Portfolio Choice in the Presence of Housing

(Cocco, 2005, Review of Financial Studies)

A Replication by Zachary Orlando *

Introduction

Why do so few individuals own stock early in life? If, as is shown in (Cocco et al., 2005), individual labor income is relatively uncorrelated with the stock market, then young individuals, who hold a great amount of future labor income, should invest as if they already are significantly exposed to safe assets (and hence allocate close to all of their investments into risky assets like stocks). Yet, only 26% of under 35 adults in the 1989 PSID reported holding stock at all.

"Portfolio Choice in the Presence of Housing" shows that individual's housing investments play an important role in explaining this puzzle. Housing risk crowds out stock market holdings. Moreover, as most housing purchases occur while young, this reduces the magnitude of liquid assets and depletes the incentive to invest in stock consistent with relatively low assumed "participation costs". Overall, stock market participation when factoring in housing, is predicted to rise over the lifecycle consistent with the data.

This replication has four parts:

- In Part I, I detail the lifecycle portfolio-choice model.
- In Part II, I (following the paper) use Panel Study of Income Dynamics (PSID) 2025 data to estimate the income and housing processes used as model inputs.
- In Part III, I then replicate the main exhibits of the paper.
- In Part IV, I extend the model to include a rent versus buy decision and non-renegotiable mortgages. I then investigate the impact of home affordability on portfolio choice.

I wanted to be up front on some results that I couldn't replicate:

- My simulation predicted far higher stock market participation rates than reported in the model, particularly early in life. This was robust across all the potential model versions I tried [1].
- I also found that house price risk, as it is parameterized to be perfectly correlated with persistent income risk, instead somewhat *crowds-in* stock market investment as a hedge.

The lack of available code and replications of this paper meant it was difficult to exactly pin down what generates differences between my results and the paper. I will aim to give my thoughts.

*UW Madison, zorlando@wisc.edu. Code is available on GitHub.

[1] A subset of reasonable parameterizations are detailed in Appendix A.

Part I: The Model

Consider the partial-equilibrium problem of an agent who receives uninsurable labor market earnings for T (10, every model-period represents five years) periods and must decide much to invest in the stock market versus bonds and additionally how big of a home to own.

The dynamic problem can be recursively represented as:

$$V_t(\text{State}_t) = \max_{C_t, H_t, D_t, FC_t, \alpha_t} \left[\frac{\left(C_t^{1-\theta} H_t^{\theta}\right)^{1-\gamma}}{1-\gamma} + \beta\, E_t[\, V_{t+1}(\text{State}_{t+1})] \right], \qquad \forall\, t \leq T.$$

Where C_t is consumption, H_t is housing today, D_t is mortgage debt held today, FC_t is whether the investor pays the fixed stock market entry fee at period t, α_t is the investor's stock market holdings today.

Note in this model, housing is a **continuous** variable - households are forced to hold a positive minimum amount of housing.

State_t consists of "Cash-on-Hand" ("Cash"), X_t, housing chosen yesterday, H_{t-1}, whether they are already a stock market participant IFC_t, an "involuntary movement" state Inv_Move_t and the aggregate labor market state today η_t.

Housing depreciates at a rate δ (1% per annum), there is a one-time stock market participation fee F ($1000) and a house-sale adjustment cost fee of λ(8)%.

The period budget constraint is:

$$S_t + B_t = \begin{cases} X_t - C_t - FC_t\, F - \delta P_t H_{t-1} + D_t, & \forall\, t \text{ s.t. No House Trade,} \\ X_t - C_t - FC_t\, F - \delta P_t H_{t-1} + D_t + (1-\lambda) P_t H_{t-1} - P_t H_t, & \forall\, t \text{ s.t. House Trade.} \end{cases}$$

S_t is the total value of their stock market holdings in period t, B_t is the value of their holdings of a risk-free bond and P_t is the house price.

In any period, an investor's "Liquid Wealth" refers to their total earnings from their portfolio (net of debt) last period:

$$LW_t = R_t S_{t-1} + R^F B_{t-1} - R^D D_{t-1}$$

Here, the risk-free rate is 2%, the mortgage rate is 4% and the stock market return is a random variable (with annual mean 10%). "Cash" in any period is the sum of liquid wealth and labor market earnings:

$$X_t = LW_t + Y_t$$

The investor additionally values their bequest at T + 1,

$$W_{T+1} = X_{T+1} \; - \; \delta H_T P_{T+1} \; + \; (1-\lambda) H_T P_{T+1}$$

as if it was another period of life:

$$U_1 = E_1 \left[\sum_{t=1}^{T} \beta^{\,t-1} \frac{\left(C_t^{1-\theta} H_t^{\theta}\right)^{1-\gamma}}{1-\gamma} + \beta^T \frac{W_{T+1}^{1-\gamma}}{1-\gamma} \right].$$

Investors face three main sources of uncertainty:

1. Labor Income risk
2. Stock Market risk
3. Housing risk

Labor income

An investor at age t's labor income (in logs) follows an exogenous process given by:

$$\tilde{y}_{it} = \begin{cases} f\big(t, Z_{it}\big) + \tilde{u}_{it}, & \text{for } t \leq K, \quad \text{(Working)} \\ f\big(t, Z_{it}\big), & \text{for } t > K. \quad \text{(Retirement)} \end{cases}$$

where K is the date of retirement (here, 65 years). $f\big(t, Z_{it}\big)$ is a deterministic income process and risk, $\tilde{u}_{it}$, is given by the sum of a persistent aggregate process and a transitory shock.

$$\tilde{u}_{it} = \eta_t + \omega_{i,t}$$

$$\eta_t = \phi\eta_{t-1} + \epsilon_t$$

$\omega_{i,t} \sim N(0, \sigma_\omega^2), \epsilon_t \sim N(0, \sigma_\epsilon^2),$. After retirement, individuals receive a deterministic pension.

Housing

The baseline model only considers the problem of homeowners. All investors must purchase at least $H \geq H_m in$ (\$20,000 in 1992 dollars) worth of housing in the first period and retain at least this amount their entire lives.

House prices follow a log-linear time trend with a residual component:

$$log(P_t) = p_t = bt + p_t'$$

To simplify computation, Cocco imposes that p_t', the risk in house prices, is perfectly correlated with the aggregate labor income state η_t.

An investor may also be hit with an"'Involuntary Movement" shock π that forces them to sell their home. Because selling housing is subject to a proportional adjustment cost λ, individuals lose resources when forced to trade when they otherwise would not have chosen to do so.

Stock Market and Debt

Stock market returns are log-normally distributed:

$log(R_t) = \mu + \iota_t$

$\iota_t \sim N(0, \sigma_\iota^2)$.

Households can costlessly adjust their mortgage debt but must respect a down-payment constraint:

$$D_t \leq (1 - d)P_t H_t$$

Table 1. (Annual) [2] Parameters used for the replication.

