Wealth Proxies in the PSID

A Wealth Series from 1969

Christophe Van Langenhove

September 2026

1. Introduction

1. Introduction

The PSID runs from 1968 and records household wealth only from 1984. Two published wealth proxies predict wealth in the pre-1984 waves.

- **PK-proxy.** Pfeffer and Killewald (2018): the housing value, renters set to zero, one of the sixteen characteristics the PSID records in every wave since 1969
- **FJ-proxy.** Fisher and Johnson (2023): the housing value and interest and dividend income, two of the sixteen characteristics
- Both are compared against actual wealth on one performance metric, the parent-child rank-rank slope: the child's wealth rank regressed on the parent's

1. Introduction

I build the GB-proxy with gradient boosting from all sixteen characteristics, and compare the three wealth proxies against actual wealth on five dimensions.

1. **The overall wealth rank distribution.** The wealth rank error falls from 21 and 19 rank units to **14**

1. Introduction

I build the GB-proxy with gradient boosting from all sixteen characteristics, and compare the three wealth proxies against actual wealth on five dimensions.

1. **The overall wealth rank distribution.** The wealth rank error falls from 21 and 19 rank units to **14**
2. **Intragenerational wealth mobility.** The GB-proxy overstates wealth persistence, the one dimension where the published wealth proxies come closer

1. Introduction

I build the GB-proxy with gradient boosting from all sixteen characteristics, and compare the three wealth proxies against actual wealth on five dimensions.

1. **The overall wealth rank distribution.** The wealth rank error falls from 21 and 19 rank units to **14**
2. **Intragenerational wealth mobility.** The GB-proxy overstates wealth persistence, the one dimension where the published wealth proxies come closer
3. **Intergenerational wealth mobility.** Its parent-child slope is indistinguishable from actual wealth's

1. Introduction

I build the GB-proxy with gradient boosting from all sixteen characteristics, and compare the three wealth proxies against actual wealth on five dimensions.

1. **The overall wealth rank distribution.** The wealth rank error falls from 21 and 19 rank units to **14**
2. **Intragenerational wealth mobility.** The GB-proxy overstates wealth persistence, the one dimension where the published wealth proxies come closer
3. **Intergenerational wealth mobility.** Its parent-child slope is indistinguishable from actual wealth's
4. **Wealth inequality.** The Gini distance falls from 0.16 and 0.17 to **0.06**

1. Introduction

I build the GB-proxy with gradient boosting from all sixteen characteristics, and compare the three wealth proxies against actual wealth on five dimensions.

1. **The overall wealth rank distribution.** The wealth rank error falls from 21 and 19 rank units to **14**
2. **Intragenerational wealth mobility.** The GB-proxy overstates wealth persistence, the one dimension where the published wealth proxies come closer
3. **Intergenerational wealth mobility.** Its parent-child slope is indistinguishable from actual wealth's
4. **Wealth inequality.** The Gini distance falls from 0.16 and 0.17 to **0.06**
5. **Wealth accumulation trajectories.** The GB-proxy comes closest, and overstates

1. Introduction

I build the GB-proxy with gradient boosting from all sixteen characteristics, and compare the three wealth proxies against actual wealth on five dimensions.

1. **The overall wealth rank distribution.** The wealth rank error falls from 21 and 19 rank units to **14**
2. **Intragenerational wealth mobility.** The GB-proxy overstates wealth persistence, the one dimension where the published wealth proxies come closer
3. **Intergenerational wealth mobility.** Its parent-child slope is indistinguishable from actual wealth's
4. **Wealth inequality.** The Gini distance falls from 0.16 and 0.17 to **0.06**
5. **Wealth accumulation trajectories.** The GB-proxy comes closest, and overstates

The wider set of input variables, not the estimator, explains the outperformance. The GB-proxy gives a **1969 to 2023** wealth series with three biases of known size.

1. Introduction

Three strands of literature.

