Precautionary Saving Puzzle: How Marginal Utility Shocks Can Cause Excessive Saving at the Tail

Amin Shirazian

Department of Economics Brownbag Series

Toronto Metropolitan University

Fall 2025

Outline

Background

- When markets are complete, no incentives exist for precautionary saving (Sargent and Ljungqvist, 2012)
  - Perfect consumption smoothing is possible
  - If people born alike, they remain alike $\rightarrow$ Cannot generate heterogeneity/inequality
- Bewley (1977); Huggett (1993); Aiyagari (1994) introduce market incompleteness
  - Lack of full risk-sharing creates precautionary saving motive against uninsurable idiosyncratic income risk
  - $\rightarrow$ Individual's wealth depends on entire history of idiosyncratic shocks
  - $\rightarrow$ Gives rise to heterogeneity in wealth
- BHA models collectively became workhorse models of inequality research in modern macro (aka "income fluctuations models")

Background

- But BHA models cannot account for top inequality:
  - Wealth distribution is Pareto in US, Canada, Europe
  - Even for extreme distributions of idiosyncratic risk, BHA models predict a thin-tailed wealth distribution
- Underlying reason?
  - Even without Arrow securities, becomes possible to smooth consumption "almost perfectly" with "large enough" wealth
  - $\rightarrow$ BHA models predict a threshold after which individuals cease saving—saving rate becomes negative

This Paper: Theoretical Contribution

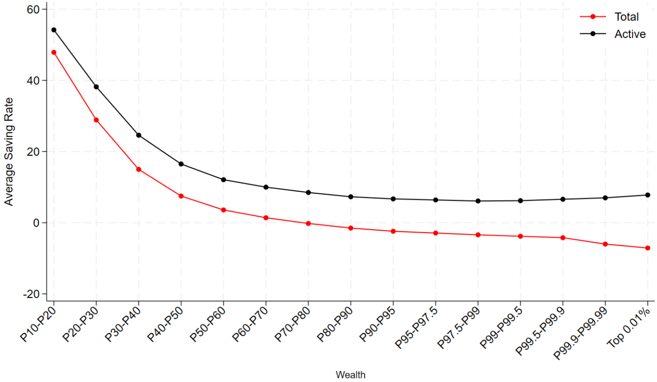
- Identify necessary and sufficient condition for precautionary saving motive to diminish
  - Does optimum achieve consumption smoothing *above a threshold?*
- Introduce new class of models that can imply ever-increasing saving at upper tail of wealth distribution
  - Inequality at top arises from *precautionary saving* of wealthy individuals
  - Not, *e.g.*, because of idiosyncratic capital income risk, *á la* Benhabib et al. (2015, 2017)
- Preview of our "example" mechanism:
  - Shock affects marginal continuation value
  - Keeping marginal utility from dropping is "luxury" compared to consumption (*á la* Eslami (2019))
    - Optimal to maintain marginal utility, instead of smoothing consumption
    - $\rightarrow$ Saving can be bounded away from zero, for certain parameter values

This Paper: Empirical Contribution

- Empirical contribution:
  - We show saving non-monotonicity is robust feature of Health and Retirement Survey data (even though HRS does not capture top wealth well)
  - Develop and calibrate empirical model with chronic health shocks to match HRS
    - How much of top shaving can be explained through lens of marginal utility shocks? (in progress)

Outline

Observation: Positive Saving at the Upper Tail



Source: Figure 4 in Bach et al. (2017), using administrative data from Sweden

Definitions: Saving and Saving Rate

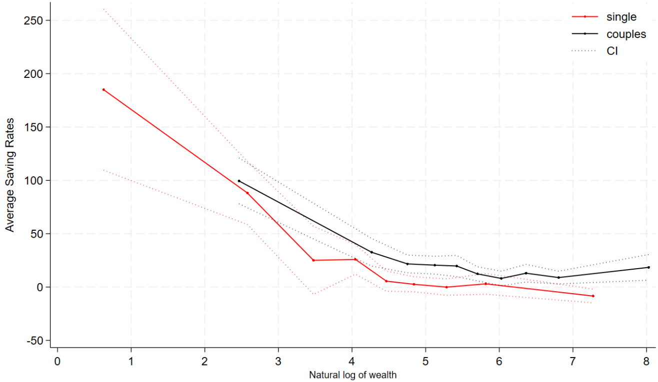
- We identify same pattern in Health and Retirement Study (HRS).
- Saving is defined as the first difference of wealth:

$$S_{it} = b_{it+1} - b_{it}$$

- Saving rate defined as:

$$s_{i,t} = \frac{S_{i,t}}{W_{i,t}}$$

Average Saving Rate as Function of Current Wealth



Sample: 55-65 years old HRS households from 1994 to 1996.