Description	Parameter	Value
Risk aversion	γ	5
Discount factor	β	0.96
Preference for housing	θ	0.10
Down payment	d	0.15
Depreciation rate	δ	0.01
Involuntary move	π	0.032
Transaction cost	λ	0.08
Riskless rate	$\hat{R}_F - 1$	0.02
Mortgage rate	$R_D - 1$	0.04
Mean stock return	$\exp(\mu + \sigma_\eta^2/2) - 1$	0.10
Std. of log stock return	σ_η	0.1674
One-time entry cost	F	\$1000
Minimum House Size	H_{min}	\$20000

(a) Preferences and stock returns

Description	Parameter	Value
House-price growth	b	1%
SD of house prices	$\sigma_{\hat{p}}$	0.062
Aggregate persistence	ϕ	0.748
SD idiosyncratic income shocks		
No high-school degree	$\sigma_{\omega,1}$	0.136
High-school degree	$\sigma_{\omega,2}$	0.131
College	$\sigma_{\omega,3}$	0.133
SD aggregate income shocks	σ_η	0.019

(b) Housing and income-process parameters

Table 1: Parameters used in the model

Part II: Empirical Replication

A critical difficulty in replicating this paper is the only document that details how to obtain the results is what appears in the paper itself. This meant I often needed to fill in the gaps when practically running regressions.

To ensure that differences in approach do not generate differences in results, in Part III, I used the parameters provided by the paper (although I needed to choose a five-year scaling detailed in Appendix A).

Income Process Estimation

The paper estimates the parameters of the investor's labor income process using data from the Panel Survey of Income Dynamics (PSID) from 1970 to 1992.

It considers only households with a male head and drops any household head-five year age bin (e.g. age 25-29) observation with missing years of labor market income. It also drops the Survey of Economic Opportunity (SEO) subsample of the PSID. I retain these filtering approaches.

In order to allow for self-insurance against labor income risk, it defines labor income as: total reported annual labor income, plus unemployment compensation, workers compensation, social security, supplemental social security, other welfare, child support, and total transfers for both the head and the spouse. Labor market income is deflated into 1992 dollars.

I use 1970-1992 data from the PSID "Equivalent" (Cross-National Equivalent File Project, 2015), which provides most income measures the paper uses at the "head + spouse" level instead the household level, which will likely lead to an overestimate of income. Labor income is the sum of: head + spouse labor income, HH private

[2] Panel b) is taken from a different table where the time-period is unspecified. The housing values are annual but it is possible the income process parameters are five-year values. I cover this possibility in Appendix A. Appendix D covers how I scaled these values into five-year parameters.

transfer income, HH public transfer income, HH social security, HH private retirement income. I also deflate to 1992 dollars using the CPI.

The paper states that "Five-year labor income is for each household equal to the sum of discounted labor income over the relevant age group". It uses an annual discount rate of 5%. It obtains a deterministic life-cycle profile by regressing household labor income against age dummies, a family fixed effect, marital status and household composition.

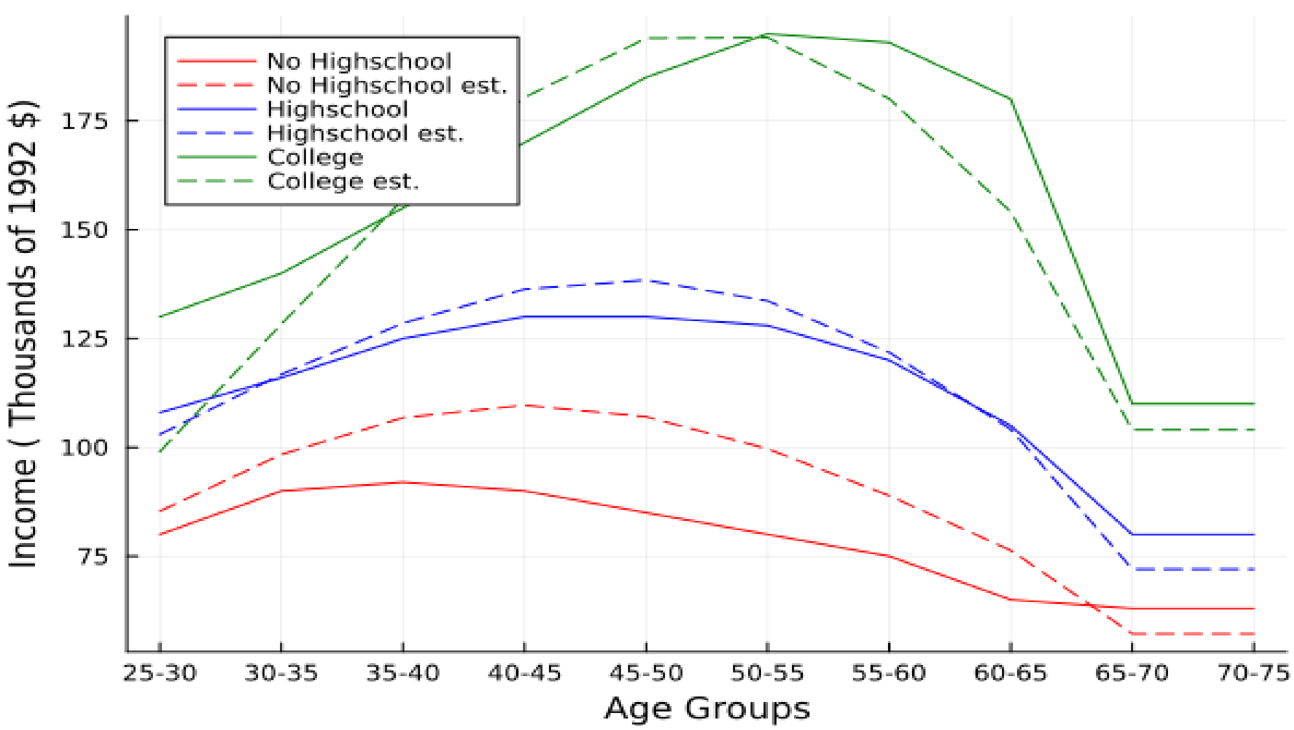
Deterministic income in the model corresponds to the estimated age-group dummies in this regression. It constructs three different profiles corresponding to different levels of education of the head (did not graduate highschool, highschool graduate and college graduate).

In order to specify the regression. I use the method detailed in Cocco, Gomes and Mahenout (2005), which constructs an identical object. Their procedure is the following

- Regress **annual** [3] labor market income on age dummies, a household fixed effect, a martial dummy and the additional number of members in a household above and beyond and the head and the spouse (if present).
- Regress the estimated dummies on a cubic polynomial in age.
- Income after age 64 is the average of the age group dummies between 65 and 75.

After following this procedure, I then aggregate the annual dummies into five-year income using the 5% discount rate.

Figure 1: Five-year labor income process, Paper vs. Estimated



[3]If instead I used five-year income, as is suggested by the paper, I obtain much higher deterministic income path than reported in the paper (see Appendix C).

Figure 1. shows that the estimated deterministic processes closely match those reported in the paper, with two exceptions: early-life college graduates and those who do not graduate high school. Early in life, household formation is the strongest so it is not surprising there is some difference between my largely HH-level estimates and the paper's head + spouse estimates. Furthermore, it is possible that during mid-life, no highschool households receive substantial transfers to dependents that aren't captured in the paper's measure. The no-highschool group represent only 17% of the 1989 PSID sample (as shown in Appendix E), so the sizeable differences for this group shouldn't translate to large differences in the simulation output.

The paper then takes the residuals from this regression to estimate the variance and persistence of the aggregate state η_t and the variance of each education group's transitory shock $\omega_{i,t}$.

