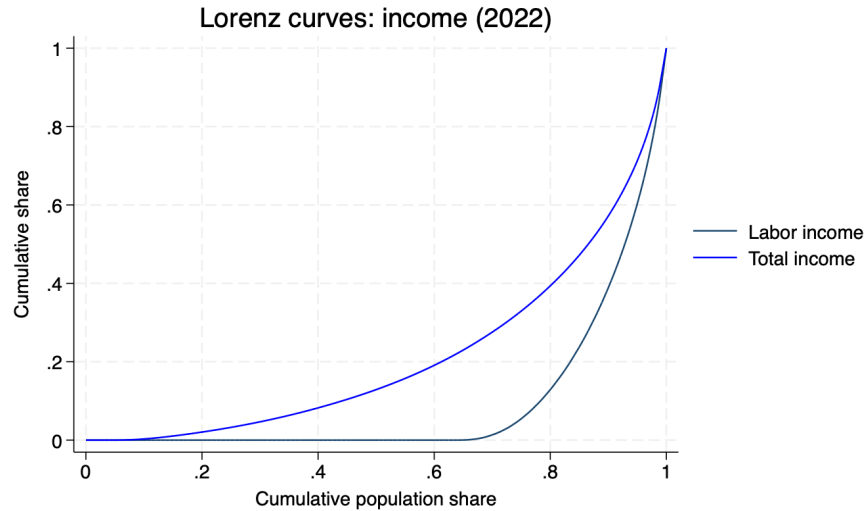


Figure 16: Lorenz: income (2022)

For wealth, holdings in core assets and retirement assets are more unequally distributed than holdings in residential assets. This is in line with common knowledge that home ownership is the most common form of asset ownership in the U.S.

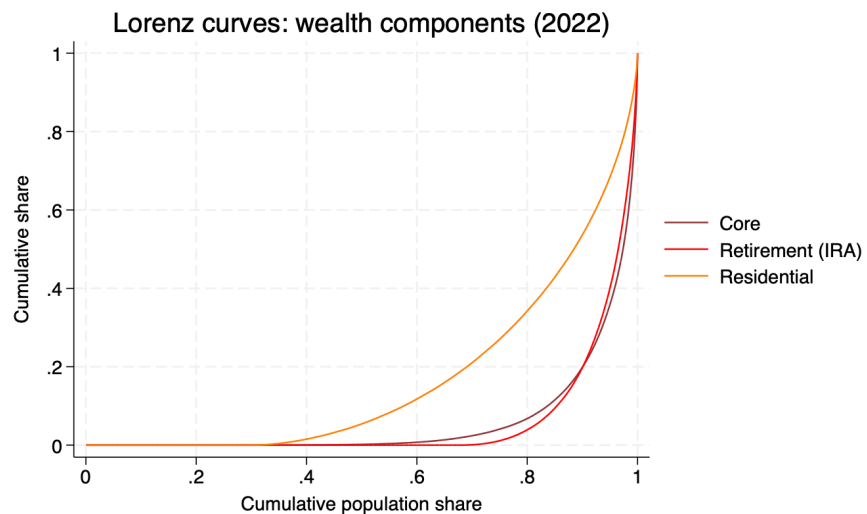


Figure 17: Lorenz: wealth components (2022)

Interestingly, the distribution of net wealth become less unequal when it is extended to incorporate retirement assets on top of core assets. this can be seen in the following figure .

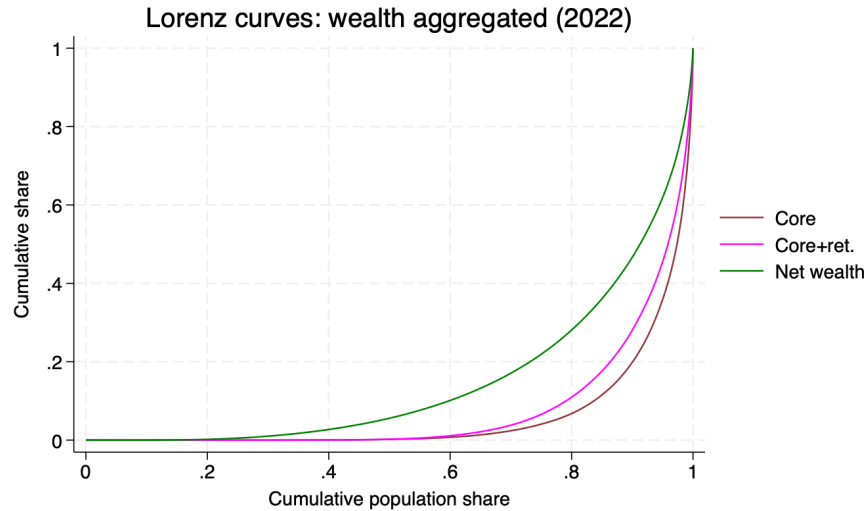


Figure 18: Lorenz: wealth aggregated (2022)

3.2.5 Portfolio composition

To describe the composition of portfolio during the period, I document mean portfolio shares for the relevant asset classes.

As you can see from figure , residential assets typically dominate portfolios for those who hold assets. The relative composition of portofolio (i.e. allocation between core, retirement, and residential assets) does not change much over the period.

Table 6: Mean portfolio share by asset class and year

Year	Core	Residential	Retirement
2000	0.168	0.440	0.093
2002	0.161	0.451	0.085
2004	0.156	0.460	0.085
2006	0.147	0.472	0.087
2008	0.143	0.464	0.091
2010	0.128	0.442	0.094
2012	0.124	0.435	0.096
2014	0.125	0.441	0.098
2016	0.118	0.448	0.094
2018	0.114	0.461	0.101
2020	0.113	0.465	0.111
2022	0.100	0.476	0.106

Core = bonds, stocks, real estate, business; Residential = primary + secondary; Retirement = IRA.

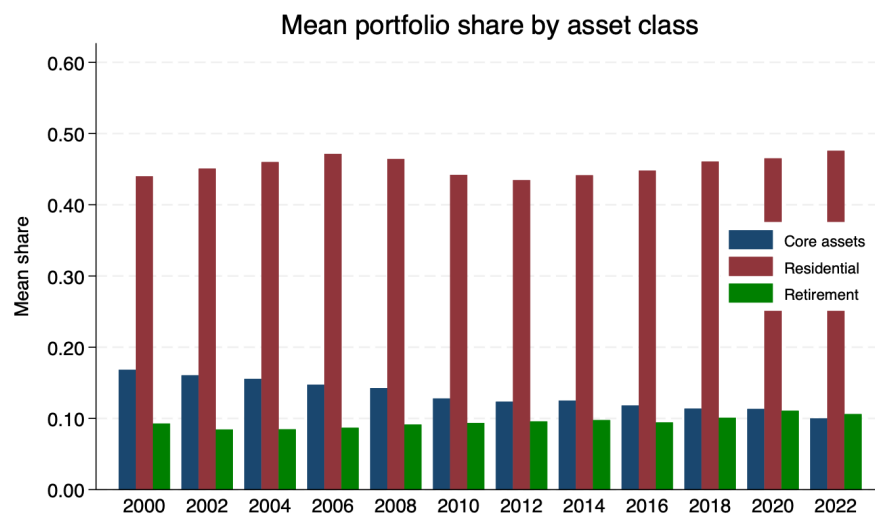


Figure 19: Share core by wealth percentile (2022)

Investment behavior is likely to vary with income and wealth. To see this relationship, I start by considering the share of core assets to gross wealth conditional on percentiles of the income/wealth distribution the respondent is in. As you can see in figure, those with labor earnings below the 40th percentile still have a significant portion of their wealth in core assets.

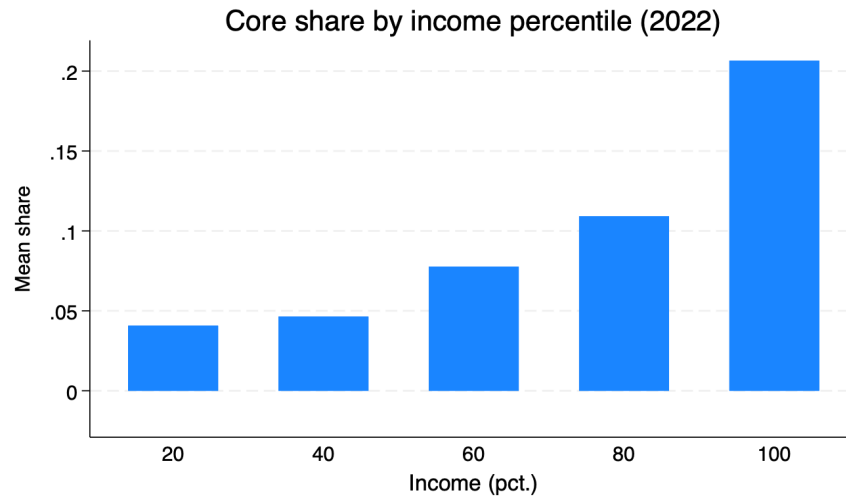


Figure 20: Share core by income percentile (2022)

This, however, is not true for shares conditional on wealth percentiles, suggesting that shares in core assets are much more unequal. Individuals below the 40th percentile hold virtually no core assets. That said, in both cases, it is clear: high earners and high net worth individual have higher shares of their assets invested into their defined portfolios.

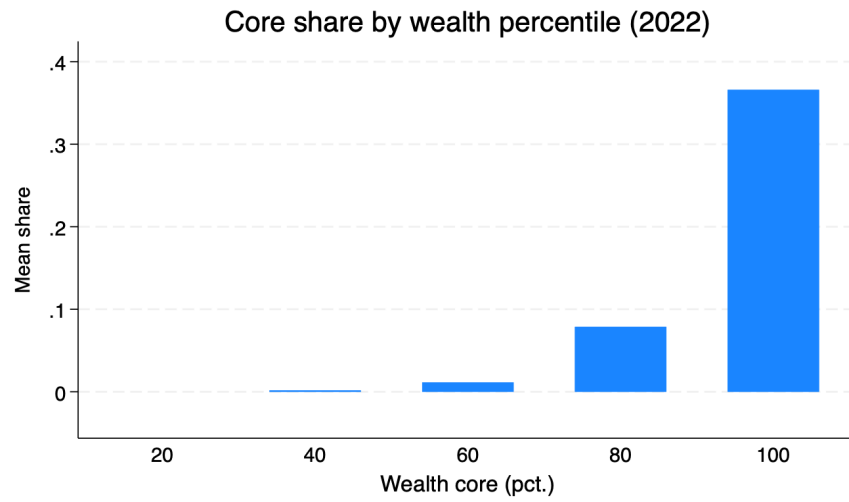


Figure 21: Share core by wealth percentile (2022)

The pattern persists when considering the share of core and retirement assets to gross wealth. The relationship is striking for wealth: top 20% income earners have 40% of their assets in core and retirement assets, while the top 20% of the wealth distribution have almost 60% of their assets in core and retirement assets!