- **The wealth the PSID records.** Hurst, Luoh and Stafford (1998), Pfeffer et al. (2016), Cooper et al. (2019), Si et al. (2023)
- **Supervised learning for distributional estimates.** Bloise et al. (2021), Salas-Rojo and Rodríguez (2022), Cornwall and Gindelsky (2025). A low prediction error is not enough
- **Housing values as a wealth proxy outside the PSID.** Kalsi and Ward (2025), Binder et al. (2025). One asset, and renters assigned none

2. PSID sample and three wealth proxies

2. PSID sample and three wealth proxies

All three wealth proxies run on one PSID sample, so a comparison between them moves the wealth proxy and nothing else.

- **The 1984 to 2023 sample:** the 16 waves that record actual wealth, 67,665 household-years
- **Actual wealth:** the PSID's own wealth variable, assets less debts, vehicles excluded
- The Survey Research Center (SRC) sample, not survey-weighted
- The GB-proxy covers every wave: **173,036** household-years over 1969 to 2023, 48,859 of them before 1984

2. PSID sample and three wealth proxies

The PK-proxy, Pfeffer and Killewald (2018), is the housing value times a factor.

$$\hat{w}_{i,t}^{\mathrm{PK}} = \begin{cases} f_h \, h_{i,t} & \text{if } h_{i,t} > 0 \\ f_r \, r_{i,t} & \text{if } h_{i,t} = 0 \end{cases}$$

- $h_{i,t}$ is the gross value of the main residence and $r_{i,t}$ the rent. PK set f_h to one and f_r to zero: a homeowner's prediction is the housing value itself, a renter's is zero
- PK do not state how renters rank among themselves. I draw their order at random, and report renters tied as an alternative

2. PSID sample and three wealth proxies

The FJ-proxy, Fisher and Johnson (2023), is a linear regression through the origin on the housing value and capital income.

$$w_i(t) = \beta_h \, h_i(t) + \beta_y \, y_i(t) + \varepsilon_i(t)$$

- $y_{i,t}$ is capital income: family income less labour income, other members' taxable income and transfer income
- FJ report the two input variables but not the functional form. Fitted on the 1984 to 2023 sample, $\hat{\beta}_h$ is 2.15 and $\hat{\beta}_y$ is 5.9

2. PSID sample and three wealth proxies

The GB-proxy takes all sixteen characteristics as input variables, in three models: homeowners, renters, and the 5% reporting neither a housing value nor rent.

	Input variables
Demographics and income	Age, household size, family structure, labour income, capital income, business held, years unable to work
Bottom of the distribution	Food stamps received, income to needs, minimum wage ratio
Wealth accumulation and surplus	Waves spent owning, rooms per person, money given away
Homeowners only	Housing value, mortgage ratio
Renters only	Rent

2. PSID sample and three wealth proxies

Each model predicts a household's wealth relative to an aggregate of its wave rather than in dollars.

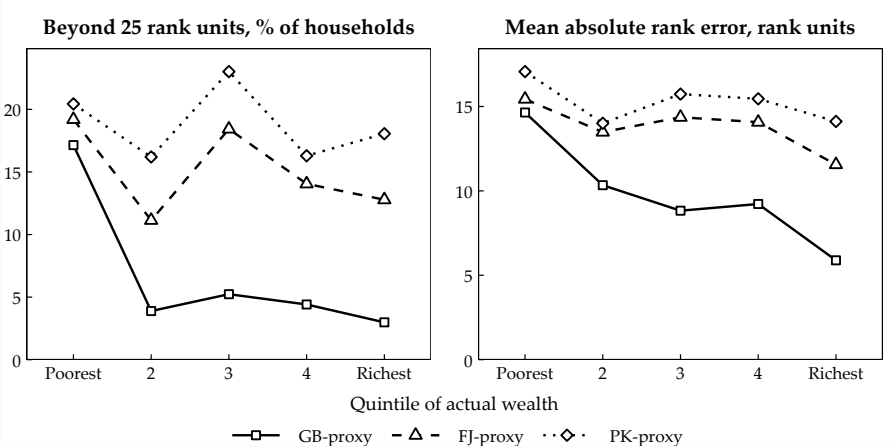
$$\tau_{i,t} = \begin{cases} \dfrac{w_{i,t}/h_{i,t}}{F(t)} & \text{if } h_{i,t} > 0 \\ \dfrac{w_{i,t}}{\bar{w}(t)} & \text{otherwise} \end{cases}$$