The estimated aggregate shock, η_t, can be obtained simply by averaging the regression residuals across all individuals in a year. Its variance is simply the variance of these 23 estimates covering 1970-1992. The estimated persistence is obtained via an AR(1) regression.

The transitory shock for each group is the variance of their deterministic regression residuals minus the variance of the persistent component.

A summary of my empirical estimates compared to those in Cocco is in Table 2 below.

House Price Process Estimation

The paper creates an annual house price series using the average log house value in the PSID each year.

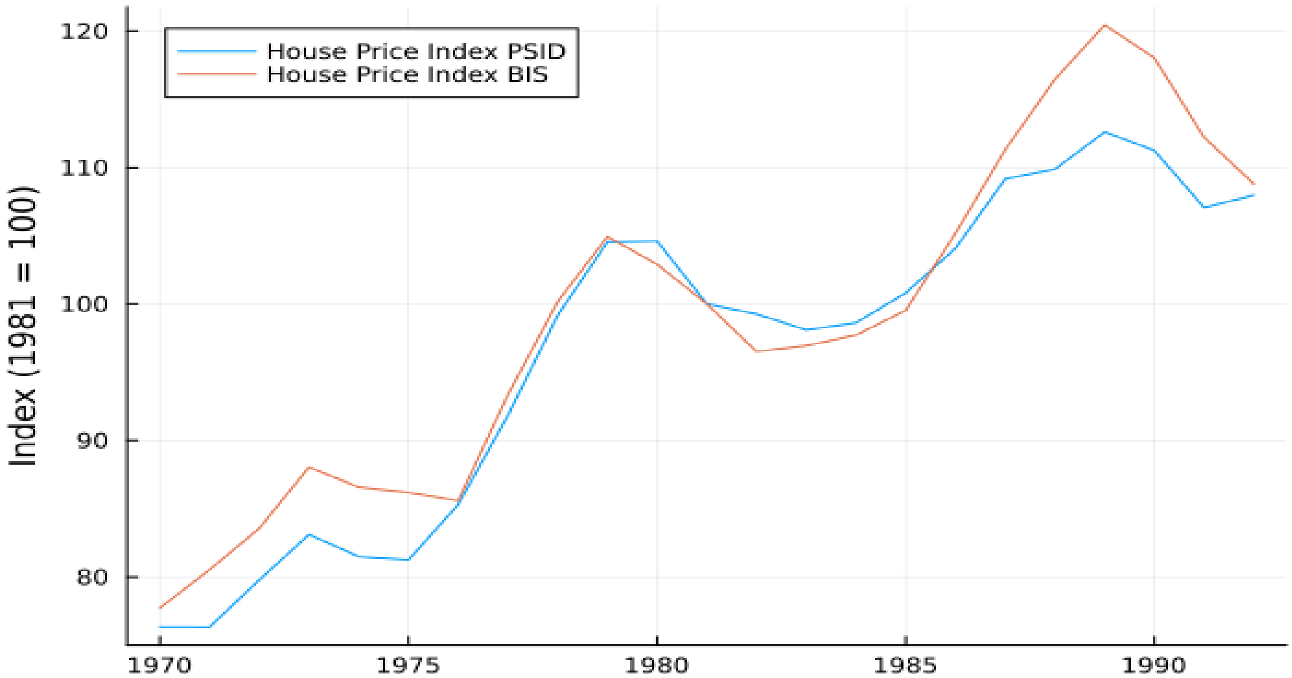
To estimate this, I use 1970-1992 data from the PSID-SHELF dataset (Daumler et al., 2025), which contains PSID data on self-reported house values but not detailed income data and find a close match (as shown in Table 1. [4]

Figure 2. shows the PSID real house price index I constructed closely tracks the Bank of International Settlement's (BIS) real house price index for the U.S. [5] .

[4] Instead using the PSID longitudinal weights produce slightly lower estimates of 1.39% annual growth and 0.048 house price SD.

[5] Cocco presents a plot of his house price index relative to other indices, but the implied growth (200% over 18 years $\sim$ 6.5% per annum) is higher than the 1.6% reported in the paper. As this is likely the corresponding nominal series, I instead checked against another real index.

Figure 2. Comparison of PSID Real House Price Index to BIS



Finally, to construct the variance of the residual component of house prices $\sigma_{\hat{p}}$, I regress log house price growth on a time trend and find the variance of the residual.

Table 2: PSID Estimates Comparison [6]

Description	Parameter	Cocco(2005) Value	Value Est.
Autoregression parameter	ϕ	0.748	0.670
SD idiosyncratic income shocks			
No high-school degree	$\sigma_{\omega,1}$	0.136	0.635
High-school degree	$\sigma_{\omega,2}$	0.131	0.516
College	$\sigma_{\omega,3}$	0.133	0.554
SD aggregate income shocks	σ_{η}	0.019	0.030
Real house-price growth	b	0.016	0.016
SD house prices	$\sigma_{\hat{p}}$	0.062	0.052

In the model, Cocco uses a slightly lower value for annual house price growth, 1%, to account for quality improvement. I do the same.

[6] It is unclear if reported income values are measured annually or on a scaled five-year level. If I use five-year labor market income, the variance estimates remain very similar, but persistence falls to 0.41.

Part III: Model Replication

It is unclear how the labor and housing market process parameters were scaled to their five-year levels in the original paper. My baseline [7]. estimates use the following:

- The variance of the housing process does not change from one to five years (this is equivalent to assuming households make housing choices once every five years).
- The variance of the transitory and persistent processes also do not change (equivalent to assuming they experience one shock every five years).
- Five-year autocorrelation is ϕ^5.

The details of the approach I used to solve and simulate the model, and how they compare to the paper's method, are contained in Appendix B.

Replicated Results

The four replicated exhibits are:

1. Portfolios as a share of assets, split by the magnitude of assets held. This compares the behavior of broadly "poor" to "rich" investors.
2. Portfolios as a share of assets, split by age-group. This compares the behavior of investors across the life cycle.
3. Tobit regressions of stock shares of assets against financial and demographic determinants (holding the path of aggregate shocks fixed over time).
4. Portfolios as a share of financial assets, comparing a model with house price risk to one with no risk. This allows us to learn about the crowding-out effect of house price risk on stock market participation.

Tables 3 and 4 are constructed in the following way: following the paper, I simulate a sample of consumption/portfolio choices by age and education in the same proportion as the age-education distribution in the 1989 PSID [8] . As the deterministic profile is constructed using age-group dummies controlling for marital status and number of additional members of the household, this exercise is for unmarried investors with no dependents.

Like in the paper, I allow the simulated households to differ in their history of both *the aggregate shock* and the other shocks. This provides the unconditional portfolio allocations generated by the model.

By examining Table 3, we see in the original paper there is a stark distinction between rich and poor investors. Rich investors almost entirely invest in stocks (as opposed to bonds) as predicted by standard portfolio choice models. However, poor investors invest much more in bills, as their portfolio is disproportionately dominated by real estate.

In my replication, the pattern is still present, but the difference between rich and poor is much less stark. My simulation predicts the stock share of poor investors to be around 82% as compared to 25% in the paper. My replication predicts that rich investors hold practically 100% in stock, compared to 95% in the original model.

In general, while the predicted values are relatively close for rich investors, my replication consistently predicts meaningfully higher amounts of stock market participation for poorer investors.