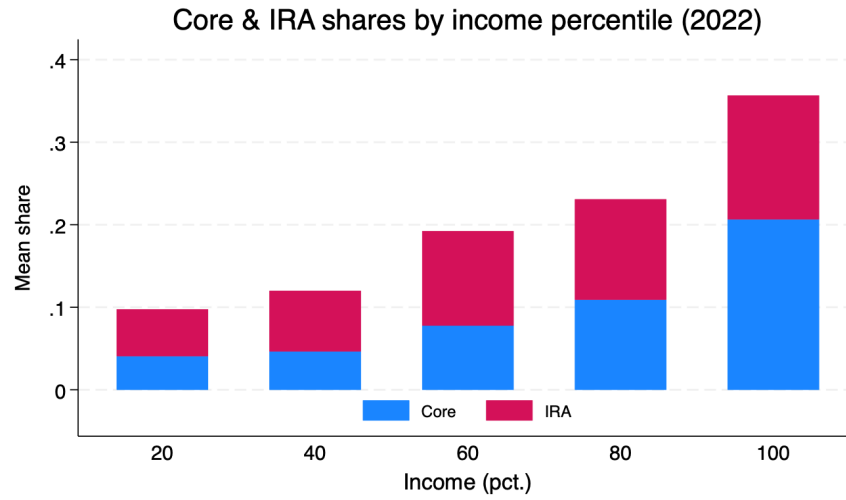


Figure 22: Share core and IRA by income percentile (2022)

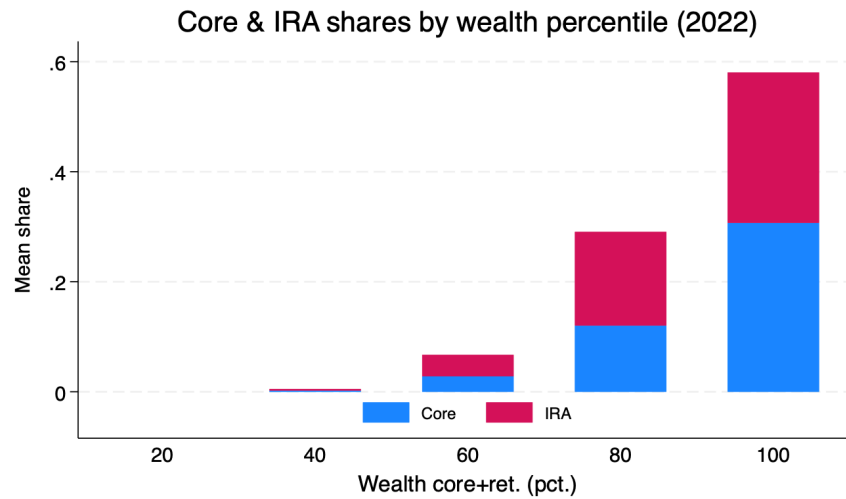


Figure 23: Share core and IRA by wealth percentile (2022)

These patterns again persist when considering the share of core, retirement, and residential assets to gross wealth. When retirement assets are added in, we see that most earners hold some wealth in one of the three asset classes: the bottom 20% of earners hold 60% of their assets in these

assets.

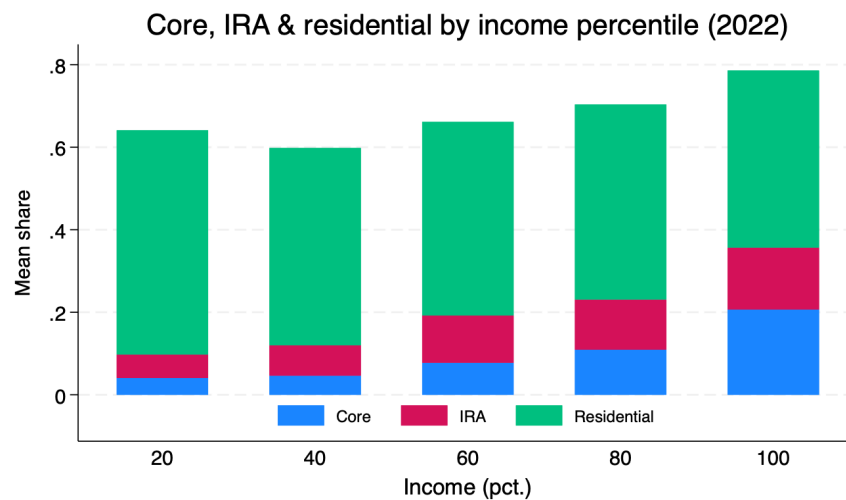


Figure 24: Share core, IRA and residential by income percentile (2022)

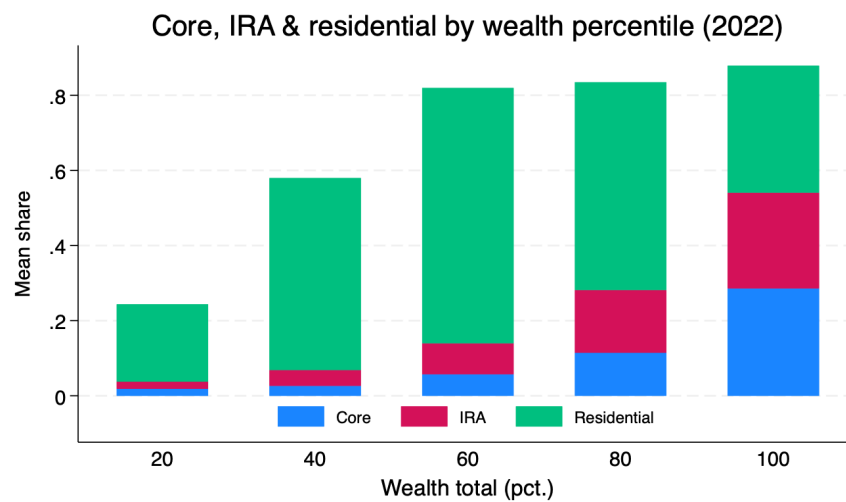


Figure 25: Share core, IRA and residential by wealth percentile (2022)

Lastly, I consider another measure of inequality based on portfolio composition. The figure shows us, for a given threshold invested in an asset class, how much of the total value of that asset does

the respondent hold. If for smaller thresholds, like 25% of your gross wealth invested in an asset class, an investor is able to hold a large share of the total value of that asset, then this suggest significant inequality within that asset class.

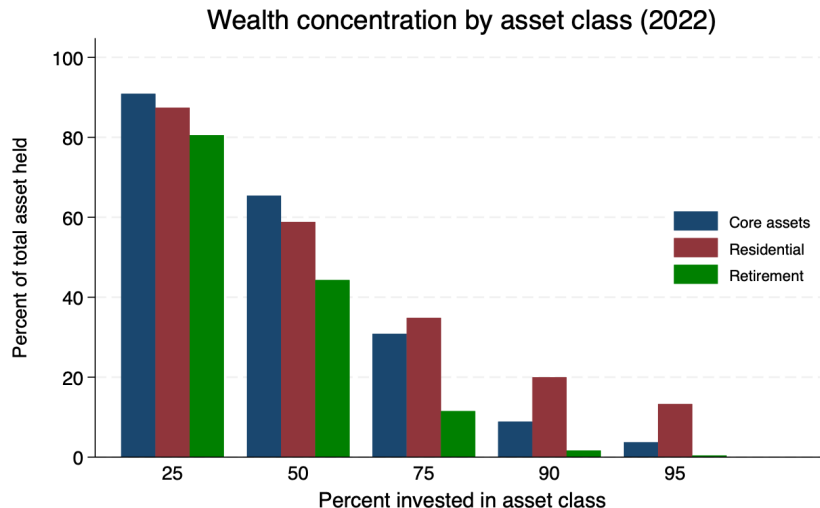


Figure 26: Share concentration (2022)

3.3 Returns

I follow recent literature using Norwegian administrative tax data at the population level and using PSID in terms of the components used to construct a measure of returns. However, I take seriously the structure of the dataset I am working with (the HRS RAND longitudinal file) in defining the portfolio/asset class for which the returns are being accrued to. As mentioned before, core assets is the most narrowly defined asset class here because interest income and dividends does not disaggregate into the asset classes: stocks, bonds, private business, IRA/Keogh.

Returns to core assets are given by

Returns to retirement assets are given by

Returns to residential assets are given by

Returns to core and residential assets are given by

Returns to net wealth is given by

I present the means for the return measures in the following figures. I group them by the returns at the asset level (core, retirement, residential) and at the portfolio level (core, core and retirement, net wealth).

I used the formula from **Daminato2024**:

$$r_t = \frac{y_t^c + cg_t - y_t^d}{A_{t-1} + .5F_t}$$

where y_t^c interest income and dividends, capital gains cg_t measured as the difference between reported stock across waves, F_t net investment flows, y_t^d payments on debt (in the RAND longitudinal

file, the variables were mentioned are all in net terms so this variable was 0), and A_{t-1} total net wealth at beginning of previous period.

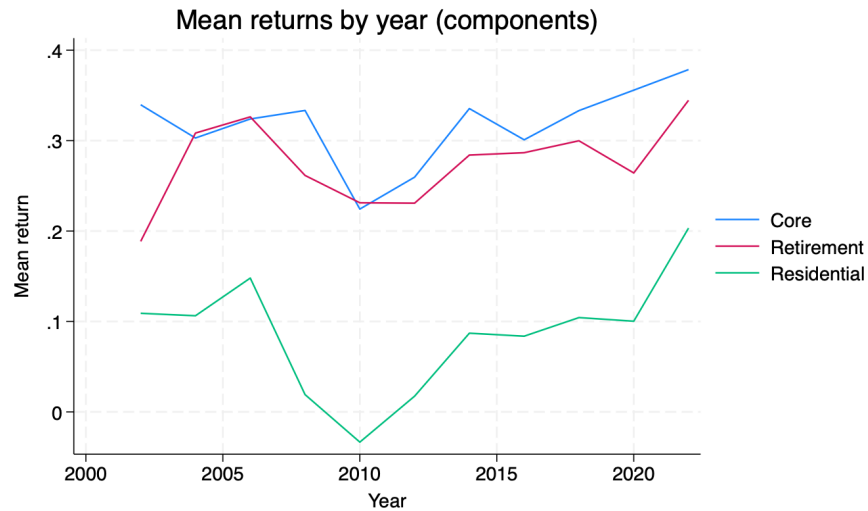


Figure 27: Returns mean by year (components)

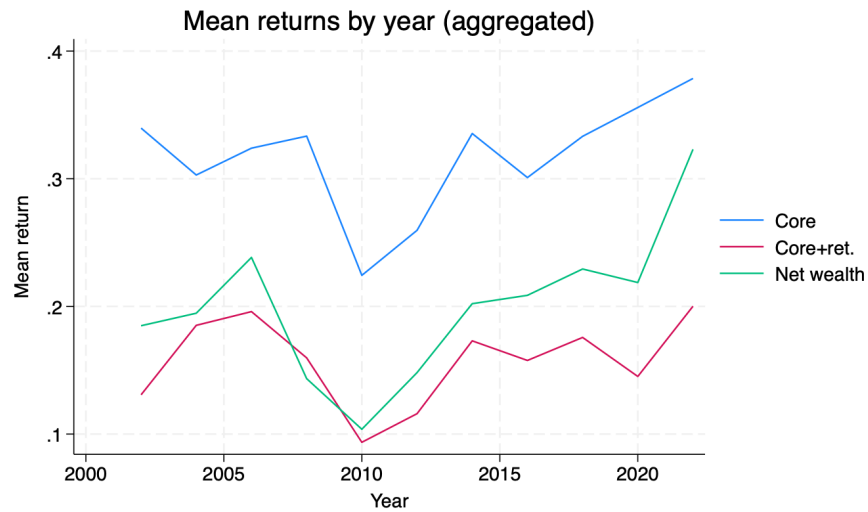


Figure 28: Returns mean by year (aggregated)

The following figures give a better idea of the distribution of the return measures across respondents.

