Rising Income Risk at the Top and Falling Interest Rates: Evidence from 50 Years of Tax Returns

J. Carter Braxton Wisconsin
Kyle Herkenhoff Minnesota
Chengdai Huang NYU
Michael Nattinger Wisconsin
Jonathan Rothbaum Census
Lawrence Schmidt MIT-Sloan

May 1, 2025

Introduction

Motivation. Large debate over two questions

(1) Has permanent income risk increased in U.S.? Bloom et al '17, Moffitt Zhang '20, Braxton et al '24

(2) Has rising income risk contributed to falling risk-free rate? Kaymak Poschke '16, Auclert et al '20

Introduction

Motivation. Large debate over two questions

(1) Has permanent income risk increased in U.S.? Bloom et al '17, Moffitt Zhang '20, Braxton et al '24

(2) Has rising income risk contributed to falling risk-free rate? Kaymak Poschke '16, Auclert et al '20

This paper. Measure income risk *over* time & *across* income distribution using 1040s

- Validate risk at top with measures of distress (bankruptcy, early 401k withdrawal, etc.)
- Estimate savings response to income risk using capitalized interest & dividend income

Introduction

Motivation. Large debate over two questions

(1) Has permanent income risk increased in U.S.? Bloom et al '17, Moffitt Zhang '20, Braxton et al '24

(2) Has rising income risk contributed to falling risk-free rate? Kaymak Poschke '16, Auclert et al '20

This paper. Measure income risk *over* time & *across* income distribution using 1040s

- Validate risk at top with measures of distress (bankruptcy, early 401k withdrawal, etc.)
- Estimate savings response to income risk using capitalized interest & dividend income

Empirical Findings.

i. Permanent income risk increased 67% for high income (P95+) since 1970s

ii. Rising permanent income risk associated with greater savings at household level

Introduction

Motivation. Large debate over two questions

(1) Has permanent income risk increased in U.S.? Bloom et al '17, Moffitt Zhang '20, Braxton et al '24

(2) Has rising income risk contributed to falling risk-free rate? Kaymak Poschke '16, Auclert et al '20

This paper. Measure income risk *over* time & *across* income distribution using 1040s

- Validate risk at top with measures of distress (bankruptcy, early 401k withdrawal, etc.)
- Estimate savings response to income risk using capitalized interest & dividend income

Empirical Findings.

i. Permanent income risk increased 67% for high income (P95+) since 1970s

ii. Rising permanent income risk associated with greater savings at household level

Quantitative Experiment: Estimate quantitative Bewley model to match (i.)-(ii.)

- ↑ permanent risk since 1970s lowered risk-free rate by 0.6pp (≈30% of observed decline)

Method

Standard income process

- $y_{i,t}$ is residual log income (e.g. AGI, WSI *more later*)
- $z_{i,t}$ is permanent component of income (random walk)
- $\omega_{i,t}$ is transitory component of income

$$y_{i,t} = z_{i,t} + \omega_{i,t} \qquad \mathbb{E}(\omega_{i,t}^2) = R_t$$
$$z_{i,t} = z_{i,t-1} + \nu_{i,t} \qquad \mathbb{E}(\nu_{i,t}^2) = Q_t$$

Method

Standard income process

- $y_{i,t}$ is residual log income (e.g. AGI, WSI *more later*)
- $z_{i,t}$ is permanent component of income (random walk)
- $\omega_{i,t}$ is transitory component of income

$$y_{i,t} = z_{i,t} + \omega_{i,t} \qquad \mathbb{E}(\omega_{i,t}^2) = R_{t,r}$$
$$z_{i,t} = z_{i,t-1} + \nu_{i,t} \qquad \mathbb{E}(\nu_{i,t}^2) = Q_{t,r}$$