[7] My appendix A reports a matrix of alternative scaling approaches I also considered, and that using them does not affect the final result)

[8] The weights I used are in Appendix E

Table 3: Portfolio shares by financial assets predicted by the model (author's estimates in **bold** parentheses)

Asset	Liquid assets		Financial assets		Total assets	
	< 100k	≥ 100k	< 100k	≥ 100k	< 100k	≥ 100k
Stocks	0.252 (**0.823**)	0.952 (**0.995**)	0.049 (**0.174**)	0.460 (**0.487**)	0.017 (**0.303**)	0.213 (**0.172**)
Bonds (Bills)	0.748 (**0.177**)	0.048 (**0.005**)	0.029 (**0.036**)	0.013 (**0.004**)	0.008 (**0.023**)	0.006 (**0.003**)
Liquid assets	1.000	1.000	0.078 (**0.210**)	0.473 (**0.491**)	0.025 (**0.326**)	0.219 (**0.175**)
Real estate			0.922 (**0.790**)	0.527 (**0.509**)	0.248 (**0.674**)	0.251 (**0.195**)
Financial assets			1.000	1.000	0.273 (**1.000**)	0.470 (**0.371**)
Human capital					0.727 (**0.000**)	0.530 (**0.629**)
Total assets					1.000	1.000
Debt			0.509 (**0.655**)	0.034 (**0.310**)	0.141 (**0.029**)	0.018 (**0.103**)
Stock mkt. part.	0.253 (**0.825**)	0.974 (**1.000**)				

Notes: The first line in each cell is the share reported in (Cocco, 2005).

The value in bold parentheses underneath is the author's own estimate obtained from simulation.

"Liquid assets" refers to the sum of stocks and bonds. "Financial assets" are liquid assets plus real estate value. Total assets includes "Human Capital" - constructed as discounted expected future labor income using an annual discount rate of 5%. I approximated this using monte-carlo simulation.

In the paper, Cocco calls this table Portfolio Shares by "Financial Net Worth", computed by subtracting debt from each assets definition. However, the table notes and values indicate all values were normalized by assets instead.

Table 4. provides average portfolios predicted by the model across the lifecycle. While stock market shares are still upward sloping over the lifetime, we can again see much higher stock market participation in my simulated data. On the extensive margin, we see that 60% of young households participate in the stock market, as opposed to 2% reported in the paper. On the intensive margin, we see that stocks represent a meaningful share of total lifetime assets for all age-groups while in the paper they represent only a small share of investment.

I also noticed that debt, as a proportion of financial assets, is relatively flat under the reported model whereas it falls in my model. To me, this suggests that agents are using debt to consumption smooth and potentially "lever up" at old age in the reported model, whereas under my simulation, debt is much more associated with the initial house purchase.

I wanted to highlight one other meaningful area of difference: the "human capital" share of total lifetime assets. Following the paper, I define an investor's human capital to be the present expected value of their future lifetime earnings using a discount rate of 5%. I calculate this expectation using monte-carlo simulation.

In Table 3., my simulation predicts that those with low total assets should have **no** human capital, while the paper reports that **over 70%** is held in future human capital. I believe this discrepancy points to a difference in how this present-value is calculated.

Table 4: Portfolio shares by age predicted by the model
(author's estimates in **bold** parentheses)

Asset	Liquid assets				Financial assets				Total assets			
	<35	35–50	50–65	≥ 65	<35	35–50	50–65	≥ 65	<35	35–50	50–65	≥ 65
Stocks	0.02 (**0.59**)	0.21 (**0.90**)	0.61 (**1.00**)	0.72 (**0.98**)	0.01 (**0.14**)	0.06 (**0.36**)	0.24 (**0.61**)	0.10 (**0.46**)	0.00 (**0.03**)	0.01 (**0.10**)	0.10 (**0.34**)	0.07 (**0.39**)
Bills	0.98 (**0.41**)	0.79 (**0.10**)	0.39 (**0.00**)	0.28 (**0.02**)	0.03 (**0.04**)	0.02 (**0.01**)	0.04 (**0.00**)	0.01 (**0.01**)	0.00 (**0.01**)	0.01 (**0.00**)	0.02 (**0.00**)	0.01 (**0.01**)
Liquid assets	1.00	1.00	1.00	1.00	0.03 (**0.17**)	0.08 (**0.37**)	0.28 (**0.61**)	0.11 (**0.47**)	0.00 (**0.03**)	0.02 (**0.11**)	0.12 (**0.34**)	0.08 (**0.40**)
Real estate					0.97 (**0.83**)	0.92 (**0.63**)	0.72 (**0.39**)	0.89 (**0.53**)	0.13 (**0.13**)	0.15 (**0.15**)	0.29 (**0.21**)	0.72 (**0.47**)
Financial assets					1.00	1.00	1.00	1.00	0.13 (**0.16**)	0.17 (**0.26**)	0.41 (**0.55**)	0.80 (**0.87**)
Human capital									0.87 (**0.84**)	0.83 (**0.74**)	0.59 (**0.45**)	0.21 (**0.13**)
Total assets									1.00	1.00	1.00	1.00
Debt					0.68 (**0.70**)	0.43 (**0.54**)	0.27 (**0.20**)	0.51 (**0.04**)	0.09 (**0.11**)	0.07 (**0.13**)	0.12 (**0.09**)	0.42 (**0.03**)
Stock mkt. part.	0.02 (**0.59**)	0.21 (**0.90**)	0.62 (**1.00**)	0.73 (**1.00**)								

Notes: First line in each cell is the model share reported by (Cocco, 2005). The value in bold parentheses underneath is the author's own estimate obtained from simulation.

Moreover, examining the deterministic income paths of Figure 1., notice that investor's deterministic labor income in each period is large, for most groups in excess of 100K. In particular, given the assumptions of the model, this stream **becomes deterministic** in retirement, so differences in the specification of shock processes will not affect this quantity. Investors are also forced to save a minimum amount in housing.

This means the greater than 100K portfolio share for total assets should represent close to the unconditional distribution for both of our models. However, while human capital in my simulation exceeds the paper's by 10 p.p. in Table 3., examining the corresponding age splits for Table 4., my predicted human capital shares are instead **consistently lower**. This means a considerable number of investors must be total asset "poor" in the original paper's simulation. The only way I can reconcile this is if the persistent aggregate shock is extremely large in the original solved model, enough to greatly reduce housing values. This would generate a number of investors whose net worth is almost entirely in their human capital, despite the strong bequest motive. Notice these poor investors hold no stock or bills, and their representative real estate portfolio share of 24% implies they must be holding close to the minimum allowable amount of housing (which would be $20,000 if there was been no growth in prices).

By contrast, the only investors who have less than 100K in total assets in my simulation are in their last period, and so necessarily have 0 expected future earnings.

Table 5 shows the determinants of stock market portfolios shares and the absolute investment in stocks. It conditions instead on a single aggregate shock history that best matches the path that actually occurred. I generated this by inspecting the path of the house price index process against a linear time trend from 1970 - 1989 [9] .

All households of the same age will have experienced the same history of shocks across their lifetimes. However, households of different ages will not have experienced the same histories (as age and cohort diverge in this simulation). Furthermore, I follow the paper in specifying the shocks experienced by those born before 1945 (i.e. pre-1970 shocks) to the unconditional mean (of 0).

Financial net-worth is constructed by subtracting debt from financial assets. Relative real-estate and mortgage variables are constructed by dividing real estate assets and debt respectively by financial net worth.