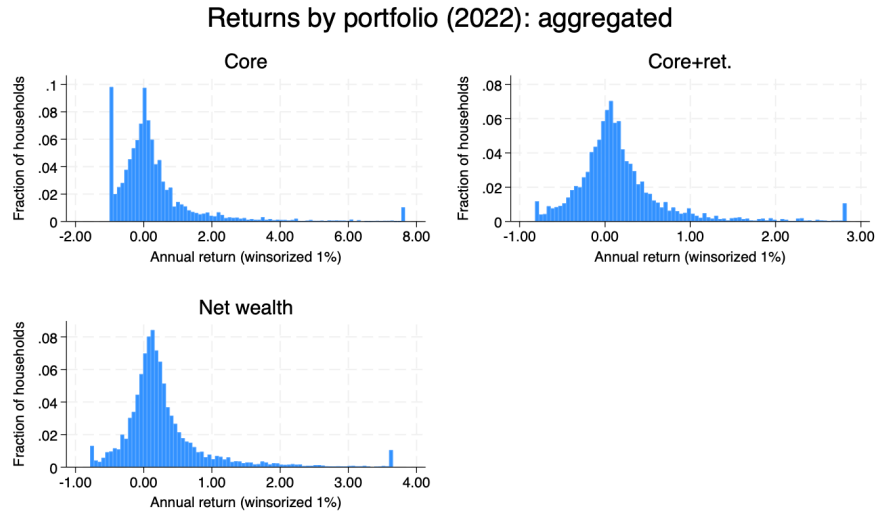


Figure 29: Returns histogram, aggregated (2022)

These are encouraging, as shape of this distribution closely resembles early estimates of the empirical distribution of individual realized returns in the Norwegian population data.

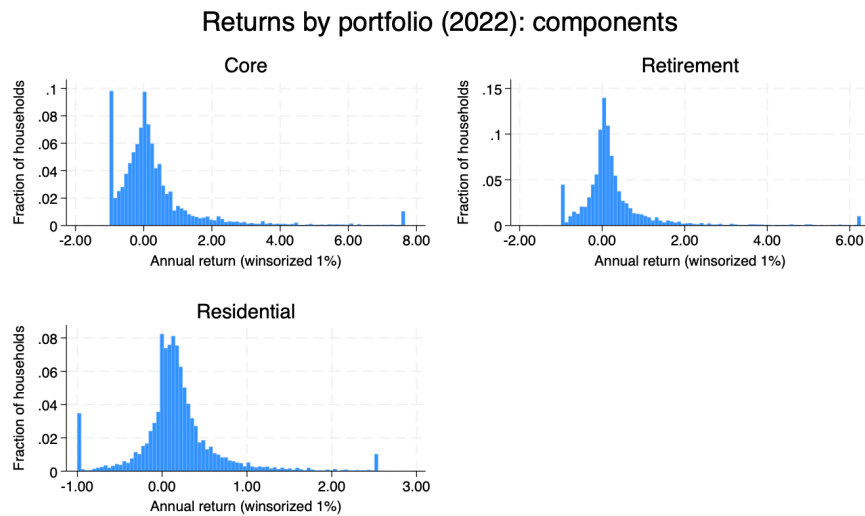


Figure 30: Returns histogram, components (2022)

3.4 Income and wealth

Income and wealth are understood to be positively correlated. To see this play out in the HRS data, I first look at mean labor income by wealth percentile. In this figure, it is clear that a positive relationship seems to hold.

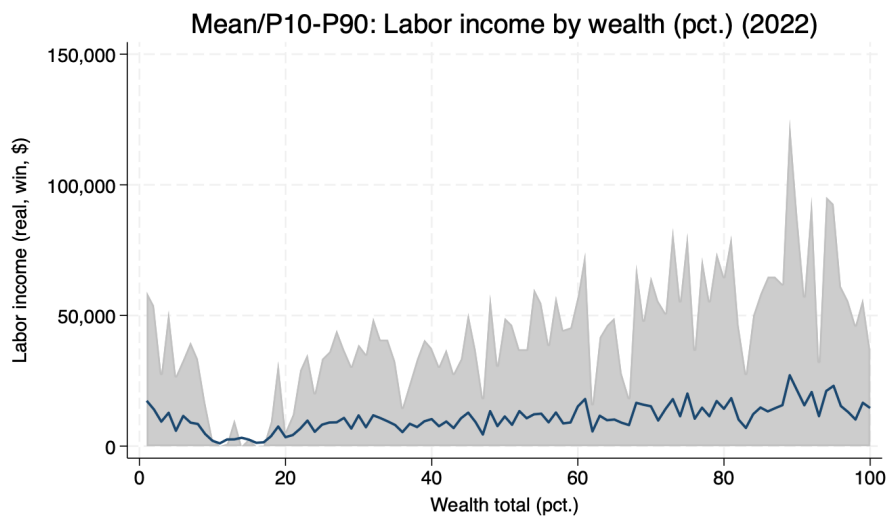


Figure 31: Labor income mean/P10-P90 by wealth percentile (2022)

The literature on income and wealth also documents significantly more inequality in the the upper tails of the wealth distribution than in the income distribution. The observations towards to top and bottom of the wealth distribution in the following figure seem to suggest a non-linear relationship between the variables.

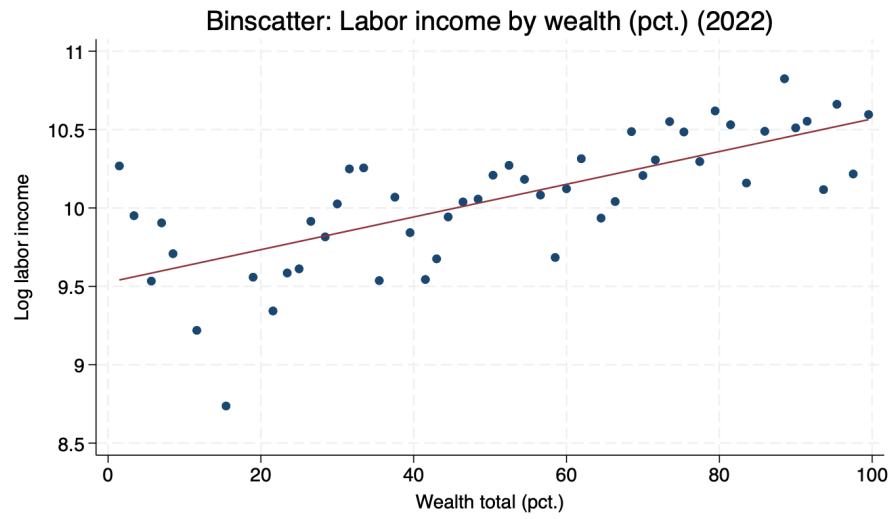


Figure 32: Log labor income binscatter (2022)

I repeat this for the measure of total income and both patterns persist.

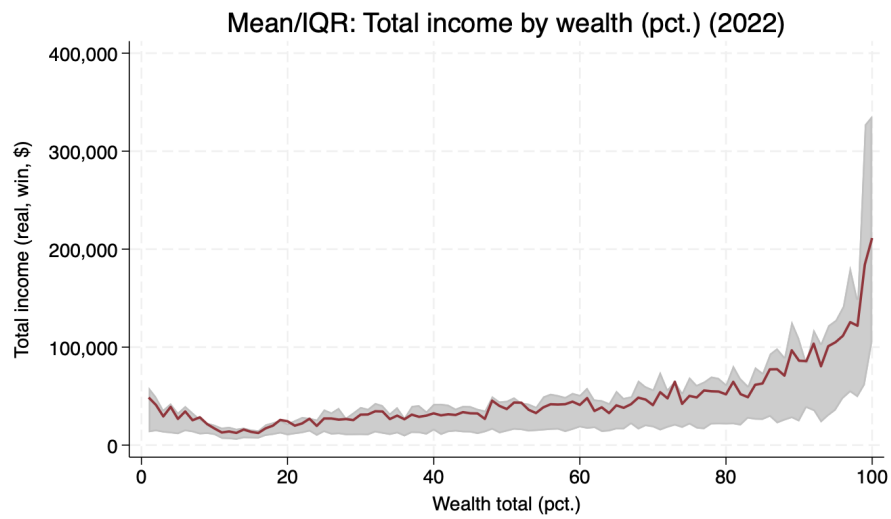


Figure 33: Total income mean/IQR by wealth percentile (2022)

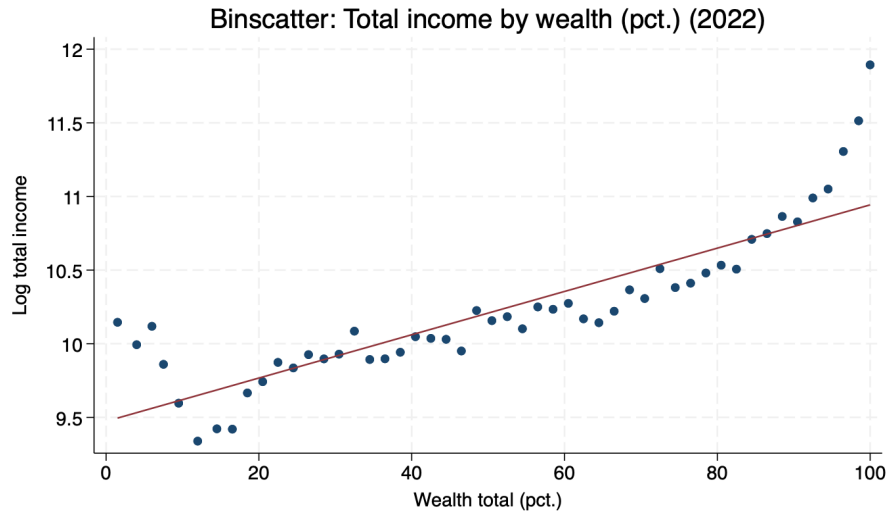


Figure 34: Log total income binscatter (2022)

3.5 Returns and wealth by portfolio

A positive relationship between wealth and returns has also been documented, referred to as *scale dependence*. In the next three figures, I show that there does seem to be a positive statistical relationship between the return measure (core, retirement, residential) and wealth.