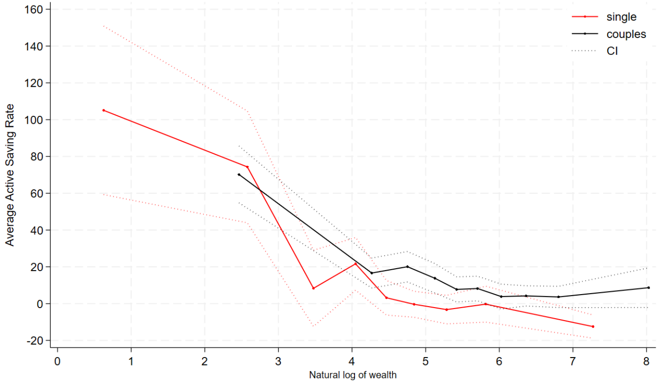
Active vs Total Saving

- Active saving rate defined as:

$$s_{it}^{act} = \frac{S_{it}^{tot} - r_{it+1} b_{it}}{W_{it}}$$

- Excludes reinvested capital income.
- Active rates lower than total across deciles.

Active Saving Rates as Function of Current Wealth



Sample: 55-65 years old HRS households from 1994 to 1996.

Outline

Bewley-Aiyagari-Huggett (BAH) Setup

- Economy with idiosyncratic income shocks, only risk-free bonds with return r, and *ad hoc* borrowing constraint $-\underline{b}$
  - For simplicity, assume r is given, and smaller than discount rate
- An individual enters period t with income $y \in \mathcal{Y} = \left\{\underline{y}, \bar{y}\right\}$, and asset holdings $b \in [-\underline{b}, \infty]$
  - Decides how much to consume (c) and hold assets for next period (b')
  - If high in current period, next period's income will be high wpr. π
- Will refer to current income states as low and high states
- CRRA utility function

Individual's Problem in BAH

- Individual's problem in high state:

$$\bar{V}(b) = \max_{c,b'} \quad \left\{ u(c) + \rho \left[\pi \bar{V}\left(b'\right) + (1-\pi)\underline{V}\left(b'\right) \right] \right\} \tag{BAH}$$
$$s.t. \quad b' = (1+r)\,b + \bar{y} - c$$
$$b' \geq \underline{b}$$

Sign of Saving: Intuition

- Euler Equation:

$$\bar{V}_b(b) = \rho \cdot (1+r) \cdot \left[\pi \bar{V}_b\left(b'\right) + (1-\pi)\, \underline{V}_b\left(b'\right)\right]$$

- Assuming concavity, if $\pi = 1$:
  - $\rho \cdot (1+r) = 1 \rightarrow b' = b \rightarrow s = 0$
  - $\rho \cdot (1+r) < 1 \rightarrow b' < b \rightarrow s < 0$
  - $\rho \cdot (1+r) > 1 \rightarrow b' > b \rightarrow s > 0$

Concavity of Value Function

Assumption 1

Individual's value functions ($\underline{V}(\cdot)$ and $\bar{V}(\cdot)$) are concave.

- Straightforward to show in BAH, but not possible with shocks to continuation value

Sign of Saving: Intuition

- More generally, for current wealth b, future wealth satisfies $b' > b$ if—and only if:

$$(1-\pi)\left[\frac{\underline{V}_b(b)}{\bar{V}_b(b)}\right] > \frac{1}{\rho \cdot (1+r)}$$

- Intuition: If $b' = b$, if we end-up in the low state in next period:
    - marginal value of wealth is too high
    - $\rightarrow$ We need to increase b' relative to b (decrease $\underline{V}_b$)

Sign of Saving

Lemma 1

Under assumption 1, when current wealth is b, it is optimal to save in the high state if, and only if:

$$\left[\frac{\bar{c}\,(b)}{\underline{c}\,(b)}\right]^{\sigma} > \left[\frac{\rho^{-1}(1+r)^{-1}-\pi}{1-\pi}\right]. \tag{1}$$

- Key point: $\rho \cdot R < 1 \Rightarrow RHS > 1$

Sign of Saving in BAH

- Similar to Theorem 2 in Huggett (1993):

Theorem 2

As long as $\rho \cdot R < 1$, in BAH set up:

$$\lim_{b \to \infty} \frac{\bar{c}(b)}{\underline{c}(b)} \to 1$$

- Therefore, by lemma (1), BAH does not generate positive savings in the upper tail
- **Key:** Consumption smoothing.