- $F(t)$ is aggregate household wealth over residential real estate, from the Financial Accounts, and $\bar{w}(t)$ mean wealth per household
- $F(t)$ runs 2.52 to 3.54 over 1984 to 2023 and 2.97 to 3.87 before 1984. Dividing by it removes the wave-level shift
- The GB-proxy is the **point prediction**, which varies less across households than actual wealth does

3. Three wealth proxies against actual wealth

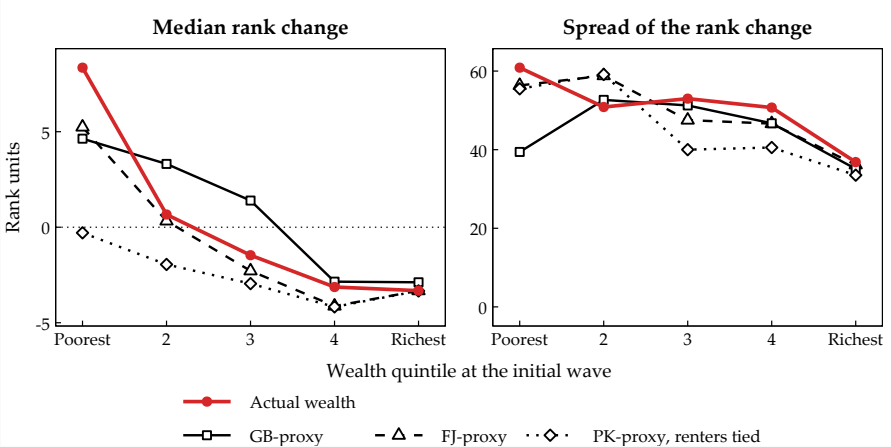
3. Three wealth proxies against actual wealth

The overall wealth rank distribution. The GB-proxy outperforms on all four performance metrics: a wealth rank error of 14 rank units against 21 and 19.



3. Three wealth proxies against actual wealth

Intragenerational wealth mobility. The GB-proxy overstates ten-year wealth persistence: a rank-rank slope of 0.76 against 0.70 in actual wealth, the largest gap.



3. Three wealth proxies against actual wealth

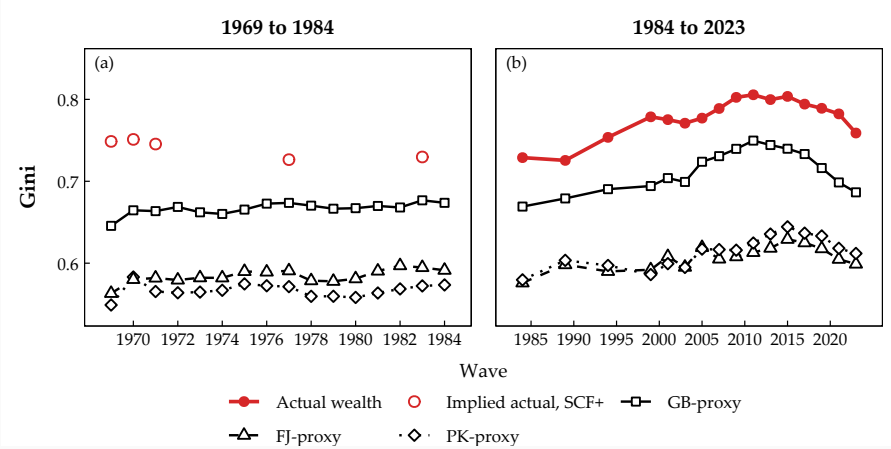
Intergenerational wealth mobility. With the parent ranked under the wealth proxy and the child under actual wealth, the GB-proxy's seven gaps are indistinguishable from zero, at a joint p-value of 0.62.