Table 5: Tobit regressions of stock holdings on investor characteristics

	Stock relative to liquid assets	Stock relative to financial assets	Stock relative to total assets	Stocks
Intercept	-8.405	0.423	0.085	-94223.460
	(0.986)	**(0.460)**	**(0.004)**	**(-115635.200)**
Total income	1.59e-05	3.73e-07	-1.67e-07	0.209
	(1.39e-06)	**(-3.54e-07)**	**(-6.63e-07)**	**(0.689)**
Financial net-worth	2.92e-04	1.58e-06	1.35e-06	
	(-6.43e-07)	**(5.97e-07)**	**(7.32e-07)**	
Age	4.608	0.021	0.014	12110.850
	(0.002)	**(0.004)**	**(0.006)**	**(2171.252)**
Relative real estate	-48.834	-0.762	-0.266	
	(-0.289)	**(-0.651)**	**(-0.147)**	
Relative mortgage	48.174	0.725	0.264	
	(0.091)	**(0.591)**	**(0.127)**	
Real estate				0.245
				(0.570)
Mortgage				-0.957
				(-1.035)

Notes: This table compares the reported results of cross-sectional Tobit regressions of stock holdings (relative to liquid, financial, and total assets) and the dollar value of stocks on investor characteristics. Values in bold parentheses underneath are my estimates. Independent variables include total income, financial net-worth, age, and measures of real estate and debt relative to financial net-worth. Real estate and mortgage variables enter the dollar regression directly.

Differences in the sign of financial net worth can be explained by differences in the magnitude of stock market participation between the reported model and my replication. This is because the underlying portfolio rule (for investing in stocks) is decreasing in financial net worth, conditional on stock market entry [10]. However, stock market entry rises with financial net worth.

Lower participation in the paper generates a difference in sign for liquid assets. My estimate picks up the

[9] The best - match was: 1970-1974 (below trend), 1975 - 1979 (above trend), 1980-9184 (on trend) and 1985-1989 (on trend). These correspond one to one with the realization of the three grid-point aggregate shock.

[10] As financial net worth rises, the implicit holdings of safe assets through future labor income fall in significance as a share of lifetime income, and therefore investors substitute towards bills.

shift towards bills by the rich, while the paper's picks up stock market entry. My sense is that normalizing by financial or total assets relative to liquid assets weights the results more towards those investors who have a lower denominator. Intuitively, those who have less financial net worth (conditional on age) tend to participate less and therefore drive a positive effect.

To understand the effect of income, the paper states "It is future and not current labor income that constitutes the implicit holdings of an asset. But the two tend to be positively correlated". However, under my construction, they will not be. This is because I compare individuals from different birth cohorts but **all in the same year** (1989). As a result, they face the same aggregate labor income level, and income varies only due to transitory shocks.

I replicate the positive correlation between mortgage size and stock shares reported in the paper. This is likely influenced by lifetime income: individuals with higher incomes over their lives can afford to take more risk.

Finally, I replicate the substitution between real estate shares and stock shares.

In Table 6., I report portfolio shares of financial assets from the baseline model compared to the same model but solved with no house price risk.

Table 6: Portfolio shares relative to financial assets: Baseline vs. No House-Price Risk

Asset	Baseline (with house-price risk)		No house-price risk	
	< 100k	≥ 100k	< 100k	≥ 100k
Stocks	0.049 (**0.174**)	0.460 (**0.487**)	0.056 (**0.163**)	0.489 (**0.473**)
Bills	0.029 (**0.036**)	0.013 (**0.004**)	0.022 (**0.038**)	0.002 (**0.006**)
Liquid assets	0.078 (**0.210**)	0.473 (**0.491**)	0.078 (**0.201**)	0.491 (**0.479**)
Real estate	0.922 (**0.790**)	0.527 (**0.509**)	0.922 (**0.799**)	0.509 (**0.521**)
Debt	0.508 (**0.655**)	0.034 (**0.310**)	0.473 (**0.662**)	0.013 (**0.326**)
Stock mkt. part.	0.253 (**0.646**)	0.974 (**0.986**)	0.311 (**0.611**)	1.000 (**0.971**)

Notes: First line in each cell is the share reported by (Cocco, 2005). Bold values in parentheses below are the author's estimates from simulation.

The reported stock market participation appears to match the values for *liquid assets* in Table 3. rather than financial assets. I interpret this as a typo.

We can see that in the reported results, (primarily poor) investors increase their stock shares when house price risk is shut-off. This is evidence for house price risk crowding out stock market investment.

However, in my simulation, the opposite occurs, under no house price risk, participation and stock shares actually **fall** somewhat. This is consistent with the perfect correlation of housing with income risk making stocks an attractive hedge - and actually crowding in stock market investment.

Part IV: Extension

To study the impact of housing affordability on portfolio choice, I revised the model into a buy vs. rent decision. Following the literature, notably (Paz-Pardo, 2024), I change on this decision, I make a number of changes to the baseline model:

1. Mortgages cannot be flexibly renegotiated and must be paid off when downsizing (or upon death). In practice this requires an additional state - the level of mortgage debt an investor had in the previous period.
2. Instead of a continuous housing choice, I use only a single housing value set to the median house price reported in 1992, \$80,000 [11].
3. Instead of a housing service "dividend" proportional to the value of the home, I add housing services, a proportional scaling of CRRA consumption utility, that owners benefit from. I also increase θ to 0.2 to match other studies which match this to National Income and Product Accounts data on budget shares (Paz-Pardo, 2024).
4. Consistent with the data, investors pay 5% of the price of a home annually if they rent. The five year %, 22%, is the present value of five annual rental payments, discounted at a rate of 5% annually.
5. As the down-payment decision is so important for entry, I increased the down payment proportion to 0.2 consistent with the literature.
6. I further followed Paz-Pardo, in increasing the mortgage interest rate to 5% for retirees to model increased credit constraints for pensioners.
7. I focus only on the high school graduate cohort, the largest education group.

The updated utility function is:

$$U_1 = E_1 \left[\sum_{t=1}^{T} \beta^{\,t-1} \frac{\left(C_t^{1-\theta} s_t^{\theta}\right)^{1-\gamma}}{1-\gamma} + \beta^T \frac{W_{T+1}^{1-\gamma}}{1-\gamma} \right].$$

and the mortgage constraint (in addition to the downpayment and the requirement to at least pay the interest on the mortgage) is:

$$M_{t+1} \leq \begin{cases} M_t, & \forall\, t \text{ s.t. Existing Home-Owner, staying in their home,} \\ 0.0, & \forall\, t \text{ s.t. Existing Home-Owner, downsizing,} \\ (1-d)P_t, & \forall\, t \text{ s.t. New Home-Owner.} \end{cases}$$

I calibrate the stock market participation rate and housing services in the model to match stock market participation rates between age 25 and 34 and home ownership rates between 45 and 49 in the PSID. The value of s was 1.15 implying a 15% scale-up of utility from consumption from owning. F, the one time cost of stock market entry, was \$2000.

[11] It would be interesting to increase the number of housing levels, e.g. to add "starter" homes, and to introduce variety in other attributes of housing, such as a spatial element. This would move in the direction of the "spatial housing ladder" model of (Fonseca et al., 2024), used to model the general equilibrium impact of mortgage lock-in.