A Theoretical Framework with Marginal Utility Shocks (MUS)

- **Observation:** *Any* economy—that does not involve consumption smoothing in optimum—can generate excessive saving at upper tail
- Similar to BAH, replace income shocks with marginal utility shocks:
  - Example: Chronic health shocks
  - Known in health economics as "high utilization state"
  - "Almost" absorbing state in health care data, with high probability of separation (*i.e.*, extremely high mortality rate)

A Theoretical Framework with MUS

- Individuals are heterogeneous in health status $h \in \mathcal{H} = \{\underline{h}, \bar{h}\}$, and asset holdings $b \in [-\underline{b}, \infty)$
  - Will keep referring to current health states as low and high states
- A high-state individual (with current $h = \bar{h}$):
  - Decides how much to consume (c) and save (b')
  - Transitions to low state in next period with probability π
- Low state is absorbing
- In low state:
  - Decides how much to spend on health (m), in addition to consumption and saving
  - Continuation value is further discounted by $\chi(m)$
    - Becker et al. (2005); Hall and Jones (2007)'s interpretation: Individual dies wpr. $1 - \chi(m)$

A Theoretical Framework with MUS

- Same CRRA utility function, with additive term:

$$u(c) = v + \frac{c^{1-\sigma}}{(1-\sigma)}$$

- Utility upon death is normalized to zero: $u^d = 0$
- Value of being alive, v :
  - determines statistical value of life
  - ensures $u(\underline{c}) > u^d$ for some minimal consumption
- Health production function, $\chi : \mathbb{R}_+ \to [0, 1]$:

$$\chi(m) = 1 - (\alpha m + \underline{h})^{-\beta}$$

Individual's Problem in High State

- Identical to individual's problem in BAH:

$$\bar{V}(b) = \max_{c,b'} \quad \left\{u(c) + \rho\left[\pi \bar{V}\left(b'\right) + (1-\pi)\,\underline{V}\left(b'\right)\right]\right\} \qquad \text{(MUS-H)}$$
$$s.t. \quad b' = (1+r)\,b + y - c$$
$$b' \geq \underline{b}$$

Individual's Problem in Low state

$$\underline{V}(b) = \max_{c,m,b'} \quad \left\{u(c) + \rho \cdot \chi(m) \cdot \underline{V}\left(b'\right)\right\} \tag{MUS-L}$$

$$s.t. \quad b' = (1+r)\,b + y - c - m$$

$$b' \geq \underline{b}$$

Theorem 2

Theorem 3

Under assumption 1, saving rate in problem (MUS-H) *cannot be negative above some threshold* $\bar{b}$*, when* $\sigma > (1+\beta)$.

Sign of Saving: Intuition

- Problem (MUS-H) is identical to high-state individual's problem in BAH $\rightarrow$ Lemma (1) applies
- $\rightarrow$ If there is some $\hat{b}$ after which consumer dissaves, must have:

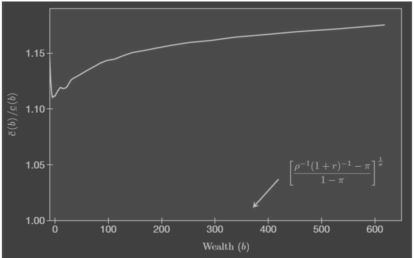
$$\lim_{b \to \infty} \frac{\bar{c}(b)}{\underline{c}(b)} < \left[\frac{\rho^{-1}(1+r)^{-1} - \pi}{1-\pi} \right]^{1/\sigma}$$

I.e., consumption in low state increases faster than wealth

- When $\sigma > (1+\beta)$, m is luxury good in low state:
  - *I.e.*, its share in income increases faster than consumption
- $\rightarrow \underline{c}(b) + m(b)$ explodes as $b \rightarrow \infty$
- Borrowing constraint binds for high enough b, violating Euler equation
- **Key:** Health spending is a luxury good

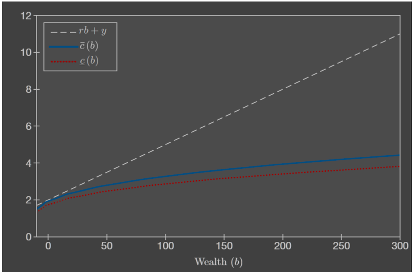
Sign of Saving: Intuition

Consumption Ratio



Panel (a). Consumption in high state to low state

Consumption vs Wealth



Panel (b). Consumption in two states vs income

Note: For $\bar{c}(b)/\underline{c}(b)$, centred moving averages are drawn to smooth out high-frequency fluctuations at lower wealth levels due to numerical error.