Two innovations in this paper

(1) Shocks can depend flexibly on lagged rank in income distribution (e.g. b-spline) ▸ Full

$$r = \text{Rank}_t(y_{i,t-x}, \text{age}_{i,t})$$

(2) Incorporate information about income risk from non-employed/non-filers

Empirics

Data. 1040 tax returns from 1969 to 2019, every 5 years

(1) 5% random sample from Internal Revenue Service (IRS) tax return data *[appendix]*

(2) Current Population Survey (CPS) linked to IRS data ($\approx 10mil$ obs) *[today]*

Empirics

Data. 1040 tax returns from 1969 to 2019, every 5 years

(1) 5% random sample from Internal Revenue Service (IRS) tax return data *[appendix]*

(2) Current Population Survey (CPS) linked to IRS data ($\approx 10\textit{mil}$ obs) *[today]*

Restrictions. (GKOS '21, BHRS '24)

- Non-employed/non-filer threshold of $10,000 (2005 PCE dollars)
- Must earn more than 10k twice between 1969 and 2019
- Half of observations must exceed 10k between first & last year above cutoff

Empirics

Data. 1040 tax returns from 1969 to 2019, every 5 years

(1) 5% random sample from Internal Revenue Service (IRS) tax return data *[appendix]*

(2) Current Population Survey (CPS) linked to IRS data (≈ 10*mil* obs) *[today]*

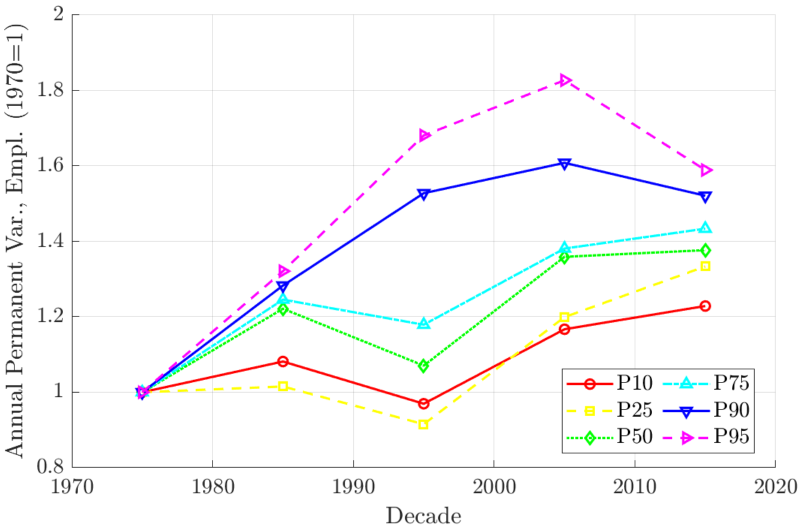
Restrictions. (GKOS '21, BHRS '24)

- Non-employed/non-filer threshold of $10,000 (2005 PCE dollars)
- Must earn more than 10k twice between 1969 and 2019
- Half of observations must exceed 10k between first & last year above cutoff

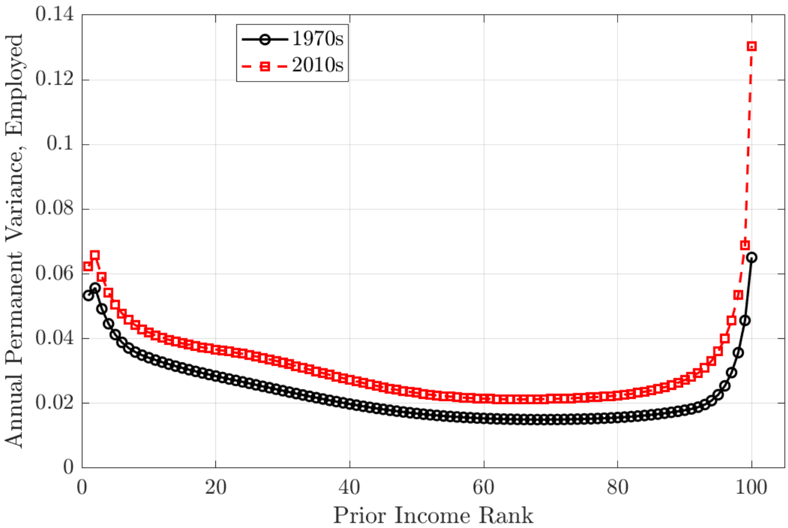
Income.