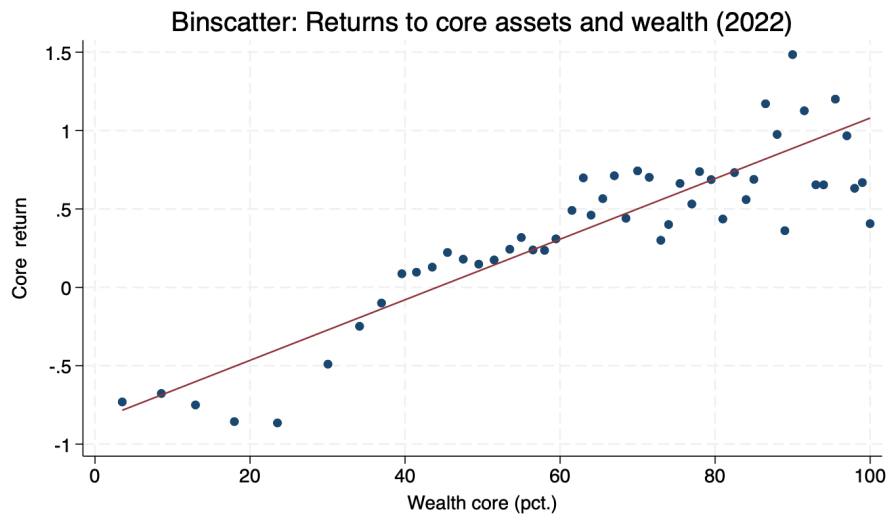


Figure 35: Core return binscatter (2022)

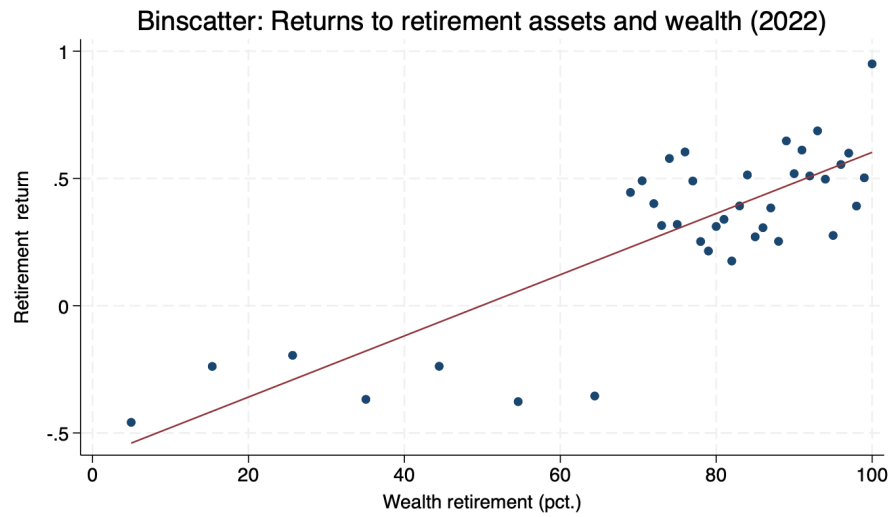


Figure 36: Retirement return binscatter (2022)

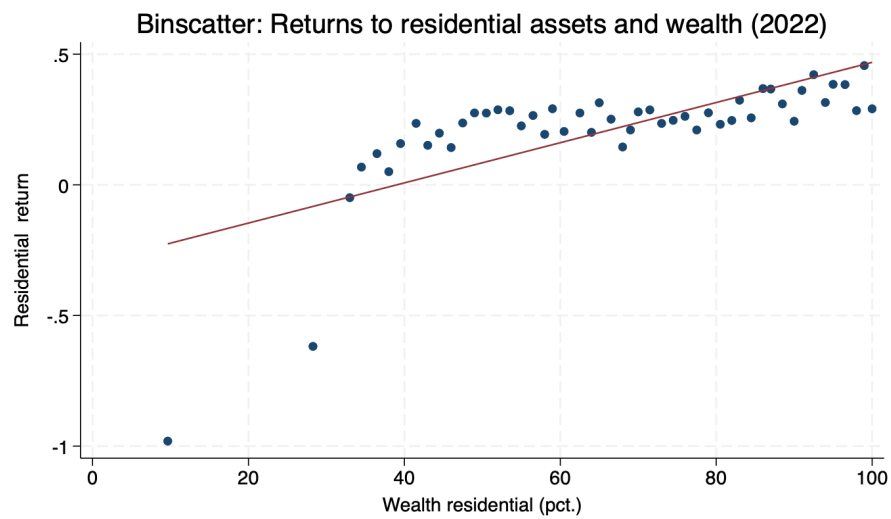


Figure 37: Residential return binscatter (2022)

I also show the positive statistical relationship between returns and wealth at the portfolio level, when comprised of core assets and retirement assets.

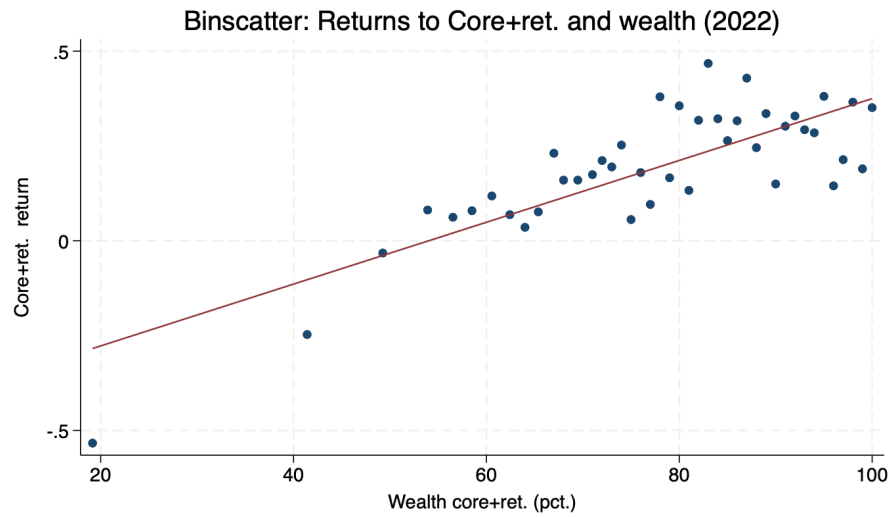


Figure 38: Core+IRA return binscatter (2022)

Interestingly, the relationship between wealth and returns to total net wealth does not appear positive. In fact, they correlation is slightly downward-sloping.

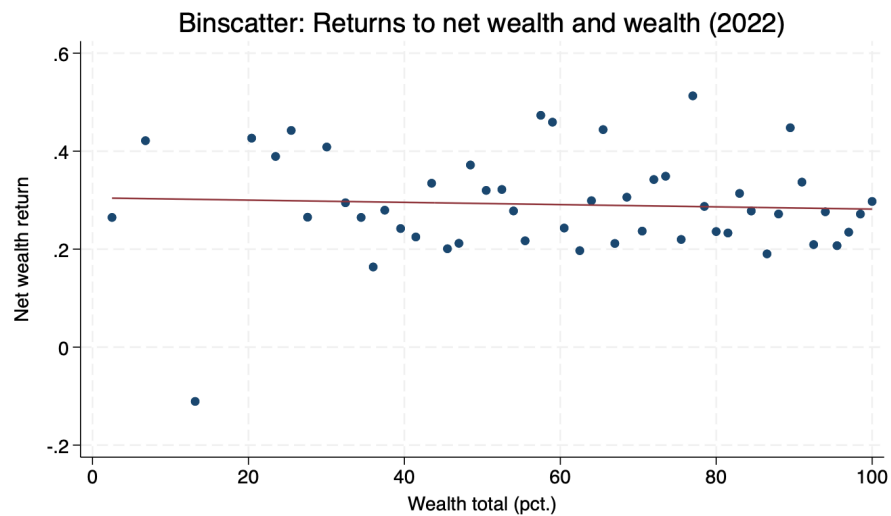


Figure 39: Net wealth return binscatter (2022)

3.6 Trust

I turn my attention to the suite of trust questions in “Section V: Modules” section of the 2020 HRS data. There are a total of 8 questions, asking respondents to say on a scale of 1-10 “how much do you trust people in general?” and of their trust in other features of American life relating to healthcare, finance, and media. These questions were only asked for a single year. Although this is an issue, because the panel structure of the HRS allows me to estimate the persistent component of returns, which is an important object in this literature. That said, in the pooled setting, we can assume that trust, like education, is fixed over time.⁴

I turn my attention to the correlations between the trust variables in figure . There is significant correlation between the trust measures.

Table 7: Trust variables correlation matrix

	General trust	Social Security	Medicare	Banks	Financial advisors	Mutual funds	Insurance	Media
General trust	1.00	0.30	0.26	0.37	0.33	0.36	0.39	0.21
Social Security	0.30	1.00	0.84	0.45	0.33	0.26	0.43	0.29
Medicare	0.26	0.84	1.00	0.41	0.31	0.26	0.38	0.27
Banks	0.37	0.45	0.41	1.00	0.55	0.43	0.46	0.27
Financial advisors	0.33	0.33	0.31	0.55	1.00	0.63	0.50	0.26
Mutual funds	0.36	0.26	0.26	0.43	0.63	1.00	0.41	0.26
Insurance	0.39	0.43	0.38	0.46	0.50	0.41	1.00	0.32
Media	0.21	0.29	0.27	0.27	0.26	0.26	0.32	1.00

Pairwise correlations between trust items (2020).

From here, I perform a principal component analysis and store the first two components for use in the statistical analysis. The loadings on each of the trust measures is given in figure .

I am interested in the relationship between trust and economic performance as the literature is, however in my setting the measure of economic performance is the return to assets. If a hump-shaped relationship is reasonable for income, it is even more plausible for returns: there is an inherent (and possibly explicit) level of trust between borrower and lender when forming credit contracts.

That said, it is important to understand the nature of the trust measure in the HRS sample. I begin to do this by considering each trust measure conditional on race/ethnicity. There seems to be significant group variation in means.

⁴Literature on trust talks about how history of being cheated or treated fairly form individuals’ trust over time. If this is true and at some point, one learns enough and forms their trust level, then a sample with an overrepresentation of older household is a reasonable environment to assume constant trust levels.

Table 8: Trust PCA loadings (first two components)

Trust item	PC1	PC2
General trust	0.3002	0.1808
Social Security	0.3835	-0.5451
Medicare	0.3657	-0.5680
Banks	0.3887	0.1014
Financial advisors	0.3844	0.3757
Mutual funds	0.3480	0.4309
Insurance	0.3786	0.0971
Media	0.2565	-0.0317

Principal components on trust variables (2020).

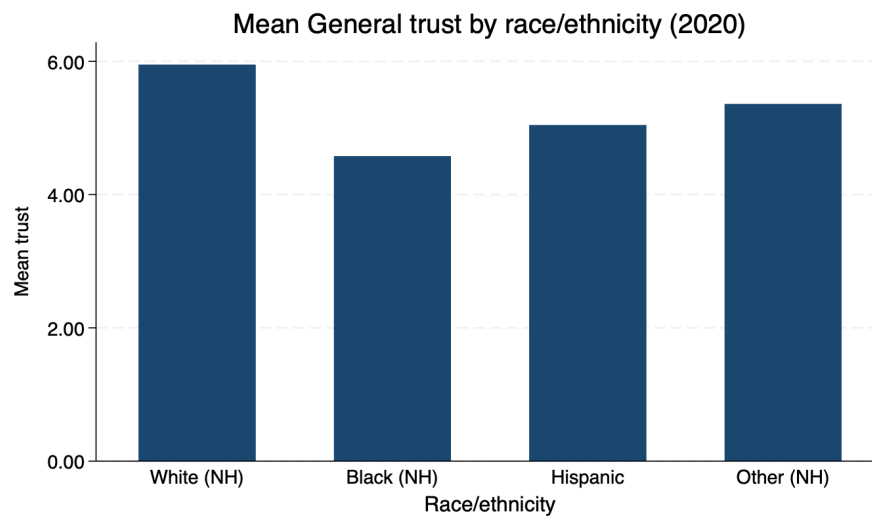


Figure 40: Mean general trust by race/ethnicity (2020)

3.6.1 Trust and income

As discussed before, the literature on trust and economic performance suggests there’s a “right amount of trust”. I want to see if this hump shaped relationship holds between the measures of trust and income in the HRS. After deflating and, more importantly, winsorizing labor income at the top and bottom 1%, the relationship seems to be hump shaped.

Income in the HRS is highly right-skewed and features a large mass at zero labor income (reflecting retirement and non-participation). In applied work, a common baseline is to use $\log(\text{income})$ because it reduces the influence of extreme values and yields coefficients interpretable as semi-elasticities. However, $\log(\cdot)$ is undefined at zero (and problematic with negative values when they arise from measurement or netting conventions), so using $\log(\text{income})$ in the HRS mechanically drops a large fraction of observations and can substantially change the visual and econometric relationship.