Once I solved and calibrated the model, I simulated a large of households and investigated the impact of a a "small" (representing moving from a low aggregate state to the medium state) and a "large" (representing moving from the low aggregate state to the high state) shock to house prices.

My sample was investors at age 35 (similar to millennials today) who had rented from age 25 up to 35 and had experienced only low aggregate states up to that point. The control group continued to stay at the low state, while the treated group's house prices relatively rose by either 9% or 18%. [12].

Figure 3. shows the coefficients of a two-way fixed effect specification (with five years prior being the reference period) we see evidence that house price rises do have a negative impact on home ownership as expected. The impact for the large housing shock is rather extreme at 100%: likely due to approximating a probability of purchasing a house using a linear probability model.

Figure 3. Impact of House Price Shock on Home Ownership

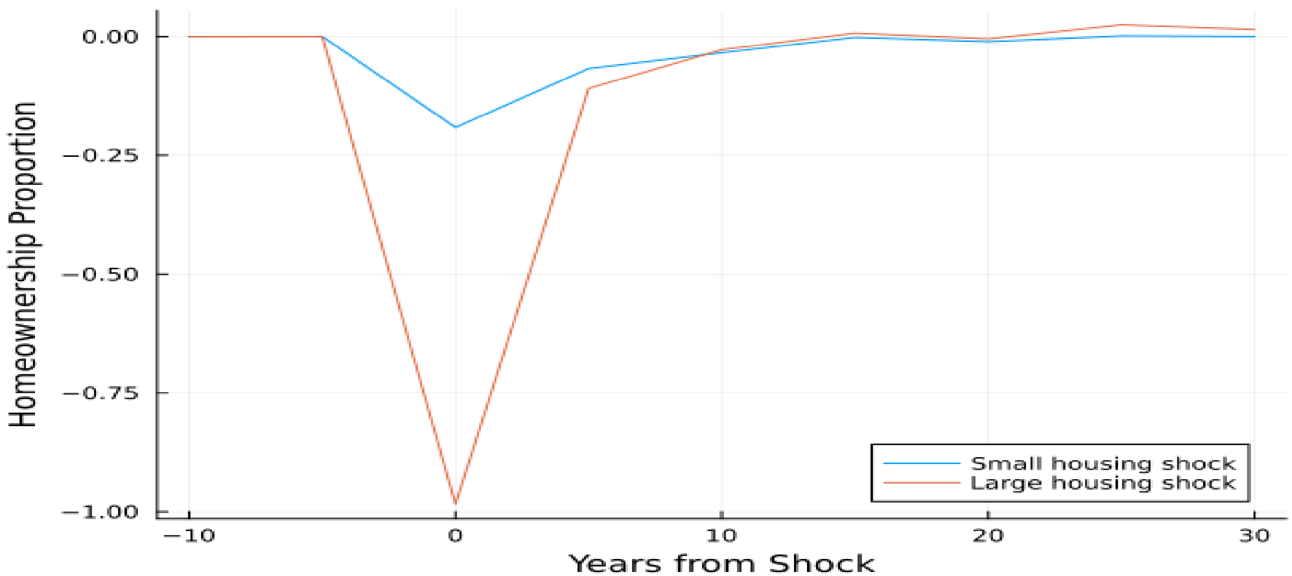


Figure 4. shows a counter-intuitive result: while those treated with a small house price shock markedly reduce their stock market participation, those treated with a large shock do not adjust their participation at all. My interpretation of this is that those with a small shock primarily reduce their saving in financial assets to purchase a home. However, a large shock forces many to continue renting and save in stocks.

[12]In practice, as there is mean reversion in the house price state, this mixes a relatively small negative realization compared to expectations for the control group with a larger positive shock for the treated. As the aggregate state also relates to persistent labor income, this means their labor income also rose by between 3% and 6%)

Figure 4. Impact of House Price Shock on Stock Market Participation

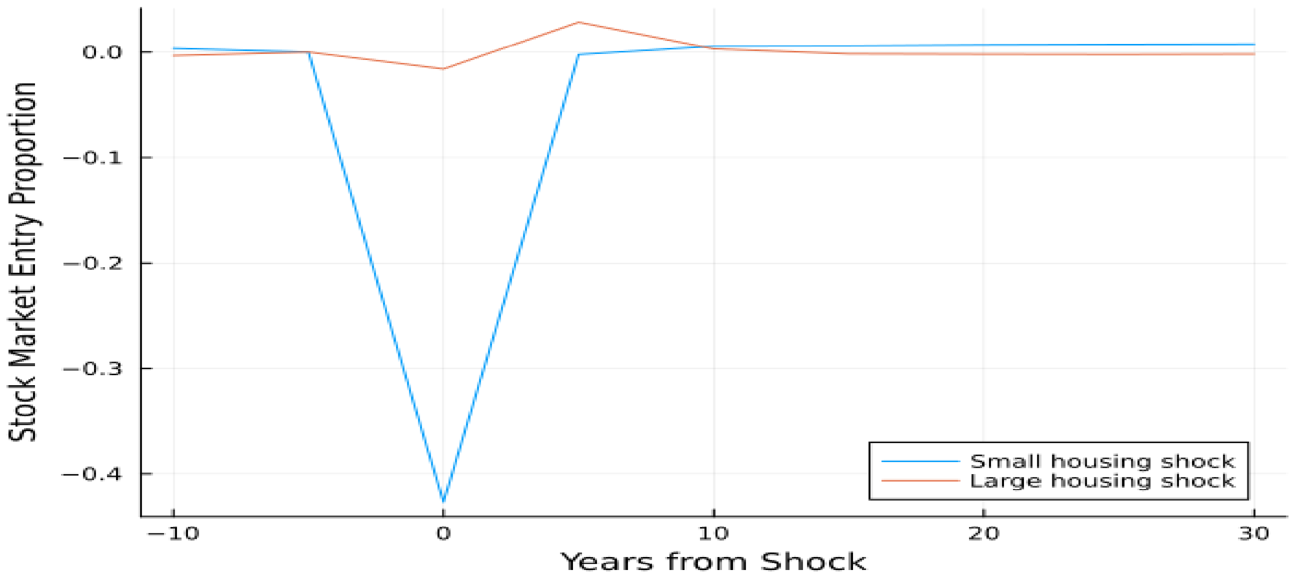
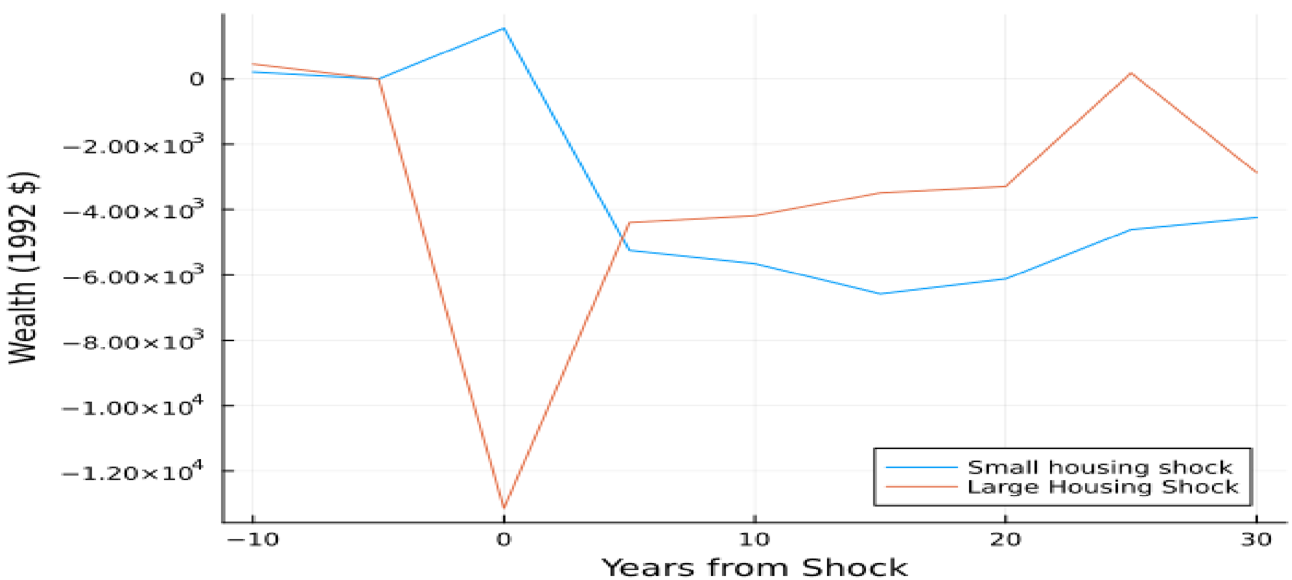