- Adjusted Gross Income (AGI): wages (WSI)+ business inc.+ interest+ cap gains etc.
- Residualize on age, time, number of adults filing (De Nardi, Fella, Paz-Pardo 2022)
- **Shocks** function of prior $(t-5)$ income rank, decade (e.g, 1970, 1980, ..., 2010) & age

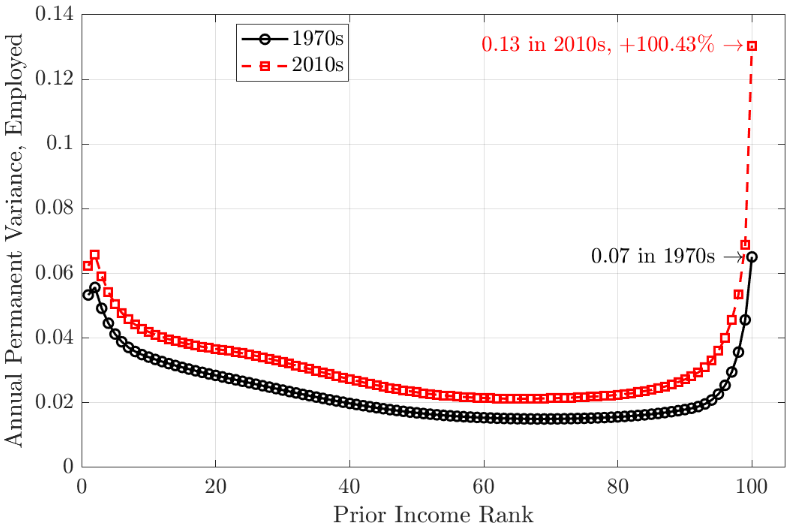
Permanent risk rose most for high income Overlay



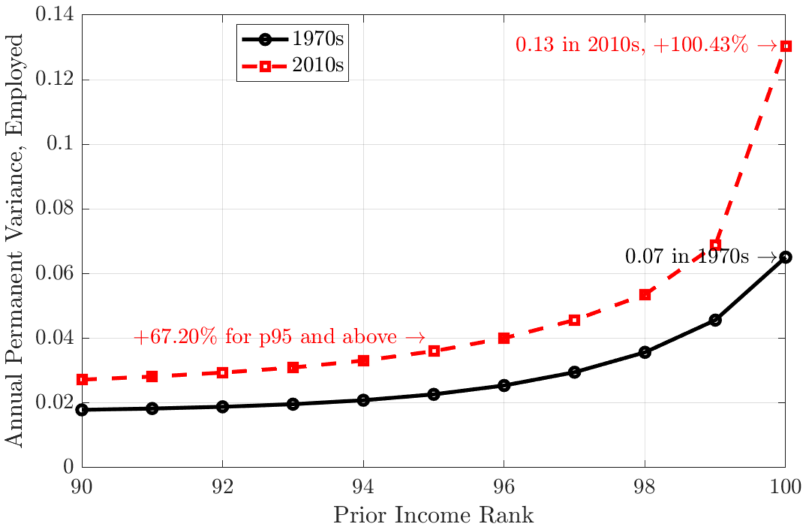
Permanent risk rose most for high income