In our data, the corresponding scatter of $\log(\text{income})$ versus trust changes markedly because so many observations have exactly zero labor income. Using $\log(1 + \text{income})$ retains zeros but imposes an ad hoc curvature near zero and, in our case, does not restore the same qualitative picture.

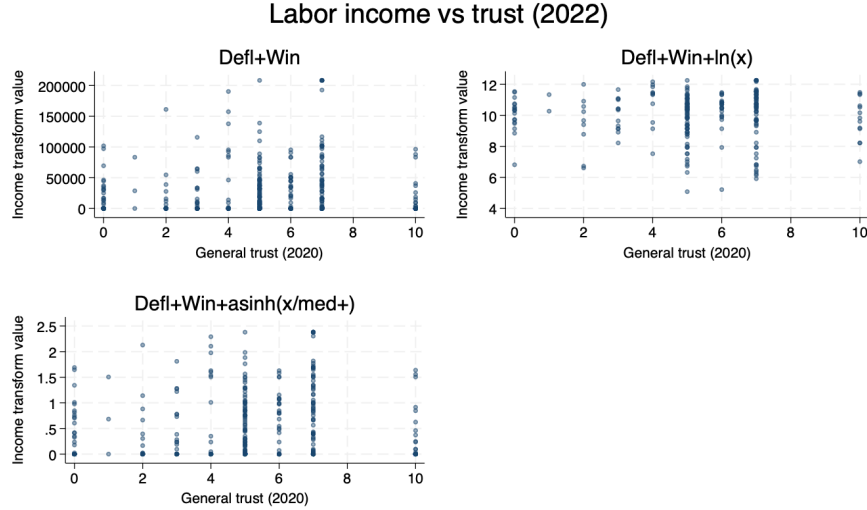


Figure 41: Labor income measures vs. trust (2022)

Motivated by this issue, we follow a large applied econometrics literature that uses the inverse hyperbolic sine (IHS, or asinh) transformation as a “log-like” mapping that is well-defined at zero and can accommodate negative values. The IHS transformation was advocated early on as an alternative to the log for heavy-tailed outcomes with extreme values and potential zeros/negatives (Burbidge, Magee, and Robb, “*Alternative Transformations to Handle Extreme Values of the Dependent Variable*”; MacKinnon and Magee, “*Transforming the Dependent Variable in Regression Models*”). More recently, Bellemare and Wichman (“*Elasticities and the Inverse Hyperbolic Sine Transformation*”) provide practical guidance on interpretation, emphasizing that $\text{asinh}(y)$ behaves similarly to $\log(y)$ for large y (since $\text{asinh}(y) \approx \log(2y)$ when y is large) while remaining defined at $y = 0$.

A related point is that the IHS can be sensitive to the units of measurement, which motivates a *scaled* IHS of the form $\text{asinh}(y/\kappa)$ (see discussions in Aihouton and Henningsen, “*Units of Measurement and the Inverse Hyperbolic Sine Transformation*”; Norton, “*The Inverse Hyperbolic Sine Transformation*”). In practice, κ is chosen as a meaningful scale (e.g., the median positive income), which makes the transformation more comparable across samples and improves interpretability near zero. Empirically, applying the (scaled) IHS to our deflated and winsorized income restores a scatter with trust that is visually well-behaved while retaining the economically important mass at zero labor income. This can be seen for the measure of total income as well in the following figure .

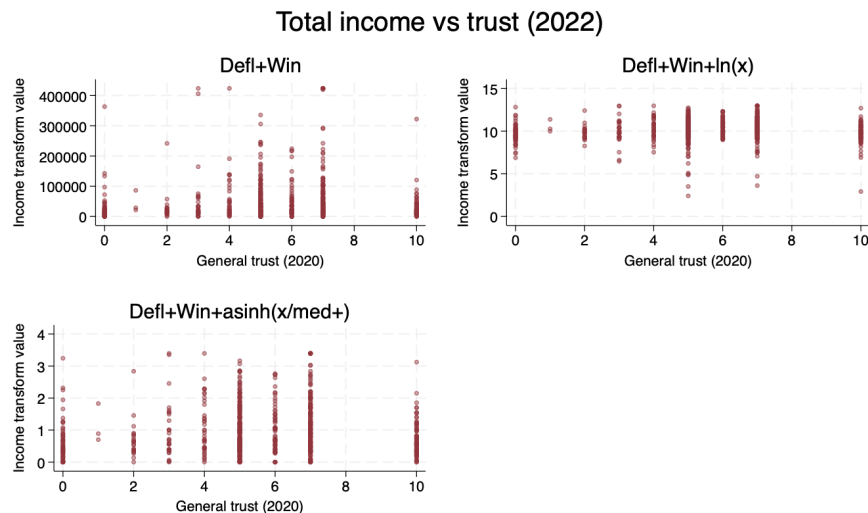


Figure 42: Total income measures vs. trust (2022)

3.6.2 Trust and returns

The scatterplots for trust and the measures of returns suggest stronger evidence for this hump shaped relationship. Especially for the smaller portfolio compositions (core, core and retirement), return values are highest in the middle and lower at the lowest and highest trust values.

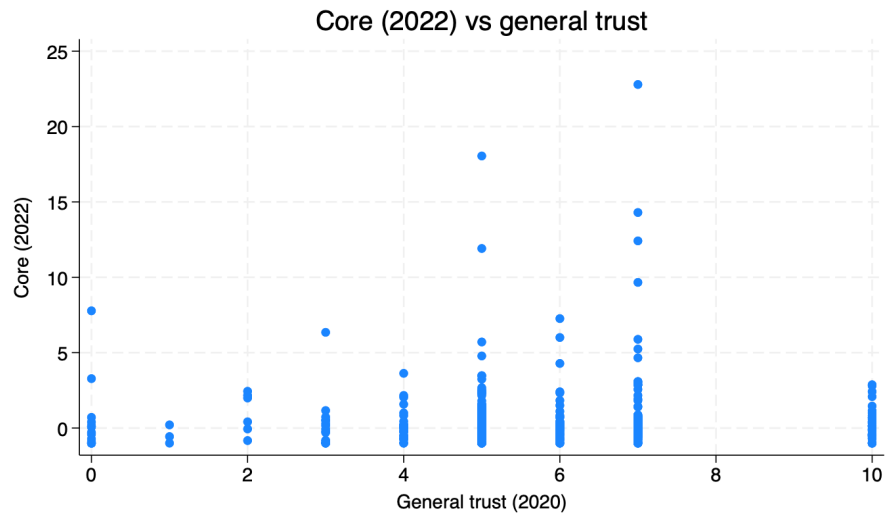


Figure 43: Returns to core assets vs. trust (2022)

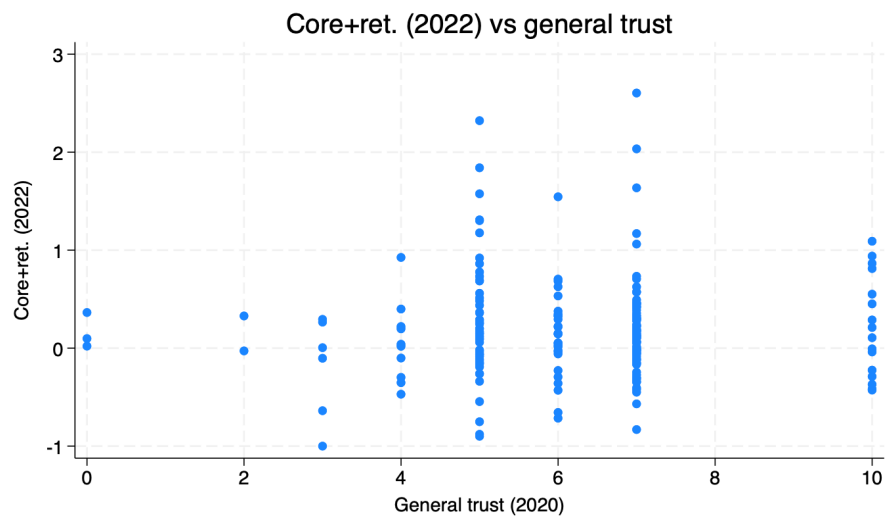


Figure 44: Returns to core and retirement assets vs. trust (2022)

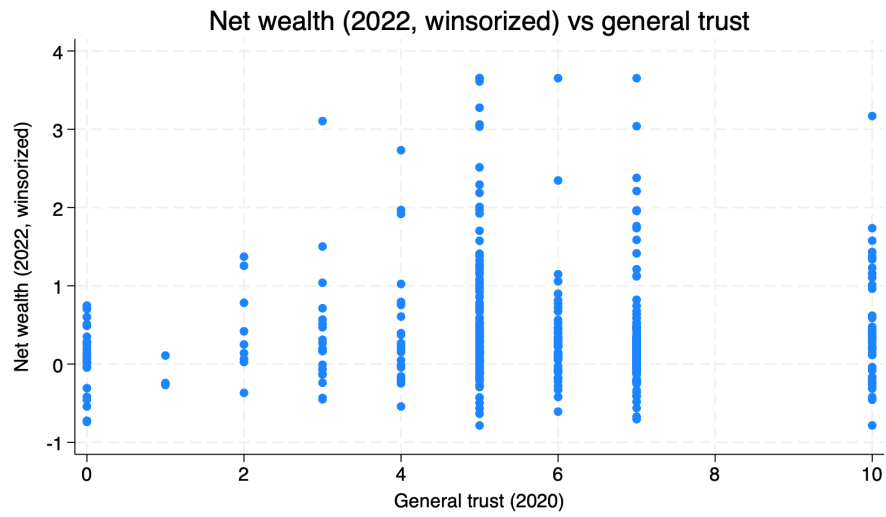


Figure 45: Returns to net wealth vs. trust (2022)

3.7 Demographics

As the HRS oversample older households, I take a quick look at demographic variables that will be important for the statistical analysis later.

Now, I look at the pairwise correlations between general and more specific control variables and the general trust measure in figure. This will be useful, as it will help understand what explains trust in the HRS data.

Table 9: Demographics: general controls (2020)

Variable	N	Mean	SD
Age	15723	68.089	10.849
Female	15723	0.594	0.491
Years of education	15651	12.961	3.241
Married	15685	0.537	0.499
Immigrant	15718	0.172	0.378
Born in U.S.	15718	0.828	0.378
Race: White (NH)	8865	0.565	0.496
Race: Black (NH)	3381	0.216	0.411
Race: Hispanic	2678	0.171	0.376
Race: Other (NH)	765	0.049	0.215
Working (in labor force)	15479	0.386	0.487

2020. Mean and SD; for dummies/categories mean = proportion (pct).

Table 10: Demographics: other controls (2020)

Variable	N	Mean	SD	p50
Depression	14998	1.55	2.04	1.00
Health conditions	15723	2.39	1.56	2.00
Medicare	15498	0.60	0.49	1.00
Medicaid	15383	0.14	0.35	0.00
Life insurance	15340	0.53	0.50	1.00
Times divorced	15723	0.57	0.79	0.00
Times widowed	15723	0.22	0.45	0.00

2020. Mean, SD, and median.

Table 11: Correlations of General trust with controls

Variable 1	Variable 2	Correlation
General trust	Age	0.2346
General trust	Years of education	0.0971
General trust	Female	-0.0035
General trust	Immigrant	-0.1161
General trust	Born in U.S.	0.1161
General trust	Race/ethnicity	-0.1140
General trust	Married	0.0970
General trust	Depression	-0.1673
General trust	Health conditions	-0.0144
General trust	Medicare	0.1438
General trust	Medicaid	-0.1791
General trust	Life insurance	0.0616
General trust	Times divorced	-0.0473
General trust	Times widowed	0.0266
General trust (var1) with each control.		