- Under actual wealth the parent-child slope rises from 0.34 at ages 30 to 34 to 0.49 at 60 to 64
- The PK-proxy and the FJ-proxy place it below at all seven lifecycle stages: joint p-values of 0.01 and 0.06
- **Grandparent-grandchild.** A wealth proxy adds pairs, about 1,300 against 1,000. The GB-proxy and the FJ-proxy reproduce the slope, the PK-proxy lowers it

The near-zero gap is a cancellation: -0.05 below the median parent rank, 0.07 above it.

3. Three wealth proxies against actual wealth

Wealth inequality. Every wealth proxy understates it, and the GB-proxy least: a Gini distance of 0.06 against 0.16 and 0.17.



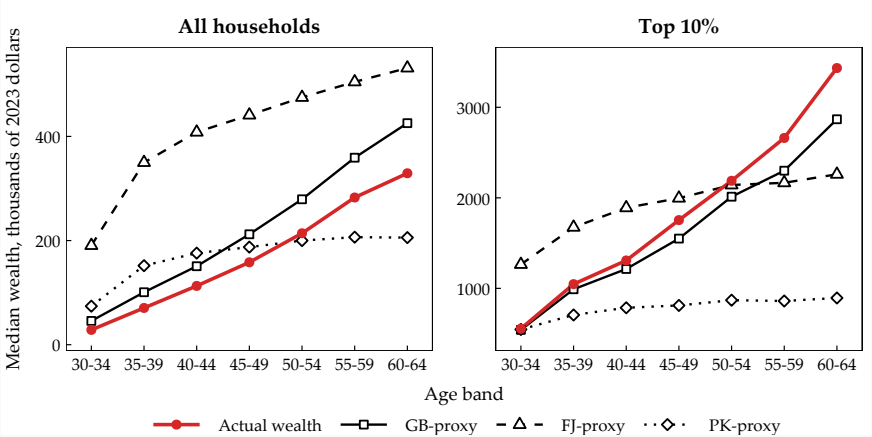
3. Three wealth proxies against actual wealth

The understatement of wealth inequality comes from the top of the wealth distribution.

- The GB-proxy misses **15%** of the top 10% wealth share, the PK-proxy 36% and the FJ-proxy 36%
- The missing wealth lands in the middle 50 to 90%: a share of 0.35 under actual wealth, 0.41 under the GB-proxy, 0.52 and 0.51 under the PK-proxy and the FJ-proxy
- The distance belongs to the wealth proxy, not the PSID. Rebuilt in the Survey of Consumer Finances, the PK-proxy leaves **0.17** against 0.16 here

3. Three wealth proxies against actual wealth

Wealth accumulation trajectories. The GB-proxy comes closest but overstates: median wealth rises 8.5 times from ages 30 to 34 to 60 to 64, 10.4 under the GB-proxy, 2.4 and 2.8 under the PK-proxy and the FJ-proxy.



4. The sources of the GB-proxy's performance

4. The sources of the GB-proxy's performance

Most of the GB-proxy's improvement comes from its set of input variables, not from its estimator: 84 to 97% on the four dimensions where it outperforms.

	Share of the improvement, %	
Dimension	Input variables	Estimator
Overall wealth rank distribution	93	7
Intergenerational wealth mobility	94	6
Wealth inequality	84	16
Wealth accumulation trajectories	97	3
Intragenerational wealth mobility	17	83

A Shapley split over four models, linear or boosted, on one input variable or on all, fitted out of fold. On intragenerational wealth mobility the estimator widens the gap, from 0.02 to 0.06.

4. The sources of the GB-proxy's performance

The GB-proxy's wealth rank error is largest where a thousand dollars moves a household the most rank units.

- One rank unit spans **$5 thousand** in the poorest wealth quintile against $158 thousand in the richest
- Renters' error is 14 units against homeowners' 9 because renters are poorer: **47%** sit in the poorest wealth quintile against 6% of homeowners
- Composition alone gives 10 units of the 5 observed. Inside the poorest wealth quintile renters rank better, 12 against 33
- The housing value and the mortgage ratio account for 43% of the gap

Out of fold, every household ranked against its whole wave, mean absolute error.