Stationary Distribution

- Assume measure ψ_0 of healthy individuals enter with wealth distribution of φ_0 in each period
- Under theorem (3), MUS results in stationary distribution of wealth with Pareto upper tail

Lemma 4

Any (discrete-time,) heterogeneous-agent economy with random separation and entry has a stationary state-distribution.

Theorem 5

When $b' \geq (1+\gamma)b$, for some $\gamma > 0$, the stationary distribution of wealth follows Zipf's law at the upper tail.

Outline

1. Introduction
2. Motivation
3. Theoretical Framework
4. Empirical Model
5. Conclusion

An Empirical Framework

- Generalized MUS framework, with income shocks and policy
- Can write individual's problem recursively as:

$$V(b,y,h) = \max_{c,m,b'} \left\{ u(c) + \rho \cdot \chi(\hat{m},h) \cdot \mathbb{E}_{y',h'} \left[V\left(b',y',h'\right) \right] \right\}$$

$$\text{s.t.} \quad b' = (1+r)b + y - c - m$$

$$\hat{m} = \begin{cases} m + \underline{m}, & \textit{if } h = \underline{h} \\ m, & \textit{if } h = \bar{h} \end{cases}$$

$$b' \geq -\underline{b}$$

- We solve for optimal saving rate as function of current wealth:

$$s^*(b,y,h) := rb + y - c^*(b,y,h) - m^*(b,y,h)$$

Calibration (in progress)

- Functional forms and parameters:
  - $\rho < 1$ and $r = 0.03$ (so that $\rho \cdot \chi < (1 + r)^{-1}$ regardless of m and h):
  - $v = 20$ and $\sigma = 2$
  - $\chi(m, h) = 1 - (\alpha m + h)^{-\beta}$, for some $\alpha, \beta \geq 0$
  - $\mathcal{Y} = \{y_l, y_h\} := \{0.5, 1.5\}$ and $\Pi^y\,(y, y') = 0.2$ when $y \neq y'$
  - $\mathcal{H} = \{h_l, h_h\}$ and $\Pi^h\,(h_h, h_l) = 0.1$, for some $h_l \leq h_h$
  - $\underline{b} = 10$ (so that $r\underline{b} + y_l = \underline{c} > 0$)
- Compare two cases where there are no health shocks (low health) and where we have health shocks (high health)

Outline

Concluding Remarks

- Workhorse heterogeneous-agent models cannot generate fat tails
- We identify a necessary and sufficient condition for this
- We present an example framework where this condition can be violated
  - The mechanism is through marginal utility shock:
    - Individuals can alleviate effect of shock via health spending
    - Health spending is a luxury good
  - We show how this framework can generate stationary distribution of wealth with Pareto tail
- Document saving patterns in HRS that are consistent with our claim
  - Calibrate model to HRS data to see how much of saving at upper tail are attributable to marginal utility shocks

Concluding Remarks

- Future Work: Couples vs singles
  - Why do couples save more than singles in HRS?
  - De Nardi et al. (2023): Bequests can account for part of difference
- We show same mechanism, accounting for possibility of shock to each partner, can account for difference:
  - Couples treat marginal utility shock, differently
  - Higher combined value of life, makes it more valuable to spend on health
  - $\rightarrow$ Positive saving for couples, and negative saving for singles at upper tail of wealth distribution

References I

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics 109*(3), 659–684.

Bach, L., L. E. Calvet, and P. Sodini (2017). From saving comes having? disentangling the impact of saving on wealth inequality. *Disentangling the Impact of Saving on Wealth Inequality (December 16, 2017). Swedish House of Finance Research Paper* (18-8).

Becker, G. S., T. J. Philipson, and R. R. Soares (2005). The quantity and quality of life and the evolution of world inequality. *American economic review 95*(1), 277–291.

Benhabib, J., A. Bisin, and M. Luo (2017). Earnings inequality and other determinants of wealth inequality. *American Economic Review 107*(5), 593–597.