The literature on trust already documents a significant statistical relationship between trust and mental health. For that reason, I use the HRS RAND measure for a mental health index and show a clearly negative, linear relationship with trust.

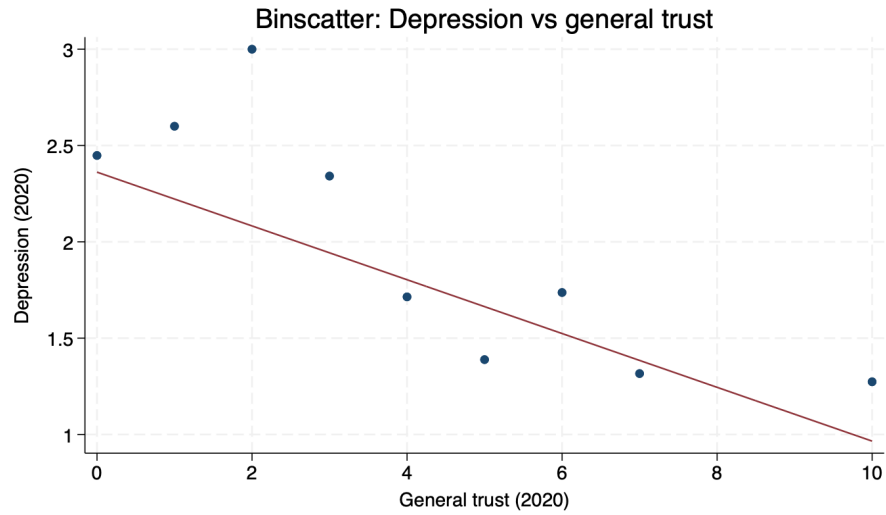


Figure 46: Depression vs. trust

3.8 Other controls

The remaining variables will be important for the statistical analysis in the next section.

3.8.1 Financial literacy

Financial literacy can be an important factor when trying to understand the persistent component of returns. The literature on measuring financial literacy using survey data has agreed up three questions regarding i) interest, ii) inflation, and iii) risk diversification to measure financial literacy for respondents.

Table 12: Financial literacy summary (2020)

Variable	N	Mean	SD	p50
Interest	1302	2.62	0.68	3.00
Inflation	1273	2.66	0.65	3.00
Risk	1072	4.02	1.72	5.00

HRS 2020: rv565 (interest), rv566 (inflation), rv567 (risk diversification).

Figure shows the cross correlations between the financial literacy measures and the general measure of trust. There is little correlation between trust and any of the financial literacy variables. More worrisome is that there is little correlation amongst the financial literacy variable. If they are all measuring the same thing, there should be some correlation.

Table 13: Financial literacy and trust correlations

	Interest	Inflation	Risk diversification	General trust
Interest	1.00	-0.01	0.03	0.01
Inflation	-0.01	1.00	0.07	-0.03
Risk diversification	0.03	0.07	1.00	0.04
General trust	0.01	-0.03	0.04	1.00

Interest, inflation, risk diversification, and general trust (2020).

3.8.2 Instrumental variables

My first attempt at identification of causal effects of trust on returns relied on literature in this direction. Specifically, inherited trust has been used as an instrument for trust before. Though I don't have a measure of inherited trust, I include some variables about the respondents' parents that is available in 2020 as an alternative. The correlations between those variables and general trust is in the figure .

Table 14: IV and trust correlations

	Parent citizenship	Parent loyalty	Population size	General trust
Parent citizenship	1.00	0.42	0.03	0.11
Parent loyalty	0.42	1.00	0.03	0.08
Population size	0.03	0.03	1.00	-0.10
General trust	0.11	0.08	-0.10	1.00
Parent citizenship, loyalty, population size, and general trust (2020).				

There was also a question in 2020 asking individuals “how large is the population of the city, village, or town where you currently live”. I plan to use this, along with the information available on what region respondents live in, to construct average trust in the location individuals live. This neighborhood trust effect may serve as a potential instrument as well.

Here is the figure of counts by region in each year.

Table 15: Observations by region and year

Year	Northeast	Midwest	South	West
2000	3300	4821	8052	3365
2002	2969	4525	7448	3180
2004	3247	4998	8008	3821
2006	2870	4628	7434	3468
2008	2637	4274	7006	3230
2010	3378	4920	9109	4557
2012	3111	4564	8528	4275
2014	2829	4100	7834	3897
2016	2992	4271	9105	4463
2018	2396	3471	7527	3673
2020	2171	3153	6853	3485
2022	1728	2602	5544	2949

Person-year observations by region (region 5 = Other omitted).

The overlap of region and population is not too sparse.

Table 16: Bin counts by region–population (2020)

Region	Population size	Obs
Northeast	Less than 1,000	3
Northeast	1,000 to 10,000	17
Northeast	10,000 to 50,000	37
Northeast	50,000 to 100,000	11
Northeast	100,000 to 1 million	20
Northeast	Greater than 1 million	20
Midwest	Less than 1,000	13
Midwest	1,000 to 10,000	27
Midwest	10,000 to 50,000	49
Midwest	50,000 to 100,000	19
Midwest	100,000 to 1 million	26
Midwest	Greater than 1 million	4
South	Less than 1,000	19
South	1,000 to 10,000	58
South	10,000 to 50,000	83
South	50,000 to 100,000	55
South	100,000 to 1 million	67
South	Greater than 1 million	34
West	Less than 1,000	6
West	1,000 to 10,000	20
West	10,000 to 50,000	33
West	50,000 to 100,000	22
West	100,000 to 1 million	50
West	Greater than 1 million	29

Sample: 2020, nonmissing general trust, region, and population.

That said, collapsing population into three possible sizes and regrouping makes the bins even more dense. This can be seen in the figure .

Table 17: Bin counts by region–population (3 bins, 2020)

Region	Population	Obs
Northeast	Small town (<10k)	20
Northeast	Small/med city (10k-100k)	48
Northeast	Large metro (100k+)	40
Midwest	Small town (<10k)	40
Midwest	Small/med city (10k-100k)	68
Midwest	Large metro (100k+)	30
South	Small town (<10k)	77
South	Small/med city (10k-100k)	138
South	Large metro (100k+)	101
West	Small town (<10k)	26
West	Small/med city (10k-100k)	55
West	Large metro (100k+)	79

I consider the group means in each of these respective scenarios in the following figures.

Table 18: Mean trust by region (2020)

Region (code)	Region	Mean trust	Obs
1	Northeast	5.1517	145
2	Midwest	5.7296	159
3	South	5.3333	399
4	West	5.4541	196
General trust (2020), nonmissing region.			

Table 19: Mean trust by population size (2020)

Population (code)	Population size	Mean trust	Obs
1	Less than 1,000	5.3659	41
2	1,000 to 10,000	5.5902	122
3	10,000 to 50,000	5.5545	202
4	50,000 to 100,000	5.1028	107
5	100,000 to 1 million	5.2805	164
6	Greater than 1 million	5.4138	87
General trust (2020), nonmissing population size.			

Table 20: Mean trust by population (3 bins, 2020)

Pop3 (code)	Population	Mean trust	Obs
1	Small town (<10k)	5.5337	163
2	Small/med city (10k-100k)	5.3981	309
3	Large metro (100k+)	5.3267	251

Small/med/large; general trust (2020).

Table 21: Mean trust by region–population (3 bins, 2020)

Region (code)	Region	Pop3 (code)	Population	Mean trust	Obs
1	Northeast	1	Small town (<10k)	5.0000	20
1	Northeast	2	Small/med city (10k-100k)	5.5417	48
1	Northeast	3	Large metro (100k+)	5.4500	40
2	Midwest	1	Small town (<10k)	5.8750	40
2	Midwest	2	Small/med city (10k-100k)	5.7647	68
2	Midwest	3	Large metro (100k+)	5.2000	30
3	South	1	Small town (<10k)	5.4156	77
3	South	2	Small/med city (10k-100k)	5.1159	138
3	South	3	Large metro (100k+)	5.3465	101
4	West	1	Small town (<10k)	5.7692	26
4	West	2	Small/med city (10k-100k)	5.5273	55
4	West	3	Large metro (100k+)	5.2911	79

General trust (2020) by region \times population (3 bins).

4 Results

4.1 Determinants of trust

First, I go beyond correlations between the variables by trying to consider which variables explain the variation in the trust measures.

Table 22: General trust (2020) on controls

	Demographics	Full controls
Age (5-yr bin)=25	0.00 (.)	0.00 (.)
Female	0.19 (0.17)	0.27 (0.18)
Years of education	0.05* (0.03)	0.04 (0.03)
Married	0.50*** (0.18)	0.17 (0.19)
NH Black	-0.99*** (0.22)	-1.07*** (0.23)
Hispanic	-0.38 (0.27)	-0.34 (0.27)
NH Other	-0.19 (0.32)	-0.19 (0.33)
Depression		-0.14*** (0.05)
Health conditions		-0.05 (0.07)
Covered by Medicare		0.15 (0.30)
Covered by Medicaid		-0.58** (0.28)
Has life insurance		0.25 (0.18)
Number of reported divorces		-0.16 (0.11)
Number of reported times being widowed		0.00 (0.25)
Constant	2.04 (1.49)	2.90 (1.84)
Observations	894.00	875.00
Adj. R-squared	0.09	0.12

Standard errors in parentheses

Robust standard errors in parentheses. Trust and controls from 2020. Age bins (5-yr) included; coefficients omitted.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2 Income and trust

After understanding the determinants of trust in the HRS data, I wanted to see if the “hump-shape” relationship between trust and income found by **jbpglg2016** was present in the HRS data. To do this I used the available income data from the 2020 wave with the trust measures from the same year.

First, I consider the statistical relationship for labor income in the following figures.

Table 23: Labor income (2020) on General trust (2020)

	1	2	3	4
General trust	0.02 (0.03)	0.19** (0.08)	0.01 (0.03)	0.07 (0.08)
General trust \times General trust		-0.02** (0.01)		-0.01 (0.01)
Age (5-yr bin)=25			0.00 (.)	0.00 (.)
Female			-0.27** (0.13)	-0.26* (0.13)
Years of education			0.11*** (0.02)	0.10*** (0.02)
Married			0.32** (0.13)	0.31** (0.13)
Born in U.S.			-0.11 (0.18)	-0.11 (0.18)
In labor force			0.67*** (0.22)	0.65*** (0.22)
NH Black			-0.10 (0.15)	-0.09 (0.15)
Hispanic			-0.21 (0.19)	-0.19 (0.19)
NH Other			-0.30 (0.23)	-0.31 (0.23)
Constant	10.23*** (0.17)	9.92*** (0.22)	7.75*** (0.49)	7.59*** (0.53)
Observations	347.00	347.00	343.00	343.00
Adj. R-squared	-0.00	0.01	0.16	0.16

Standard errors in parentheses

Age bins (5-yr) included in columns 3–4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The pattern seems to be stronger for total income, based on the figure .