5. Conclusion

5. Conclusion

- The GB-proxy outperforms the PK-proxy and the FJ-proxy on four of the five dimensions: a wealth rank error of **14** against 21 and 19, a Gini distance of **0.06** against 0.16 and 0.17
- The outperformance comes from the set of input variables, 84 to 97% of the improvement, not from the estimator
- Three biases of known size: it overstates intragenerational wealth persistence, 0.76 against 0.70, its Gini lies 0.06 below the actual one, and its within-cohort rise lies above actual wealth's
- Beyond the PSID: where a dataset records income, household structure and need as well as the housing value, a wealth proxy that takes all of them ranks households better than one built on the housing value alone

Thank you.

References I

Binder, Ariel J., Max Risch, and John Voorheis. 2025. "Housing Capital and Intergenerational Mobility in the United States." U.S. Census Bureau, Center for Economic Studies Working Paper 25-55.

Bloise, Francesco, Paolo Brunori, and Patrizio Piraino. 2021. "Estimating Intergenerational Income Mobility on Sub-optimal Data: A Machine Learning Approach." *The Journal of Economic Inequality*, 19(4): 643–665.

Cooper, Daniel, Karen Dynan, and Hannah Rhodenhiser. 2019. "Measuring Household Wealth in the Panel Study of Income Dynamics: The Role of Retirement Assets." Federal Reserve Bank of Boston Research Department Working Paper 19-6.

References II

Cornwall, Gary, and Marina Gindelsky. 2025. "Nowcasting Distributional National Accounts for the United States: A Machine Learning Approach." *AEA Papers and Proceedings*, 115: 79–84.

Fisher, Jonathan, and David S. Johnson. 2023. "Intergenerational Mobility using Income, Consumption, and Wealth." Washington Center for Equitable Growth Working Paper.

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *The Annals of Statistics*, 29(5): 1189–1232.

Hurst, Erik, Ming Ching Luoh, and Frank P. Stafford. 1998. "The Wealth Dynamics of American Families, 1984–94." *Brookings Papers on Economic Activity*, 1998(1): 267–337.

Kalsi, Priti, and Zachary Ward. 2025. "The Gilded Age and Beyond: The Persistence of Elite Wealth in American History." National Bureau of Economic Research Working Paper 33355.

References III

Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins. 2020. "Income and Wealth Inequality in America, 1949–2016." *Journal of Political Economy*, 128(9): 3469–3519.

Pfeffer, Fabian T., and Alexandra Killewald. 2018. "Generations of Advantage. Multigenerational Correlations in Family Wealth." *Social Forces*, 96(4): 1411–1442.

Pfeffer, Fabian T., Robert F. Schoeni, Arthur Kennickell, and Patricia Andreski. 2016. "Measuring Wealth and Wealth Inequality: Comparing Two U.S. Surveys." *Journal of Economic and Social Measurement*, 41(2): 103–120.

Salas-Rojo, Pedro, and Juan Gabriel Rodríguez. 2022. "Inheritances and Wealth Inequality: A Machine Learning Approach." *The Journal of Economic Inequality*, 20(1): 27–51.

References IV

Si, Yajuan, Steve Heeringa, David Johnson, Roderick J. A. Little, Wenshuo Liu, Fabian Pfeffer, and Trivellore Raghunathan. 2023. "Multiple Imputation with Massive Data: An Application to the Panel Study of Income Dynamics." *Journal of Survey Statistics and Methodology*, 11(1): 260–283.

Van Langenhove, Christophe. 2026. "Wealth Proxies in the PSID: A Wealth Series from 1969."

Backup

Robustness of the rank result

The GB-proxy's wealth rank error survives every test but one.