Benhabib, J., A. Bisin, and S. Zhu (2015). The wealth distribution in bewley models with capital income risk. *Journal of Economic Theory 159*, 459–515.

Bewley, T. (1977). The permanent income hypothesis: A theoretical formulation. *Journal of Economic Theory 16*(2), 252–292.

De Nardi, M., E. French, J. Bailey Jones, and R. McGee (2023). Why do couples and singles save during retirement? household heterogeneity and its aggregate implications.

Eslami, K. (2019). *Essays in Health Economics*. Ph. D. thesis, University of Minnesota.

References II

Hall, R. E. and C. I. Jones (2007). The value of life and the rise in health spending. *The Quarterly Journal of Economics 122*(1), 39–72.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of economic Dynamics and Control 17*(5-6), 953–969.

Sargent, T. J. and L. Ljungqvist (2012). Recursive macroeconomic theory. *Massachusetss Institute of Technology*.

Outline

6. Appendix
   - proof of Theorem 1
   - Proof of Theorem 2

Proof of Theorem 1

- Define total available resources in each state as:

$$\bar{\omega} = R \cdot b + \bar{y} + \underline{b}$$
$$\underline{\omega} = R \cdot b + \underline{y} + \underline{b}$$

- It is easy to show policy functions are increasing.
- From Budget constraint::

$$\begin{aligned} c(\bar{\omega}) - c(\underline{\omega}) &= \bar{y} - \underline{y} - \left(b'\left(\bar{\omega}\right) - b'\left(\underline{\omega}\right)\right) \\ &\leq \bar{y} - \underline{y} \end{aligned}$$

For $h(\omega) \in [0,1]$ we have:

$$c(\bar{\omega}) = c(\underline{\omega}) + h(\omega)\left[\bar{y} - \underline{y}\right] \tag{2}$$

Proof of Theorem 1

- With our notation 2 is:

$$\bar{c}(b) \quad = \underline{c}(b) + h(b)\left[\bar{y} - \underline{y}\right]$$

- There fore in any BAH setup:

$$\lim_{b\to\infty} \frac{\bar{c}\,(b)}{\underline{c}\,(b)} = \lim_{b\to\infty} 1 + \frac{\bar{y} - \underline{y}}{\underline{c}(b)} = 1$$

Proof of Theorem 2

- Define total available resources in each state as:

$$\omega = R \cdot b + y + \underline{b}$$

- Assuming concavity, it is easy to show policy functions are increasing.
- From Budget constraint:

$$\begin{aligned} \bar{c}(\omega) - \underline{c}(\omega) &= m(\omega) - \left(\bar{b}'(\omega) - \underline{b}'(\omega)\right) \\ &\leq m(\omega) + \underline{b}'(\omega) - \underline{b} \end{aligned}$$

Proof of Theorem 2

- Therefore, at the boundry:

$$\lim_{\omega\to\infty} \frac{\bar{c}\,(\omega)}{\underline{c}\,(\omega)} = 1 + \lim_{b\to\infty} \frac{m(\omega)}{\underline{c}\,(\omega)} + \lim_{\omega\to\infty} \frac{\underline{b}'(\omega)}{\underline{c}\,(\omega)}$$

- if there is a threshold $\hat{\omega}$ after which $\underline{b}'(\omega) \leq \underline{b}'(\hat{\omega})$ at the boundry:

$$\lim_{\omega\to\infty} \frac{\bar{c}\,(\omega)}{\underline{c}\,(\omega)} = 1 + \lim_{\omega\to\infty} \frac{m(\omega)}{\underline{c}\,(\omega)}$$

Proof of Theorem 2

- From F.O.C we have:

$$\underline{c}(\omega) = \frac{1}{\rho \cdot \chi(m(\omega)) \cdot \underline{V}\left(Rb'(\omega) + y\right) + \mu}^{\frac{1}{\sigma}}$$

- denote $\bar{\omega} = Rb'(\hat{\omega}) + y$, since $Rb'(\omega) + y \leq \bar{\omega}$:

$$\lim_{\omega \to \infty} \underline{c}(\omega) \leq \frac{1}{\rho \cdot \underline{V}\left(\bar{\omega} + \mu\right)}^{\frac{1}{\sigma}}$$

- since $m(\omega) = \omega - \underline{b}(\omega) - \underline{c}(\omega)$:

$$\lim_{\omega \to \infty} \frac{m(\omega)}{\underline{c}\left(\omega\right)} \to \infty$$

- Which implies condition 1 holds, and contradict the existence of any such $\hat{w}$.