Table 24: Labor income (2020) on General trust (2020), scaled asinh

	1	2	3	4
General trust	0.00 (0.01)	0.07*** (0.02)	0.00 (0.01)	−0.01 (0.02)
General trust × General trust		−0.01*** (0.00)		0.00 (0.00)
Age (5-yr bin)=25			0.00 (.)	0.00 (.)
Female			−0.08** (0.03)	−0.08** (0.03)
Years of education			0.03*** (0.00)	0.03*** (0.00)
Married			0.07** (0.03)	0.08** (0.03)
Born in U.S.			0.02 (0.05)	0.02 (0.05)
In labor force			0.56*** (0.04)	0.57*** (0.04)
NH Black			−0.01 (0.04)	−0.01 (0.04)
Hispanic			−0.04 (0.05)	−0.04 (0.05)
NH Other			−0.03 (0.07)	−0.03 (0.07)
Constant	0.34*** (0.04)	0.22*** (0.05)	−0.40*** (0.13)	−0.40*** (0.12)
Observations	900.00	900.00	890.00	890.00
Adj. R-squared	−0.00	0.01	0.41	0.41

Standard errors in parentheses

Age bins (5-yr) included in columns 3–4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.3 Returns and trust

With the hump-shape relationship between earnings and trust present in the HRS, I wanted to see if a similar relationship held for returns.

The smallest portfolio composition is of core assets only. The regression results for raw and winsorized annual returns for 2022 are given in figures and .

Table 25: Total income (2020) on General trust (2020)

	1	2	3	4
General trust	0.03** (0.01)	0.29*** (0.04)	0.00 (0.01)	0.08** (0.04)
General trust \times General trust		-0.03*** (0.00)		-0.01** (0.00)
Age (5-yr bin)=25			0.00 (.)	0.00 (.)
Female			-0.40*** (0.07)	-0.39*** (0.07)
Years of education			0.10*** (0.01)	0.09*** (0.01)
Married			0.22*** (0.07)	0.21*** (0.07)
Born in U.S.			0.08 (0.11)	0.08 (0.11)
In labor force			0.75*** (0.08)	0.74*** (0.08)
NH Black			-0.33*** (0.08)	-0.31*** (0.08)
Hispanic			-0.28*** (0.10)	-0.28*** (0.11)
NH Other			-0.35** (0.14)	-0.35** (0.14)
Constant	10.12*** (0.09)	9.63*** (0.11)	8.09*** (0.24)	7.95*** (0.25)
Observations	856.00	856.00	848.00	848.00
Adj. R-squared	0.00	0.05	0.33	0.33

Standard errors in parentheses

Age bins (5-yr) included in columns 3–4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 26: Total income (2020) on General trust (2020), scaled asinh

	1	2	3	4
General trust	0.03*** (0.01)	0.19*** (0.02)	0.00 (0.01)	0.05** (0.02)
General trust \times General trust		-0.02*** (0.00)		-0.00** (0.00)
Age (5-yr bin)=25			0.00 (.)	0.00 (.)
Female			-0.27*** (0.04)	-0.27*** (0.04)
Years of education			0.07*** (0.01)	0.07*** (0.01)
Married			0.14*** (0.04)	0.14*** (0.04)
Born in U.S.			0.02 (0.07)	0.02 (0.07)
In labor force			0.50*** (0.05)	0.49*** (0.05)
NH Black			-0.24*** (0.05)	-0.23*** (0.05)
Hispanic			-0.24*** (0.07)	-0.23*** (0.07)
NH Other			-0.27*** (0.09)	-0.27*** (0.09)
Constant	0.83*** (0.05)	0.53*** (0.06)	-0.48*** (0.14)	-0.51*** (0.16)
Observations	900.00	900.00	890.00	890.00
Adj. R-squared	0.01	0.05	0.35	0.35

Standard errors in parentheses

Age bins (5-yr) included in columns 3–4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 27: 2022 Core return (2022) on General trust (2020) (winsorized)

	1	2	3	4
General trust	0.02 (0.03)	0.17 (0.11)	0.01 (0.04)	0.14 (0.12)
(General) ²		-0.01 (0.01)		-0.01 (0.01)
Age (5-yr bin)=40			0.00 (.)	0.00 (.)
Female			0.07 (0.14)	0.10 (0.15)
Years of education			0.10*** (0.03)	0.09*** (0.03)
Married			0.39** (0.17)	0.39** (0.17)
Born in U.S.			0.11 (0.21)	0.11 (0.21)
NH Black			-0.42** (0.19)	-0.40** (0.19)
Hispanic			-0.42 (0.30)	-0.41 (0.30)
NH Other			-0.25 (0.37)	-0.24 (0.37)
In labor force			0.20 (0.17)	0.20 (0.17)
_cons	0.34 (0.21)	0.01 (0.35)	-0.04 (0.78)	-0.32 (0.81)
Observations	442.00	442.00	438.00	438.00
Adj. R-squared	-0.00	0.00	0.09	0.09

Standard errors in parentheses

Robust standard errors in parentheses. Age bins (5-yr) and wealth deciles included in columns 3–4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 28: 2022 Core return (2022) on General trust (2020) (raw)

	1	2	3	4
General trust	0.03 (0.04)	0.22* (0.12)	−0.01 (0.04)	0.18 (0.13)
(General) ²		−0.02* (0.01)		−0.02 (0.01)
Age (5-yr bin)=40			0.00 (.)	0.00 (.)
Female			0.23 (0.21)	0.27 (0.21)
Years of education			0.16*** (0.05)	0.15*** (0.05)
Married			0.55** (0.24)	0.55** (0.24)
Born in U.S.			0.18 (0.27)	0.18 (0.27)
NH Black			−0.50* (0.27)	−0.48* (0.27)
Hispanic			−0.32 (0.41)	−0.31 (0.41)
NH Other			−0.16 (0.52)	−0.16 (0.52)
In labor force			0.30 (0.22)	0.31 (0.22)
_cons	0.38 (0.23)	−0.06 (0.37)	−1.76 (1.48)	−2.16 (1.58)
Observations	442.00	442.00	438.00	438.00
Adj. R-squared	−0.00	0.00	0.09	0.10

Standard errors in parentheses

Robust standard errors in parentheses. Age bins (5-yr) and wealth deciles included in columns 3–4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The next portfolio composition includes retirement assets along with those core assets. The regression results for raw and winsorized annual returns for 2022 on these portfolios are given in figures and .

Table 29: 2022 Core+res return (2022) on General trust (2020) (winsorized)

	1	2	3	4
General trust	0.01 (0.02)	0.04 (0.05)	0.03 (0.02)	0.04 (0.05)
(General) ²		-0.00 (0.00)		-0.00 (0.00)
Age (5-yr bin)=50			0.00 (.)	0.00 (.)
Female			-0.11 (0.07)	-0.11 (0.07)
Years of education			0.03 (0.02)	0.03 (0.02)
Married			0.06 (0.09)	0.06 (0.09)
Born in U.S.			0.14 (0.08)	0.14 (0.08)
NH Black			-0.11 (0.09)	-0.11 (0.09)
Hispanic			-0.40** (0.20)	-0.40** (0.20)
NH Other			-0.03 (0.13)	-0.03 (0.13)
In labor force			0.03 (0.07)	0.03 (0.07)
_cons	0.10 (0.12)	0.03 (0.14)	-0.28 (0.43)	-0.30 (0.44)
Observations	210.00	210.00	209.00	209.00
Adj. R-squared	-0.00	-0.01	0.08	0.07

Standard errors in parentheses

Robust standard errors in parentheses. Age bins (5-yr) and wealth deciles included in columns 3–4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 30: 2022 Core+res return (2022) on General trust (2020) (raw)

	1	2	3	4
General trust	0.02 (0.02)	0.04 (0.05)	0.03 (0.02)	0.04 (0.05)
(General) ²		-0.00 (0.00)		-0.00 (0.00)
Age (5-yr bin)=50			0.00 (.)	0.00 (.)
Female			-0.11 (0.07)	-0.11 (0.07)
Years of education			0.03 (0.02)	0.03 (0.02)
Married			0.06 (0.09)	0.06 (0.09)
Born in U.S.			0.14 (0.08)	0.14 (0.08)
NH Black			-0.11 (0.09)	-0.11 (0.09)
Hispanic			-0.39** (0.20)	-0.40* (0.20)
NH Other			-0.03 (0.13)	-0.03 (0.13)
In labor force			0.04 (0.07)	0.03 (0.07)
_cons	0.09 (0.12)	0.02 (0.15)	-0.28 (0.43)	-0.30 (0.44)
Observations	210.00	210.00	209.00	209.00
Adj. R-squared	-0.00	-0.01	0.07	0.07

Standard errors in parentheses

Robust standard errors in parentheses. Age bins (5-yr) and wealth deciles included in columns 3–4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The largest portfolio composition is on net wealth, which includes the previous assets, residential assets, along with other smaller categories (and debt variables). The regression results for raw and winsorized annual returns to net wealth for 2022 are given in figures and .

Table 31: 2022 Net wealth return (2022) on General trust (2020) (winsorized)

	1	2	3	4
General trust	0.01 (0.01)	0.08*** (0.03)	0.01 (0.01)	0.05 (0.04)
(General) ²		-0.01** (0.00)		-0.00 (0.00)
Age (5-yr bin)=35			0.00 (.)	0.00 (.)
Female			-0.09 (0.06)	-0.08 (0.06)
Years of education			0.03** (0.01)	0.03** (0.01)
Married			0.08 (0.07)	0.08 (0.07)
Born in U.S.			0.06 (0.07)	0.06 (0.07)
NH Black			-0.15 (0.09)	-0.14 (0.09)
Hispanic			-0.08 (0.12)	-0.08 (0.12)
NH Other			-0.09 (0.13)	-0.08 (0.14)
In labor force			0.15** (0.07)	0.15** (0.07)
_cons	0.30*** (0.07)	0.15** (0.08)	-0.43 (0.35)	-0.40 (0.35)
Observations	496.00	496.00	406.00	406.00
Adj. R-squared	-0.00	0.00	0.24	0.24

Standard errors in parentheses

Robust standard errors in parentheses. Age bins (5-yr) and wealth deciles included in columns 3–4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 32: 2022 Net wealth return (2022) on General trust (2020) (raw)

	1	2	3	4
General trust	0.01 (0.01)	0.09*** (0.03)	0.01 (0.01)	0.06 (0.04)
(General) ²		-0.01** (0.00)		-0.01 (0.00)
Age (5-yr bin)=35			0.00 (.)	0.00 (.)
Female			-0.07 (0.07)	-0.07 (0.07)
Years of education			0.04** (0.02)	0.03** (0.02)
Married			0.09 (0.08)	0.10 (0.08)
Born in U.S.			0.07 (0.08)	0.07 (0.08)
NH Black			-0.15 (0.10)	-0.14 (0.10)
Hispanic			-0.07 (0.13)	-0.07 (0.13)
NH Other			-0.10 (0.14)	-0.09 (0.15)
In labor force			0.15** (0.07)	0.15** (0.07)
_cons	0.31*** (0.07)	0.14* (0.08)	-0.47 (0.44)	-0.43 (0.43)
Observations	496.00	496.00	406.00	406.00
Adj. R-squared	-0.00	0.00	0.22	0.22

Standard errors in parentheses

Robust standard errors in parentheses. Age bins (5-yr) and wealth deciles included in columns 3–4.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.4 Returns from 2002-2022

Next, I wanted to compute returns across several years and attempt to describe the persistent component of returns that **aflgdmlp20** point to as heterogeneity across individuals. I also use three similar specifications. First, a pooled OLS regression with a baseline set of controls. Second, a similar pooled regression with aimed at controlling for risk exposure. Third, a panel regression with individual fixed effects and year dummies.