Figure 5. shows results consistent with the large housing shock individuals saving less upon impact, while the small housing shock individuals actually save more. However, the long-term impact on wealth accumulation is the opposite of the short term effect: due to their earlier exposure to the stock market, the large housing shock individuals end up accumulating somewhat more wealth over their lifetimes than those hit with a small shock.

Figure 5. Impact of House Price Shock on Wealth Accumulation



Appendix

A: Robustness

In this section, I outline several alternative calibration approaches for the labor and house price parameters I considered and why I did not use them for the main results. None of these approaches noticeably reduced the gap between my model-predicted stock-market participation rates and those reported in the paper.

Alternative approach	**Alternative value**	**Rationale**	**Reason not selected**
Scaling labour-income process parameters to five-year values (sum of random variables)	• $\text{Var}(\eta_5) = 13.6 \text{ x } \text{Var}(\eta_1)$ • $\text{Var}(\omega_5) = 4.1 \text{ x } \text{Var}(\omega_1)$ • $\phi_5 = 0.415$	If Cocco ran his regressions at the *annual* level in the PSID (as suggested by my empirical section), this applies the same method that turns the deterministic income profile into random shocks.	1. Running the Cocco method on *five-year* discounted household income recovers five-year persistent and transitory variances directly. These are close to the annual estimates, not many multiples of them. 2. Because the aggregate and house-price states are perfectly correlated, their variances would also have to be scaled. This massively inflates house-price risk and prevents the model from matching PSID residuals (needed for Replicated Table 2).
Not scaling labour-income parameters at all	• $\text{Var}(\eta_5) = 0.019^2$ • $\text{Var}(\omega_5) = \text{Var}(\omega_{i,t})$ • $\phi_5 = 0.75$	If Cocco's regressions were actually run at the *five-year* level, no scaling of labour-income parameters is necessary.	1. Housing-process scaling would still break perfect correlation because the housing parameters are annual. 2. Five-year persistence of 0.75 for the aggregate labour-market/housing state seems high (my PSID estimate was 0.41). 3. Fails to match Cocco's deterministic lifecycle profile closely. 4. Cocco's empirical table mixes unscaled annual (housing) with five-year (income) parameters.
Starting value of the house-price index	$P_{\text{init}} = 0.7$	Matches the ratio of 1970 to 1992 house-price levels.	Essentially lowers the relative price of houses, so stock-market participation jumps even earlier because the initial cost of purchasing housing falls.

Table 2: Alternative calibration choices and reasons for rejection

B: Numerical Technique

Solution

Because this is a finite horizon model, I use backwards induction to approximate each education group's value and policy functions.

While the paper describes using grid search over equally-spaced grids for choices, I instead use equally-spaced grid-search *only* over stock market entry and two "share" choice variables (the share of stocks to liquid assets and the household's loan-to-home value ratio, which implies a unique debt choice given a price level and a choice of housing) but optimize over the choice of contemporaneous consumption and housing.

My optimization routine is as follows:

1. Conditional on choosing to move, find the optimal choice of both housing and consumption subject to the "House Trade" period budget constraint. The value of a choice is the sum of its flow-utility value and the implied probability-weighted average continuation value across the possible states the choice could lead to next period (shocks are the moving shock, the stock market shock, the transitory shock and the persistent transition).

   I used nested univariate (Brent) optimization over housing and consumption. I find the consumption policy function given any choice of housing tomorrow and then find the choice of housing which optimizes this function. Brent works by progressively reducing an interval around the optimum - I chose a tolerance of \$500 (reducing the tolerance has no noticeable effect on the solution).

2. Conditional on not moving (if it is feasible), I find the optimal choice of consumption subject to the "No House-Trade" period budget constraint.

3. The overall choice is whichever of the conditional choices has the higher implied value.

Unlike Cocco, which used cubic-spline interpolation, I found it was better to use bilinear interpolation to compute this continuation value for points not on the grid (housing and cash-in-hand) as cubic spline-interpolation led to strange behavior of the value function near the default state). A choice which leads to potential default (or land outside the grid) in the next period received a very low punishment utility.

Optimizing may be particularly important when studying stock market participation rates. As the barrier to entry is close to 1% of average labor income in each period, I didn't want individuals to spuriously enter because they had money left over in their budget due to my choice of grid. It is also possible that a particular grid for housing or consumption could lead individuals to choose a value higher than the optimum, leading to a lack of stock market participation.

A summary of my choice of grids:

- Housing: 35 grid points between \$20000 and \$700000
- Cash: 181 grid points between $-\$150000$ and \$3.5 million.
- Stock Share: 20 grid points between 0.0 and 1.0.
- Loan-to-value ratio: 15 grid points between 0.0 and $0.85 = 1 - d$.

My initial house price index started at 1.0.

Simulation

I then simulate lifecycles for each cohort and record the values each household attains of each relevant variable in each year. I simulate a number of households such that the proportion of observations by age / education group matches the PSID proportion presented in the paper.

Notes on the simulation:

- Replication Table 1.is constructed allowing both aggregate and transitory states to vary across lifecycles. However, the initial aggregate state individuals enter with must be specified. The initial state in my simulation is drawn from the stationary distribution of the aggregate state.

- Due to the two discrete choices individuals face (Stock market entry and whether to move), the linearly interpolated policy functions used in the simulation can imply fractional choices. I rounded fractional decisions down. I attempted to minimize the impact of this rounding by using as many grid points as possible.

Discretizing transitory and persistent income:

While Cocco used the (Tauchen and Hussey, 1991) approximation method, [13] due to the persistent component of income, I used the (Rouwenhorst, 1995) method and matched the unconditional mean, variance and first-order autocorrelation of the processes I discretize [14].

Noticing that all random variables (stock market return shocks, transitory shocks and aggregate shocks) are uncorrelated, I discretized them separately. House prices, as they are perfectly correlated with the aggregate state, do not need to be discretized and instead house-price grid points are obtain via $\frac{1}{\kappa_\eta}\eta$, where κ_η is simply the ratio of the variance of the aggregate state over the variance of the persistent component.