- **Waves.** Dropping any one of the 16 waves leaves it between 13.6 and 13.9 units, first of the three whichever wave is dropped
- **Weights.** The PSID's longitudinal weight and a reweighting for attrition each move it by less than one unit
- **Rank correlation.** A plain rank correlation with actual wealth gives the same ordering as the four performance metrics
- **The exception.** On the low-income oversample alone the PK-proxy with renters tied ranks best. On the two samples pooled the GB-proxy is first again

Robustness on the other four dimensions

- **Intragenerational wealth mobility.** The overstatement survives six changes to the sample. Dropping each initial wave in turn keeps the GB-proxy's slope between 0.76 and 0.77, against 0.69 to 0.70 in actual wealth
- **Intergenerational wealth mobility.** A Holm correction leaves no gap distinguishable from zero. A bootstrap over the 1968 families places the GB-proxy's gap between -0.02 and 0.02, and a balanced panel leaves it at 0.01
- **Wealth inequality.** Four further measures, a concept without employer retirement balances, negative wealth kept and ages 25 to 64 give the same ordering. Dropping each wave in turn leaves the GB-proxy's Gini distance between 0.06 and 0.07
- **Wealth accumulation.** Two of seven tests reverse which wealth proxy comes closest: the PSID's longitudinal weight, under which the PK-proxy's ratio rises to 9.1 times against 9.6 in actual wealth, and a concept without employer retirement balances

The three renter orderings of the PK-proxy

PK do not state how renters rank among themselves. The paper draws their order at random and reports two alternatives: all renters tied, and renters ordered by rent.

- On the overall wealth rank distribution the wealth rank error is 21 rank units with renters at random, 20 tied and 20 by rent, against 14 for the GB-proxy. The PK-proxy beats the GB-proxy only with renters tied, only in the poorest wealth quintile, and by less than sampling error
- On intragenerational wealth mobility the PK-proxy's ten-year slope is 0.66 with renters at random and 0.71 with renters tied, against 0.70 in actual wealth. With renters at random every renter's rank is redrawn each wave, so the rank change is not movement
- The FJ-proxy's closer aggregate slope comes from errors that offset: 0.83 against 0.68 among households that own at both dates, and 0.34 against 0.47 among households that rent at both dates

The Shapley split behind the decomposition

- Four models, each built and tuned as the GB-proxy is and fitted out of fold on five folds: linear or boosted, on one input variable (the housing value or the rent) or on all of them: LA, LF, BA and BF. Model BF is the GB-proxy refitted
- $\mathcal{D}$ is a model's distance from actual wealth on a dimension's performance metric. Each source's value averages what it adds with the other absent and with the other present

$$\phi_{\text{input variables}} = \tfrac{1}{2}\left[(d_{LA} - d_{LF}) + (d_{BA} - d_{BF})\right]$$
$$\phi_{\text{estimator}} = \tfrac{1}{2}\left[(d_{LA} - d_{BA}) + (d_{LF} - d_{BF})\right]$$

- The housing value or rent alone leaves a wealth rank error of 20 units, all the input variables in a linear regression 16, the GB-proxy 15. The Gini distance falls from 0.16 to 0.09 to 0.06

The estimator

- A gradient-boosting model, Friedman (2001), one for homeowners, one for renters and one for households reporting neither a housing value nor rent. The homeowner model adds the housing value and the mortgage ratio, the renter model the rent
- Each target is trimmed at its 2.5th and 99.5th percentiles before fitting. Every result uses the point prediction. A robustness series adds a random draw of the unexplained part back, which restores the dispersion of actual wealth and comes closer on wealth inequality
- Out of fold, the wealth rank error is 15 units for homeowners, 32 for renters and 29 for households reporting neither, each ranked within its group
- For homeowners the mortgage ratio accounts for 27% of the summed absolute Shapley contributions, and with capital income and income to needs the three leading input variables for 53%. For renters income to needs accounts for 29%

What PK and FJ themselves report

- Neither published comparison isolates the wealth proxy's own error
- Pfeffer and Killewald (2018) take a rank slope of 0.37 in housing values against 0.39 in wealth as evidence that housing values serve as a wealth proxy. That closeness averages gaps of one sign at every lifecycle stage
- Fisher and Johnson (2023) report 0.30 on 395 offspring aged 30 to 32 and 0.21 on the 2,853 pairs formed from the waves that record wealth alone