5 The pooled regression with constant trust

After looking at those specifications for the returns regressions, I wanted to see if the trust measure was significant in the pooled setting.

5.1 Trust and average income

5.2 Trust and average returns

5.3 Estimated fixed effects for returns

Appendix

.1 Trust and race

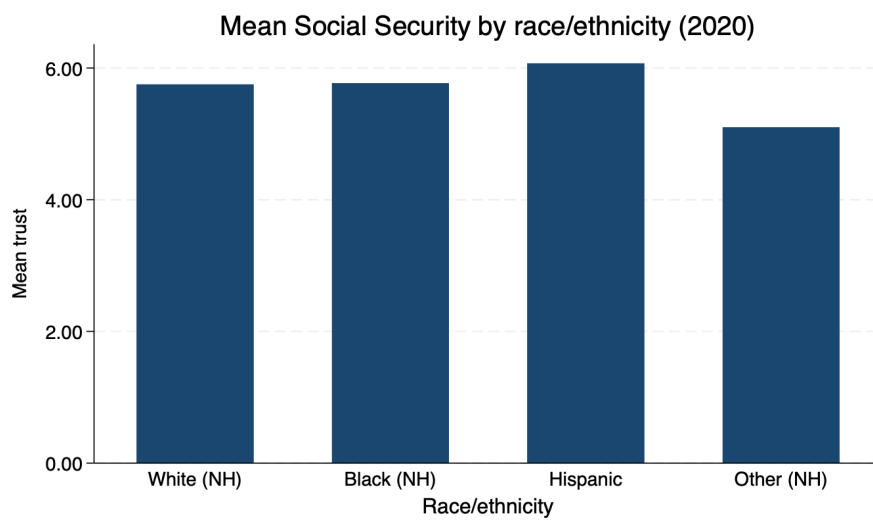


Figure 47: Mean Social Security trust by race/ethnicity (2020)

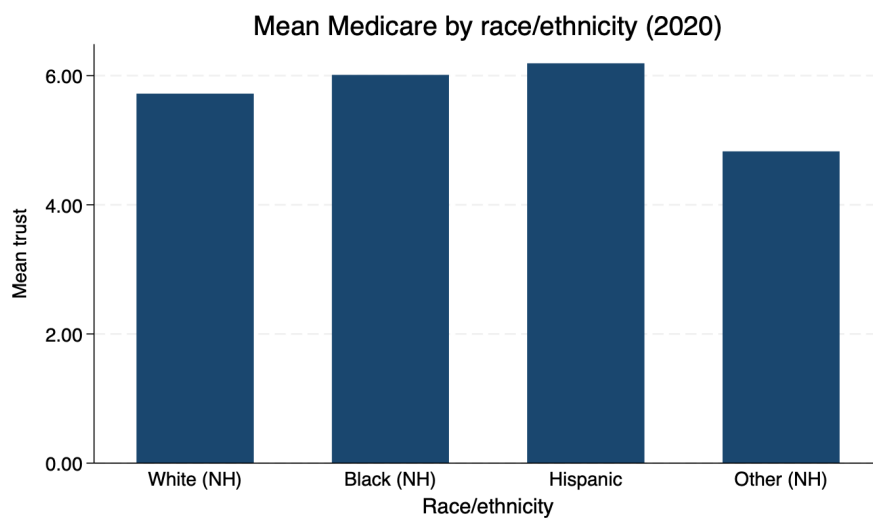


Figure 48: Mean Medicare trust by race/ethnicity (2020)

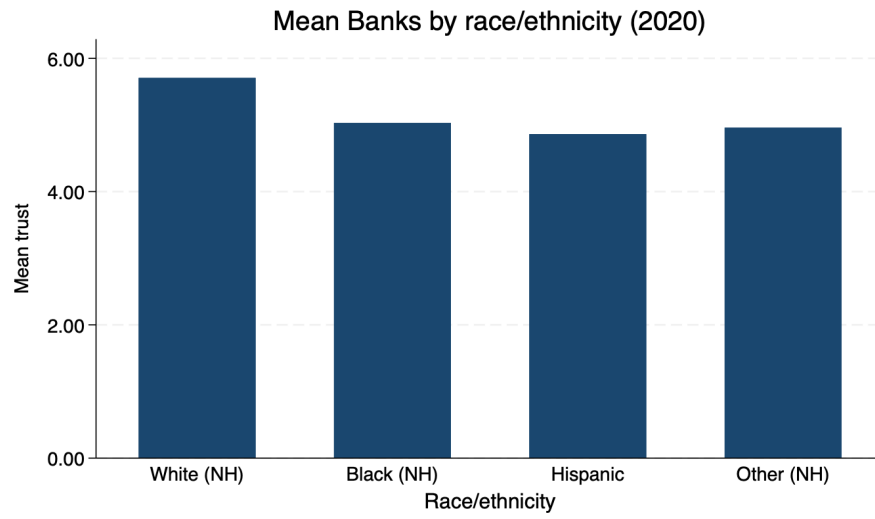


Figure 49: Mean banks trust by race/ethnicity (2020)

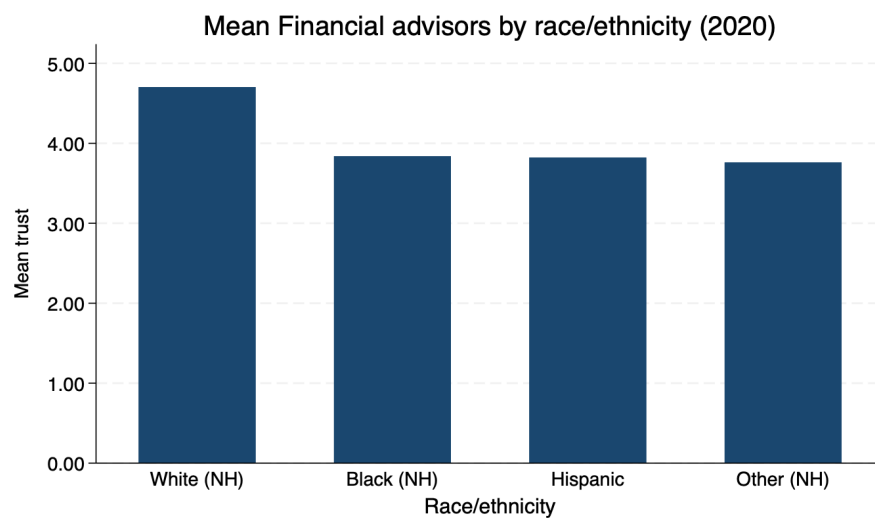


Figure 50: Mean financial advisors trust by race/ethnicity (2020)

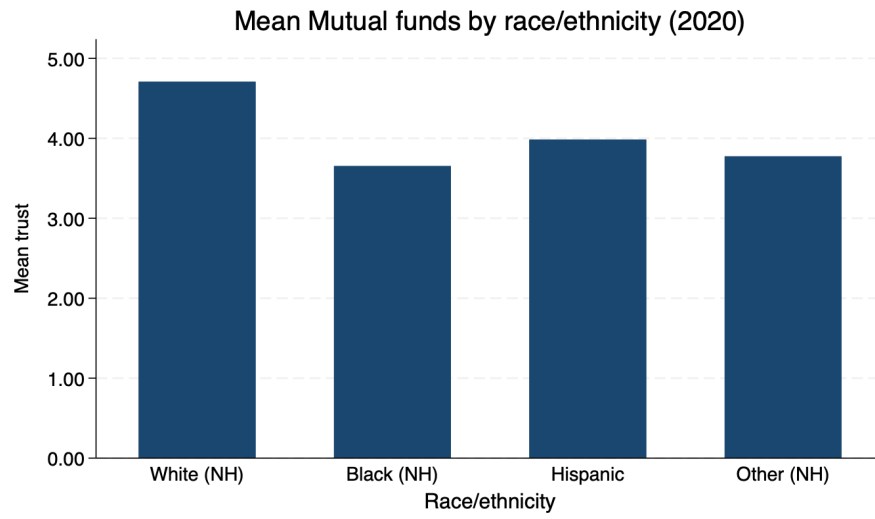


Figure 51: Mean mutual funds trust by race/ethnicity (2020)

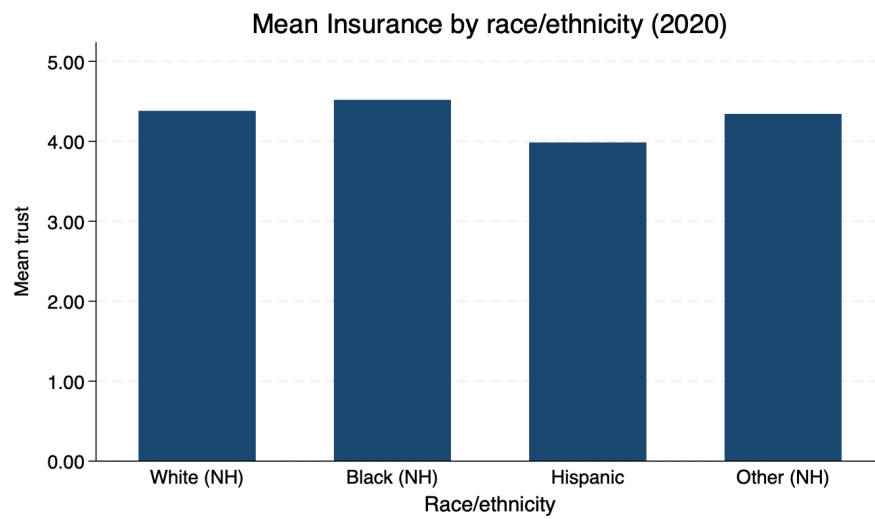


Figure 52: Mean insurance trust by race/ethnicity (2020)

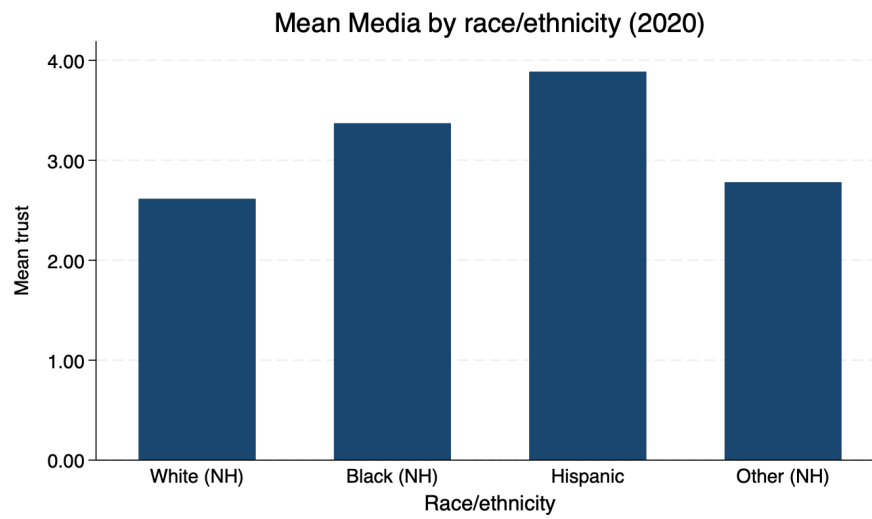


Figure 53: Mean media trust by race/ethnicity (2020)

.2 Other trust measures

.3 Cross section: Income and Trust

.4 Cross section: Returns and Trust