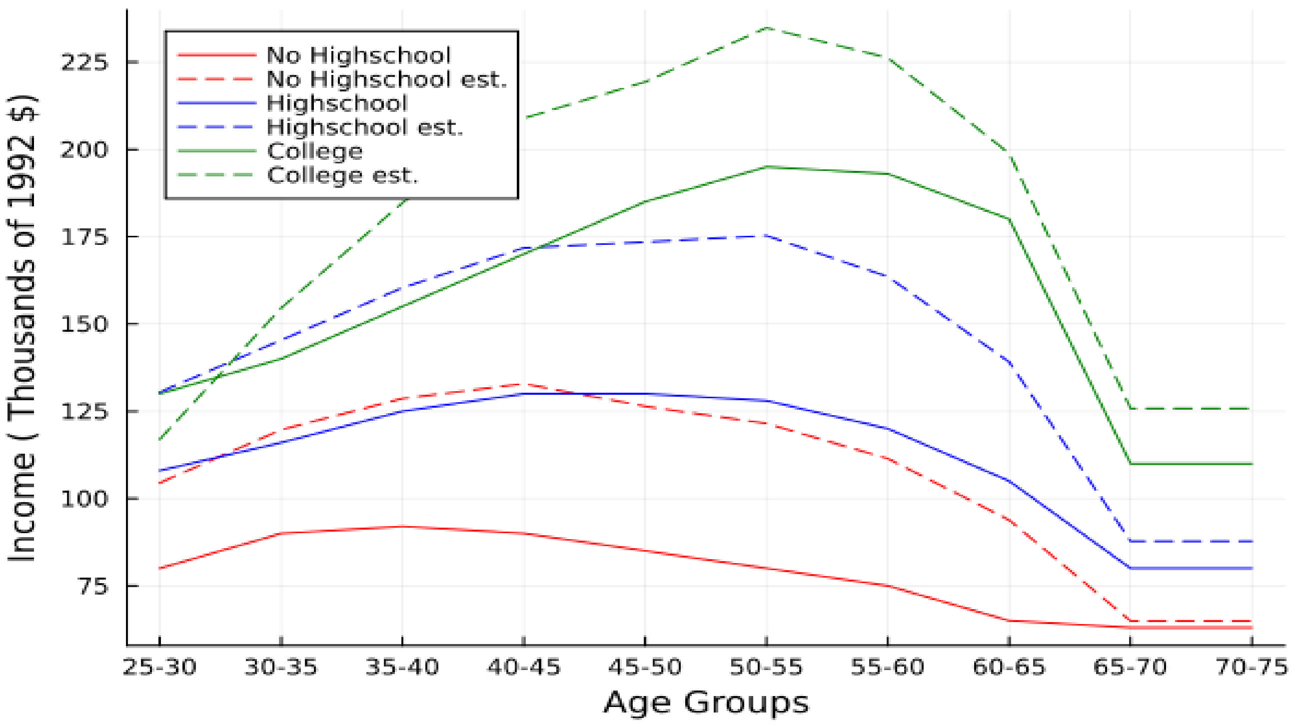
Like Cocco, I used 3 grid points for the aggregate state. I used 11 grid points for both the transitory and stock market returns shocks (these were not specified).

The distribution of the initial persistent component, η_0, is taken from the stationary distribution of η_t found using Monte-Carlo Simulation.

[13] He did not specify the maximum value he used, if the maximum value is 3 standard deviations above/below 0, then this would pose considerable issues for his exercise simulating the single path of shocks which actually occurred from 1970 to 1989 using a three-state grid as all values would be at the unconditional mean.

[14] The maximum value is then chosen to match the unconditional variance given the number of grid points.

C: Five-Year Deterministic Income Profile



D: Five-Year Scaling

Table 1. Five-Year Parameters used for the replication

Description	Parameter	Scaling
Risk aversion	γ	
Discount factor	β	β^5
Preference for housing	θ	
Down payment	d	
Depreciation rate	δ	$1-(1-\delta)^5$
Involuntary move	π	$1-(1-\pi)^5$
Transaction cost	λ	
Riskless rate	$\hat{R}_F - 1$	$\hat{R}_F^5$
Mortgage rate	$R_D - 1$	R_D^5
Mean stock return	$\exp(\mu + \sigma_\eta^2/2) - 1$	5μ
Std. of log stock return	σ_η	$5\sigma_\eta$
One-time entry cost	F	
Minimum House Size	H_{min}	

(a) Preferences and stock returns

Description	Parameter	Scaling
House-price growth	b	$5b$
SD of house prices	$\sigma_{\hat{p}}$	
Aggregate persistence	ϕ	ϕ^5
SD idiosyncratic income shocks		
No high-school degree	$\sigma_{\omega,1}$	
High-school degree	$\sigma_{\omega,2}$	
College	$\sigma_{\omega,3}$	
SD aggregate income shocks	σ_η	

(b) Housing and income-process parameters

Table 3: Parameters used in the model

Notes: Blank values were left unscaled. Agents are restricted to choose housing once every five years. It is also assumed income shocks only hit once every five years. Alternative scalings for these processes are summarized in Appendix A.

E: Age-Education Weights

Table 4: Age–education distribution

Age group	**No highschool**	**Highschool**	**College**	**Total**
25–29	0.01	0.06	0.02	0.09
30–34	0.02	0.08	0.04	0.14
35–39	0.01	0.10	0.06	0.17
40–44	0.02	0.08	0.07	0.16
45–49	0.01	0.05	0.03	0.09
50–54	0.02	0.04	0.02	0.07
55–59	0.02	0.04	0.02	0.08
60–64	0.02	0.03	0.02	0.08
65–69	0.02	0.04	0.01	0.07
70–74	0.02	0.03	0.01	0.06
Total	0.17	0.54	0.29	1.00

Notes: Each cell reports the share of the total sample in the indicated age–education cell; row totals equal the final column.
Taken from (Cocco, 2005).

References

(2025). Panel study of income dynamics. Accessed: May, 2025. Public Use Data.

Cocco, J. F. (2005). Portfolio choice in the presence of housing. *The Review of Financial Studies*, 18(2):535–567.

Cocco, J. F., Gomes, F. J., and Maenhout, P. J. (2005). Consumption and portfolio choice over the life cycle. *The Review of Financial Studies*, 18(2):491–533.

Cross-National Equivalent File Project (2015). Cross-national equivalent file (cnef).

Daumler, D., Friedman, E. M., and Pfeffer, F. T. (2025). *PSID-SHELF User Guide and Codebook, 1968–2019, Beta Release*. Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI.

Fonseca, J., Liu, L., and Mabille, P. (2024). Unlocking mortgage lock-in: Evidence from a spatial housing ladder model. Working paper.

Paz-Pardo, G. (2024). Homeownership and portfolio choice over the generations. *American Economic Journal: Macroeconomics*, 16(1):207–237.

Rouwenhorst, K. G. (1995). Asset pricing implications of equilibrium business cycle models. In Cooley, T. F., editor, *Frontiers of Business Cycle Research*, pages 294–330. Princeton University Press, Princeton, NJ.

Tauchen, G. and Hussey, R. (1991). Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica*, 59(2):371–396.