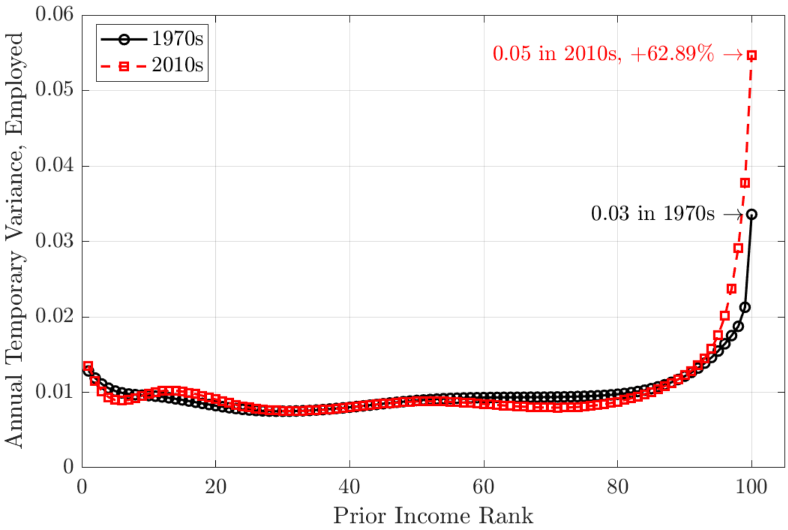
Permanent risk rose most for high income



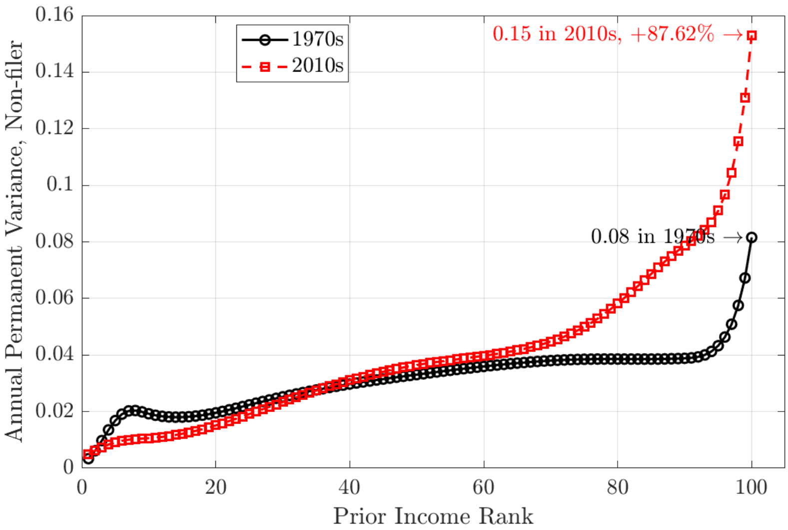
Permanent risk rose most for high income



Transitory risk rose most for high income



Non-filer/non-employed permanent risk rose most for high income



Additional results

Since the 1970s:

(1) *Scarring* effect on permanent earnings from non-filer/non-employed has increased

(2) Initial dispersion in permanent earnings has increased

Additional results

Since the 1970s:

(1) *Scarring* effect on permanent earnings from non-filer/non-employed has increased

(2) Initial dispersion in permanent earnings has increased

Robustness:

(1) Wages and salary income

(2) 5% random sample of 1040s

(3) Alt. minimum earnings

Additional results

Since the 1970s:

(1) *Scarring* effect on permanent earnings from non-filer/non-employed has increased

(2) Initial dispersion in permanent earnings has increased

Robustness:

(1) Wages and salary income

(2) 5% random sample of 1040s

(3) Alt. minimum earnings

Next:

i. It is actually risk at the top?

ii. How does savings respond to changes in permanent income risk?

Time series

Is it risk?

Question: Do those with large permanent shocks behave like “bad” shock occurred?

Is it risk?

Question: Do those with large permanent shocks behave like "bad" shock occurred?

Approach: Our filter produces two extra columns in the tax data $(\widehat{z}_{i,t|T}, \widehat{\omega}_{i,t|T})$

- Identify large declines in permanent income
- Does "economic distress" occur in response to large permanent shocks?
- Even among very high income?

Is it risk?

Question: Do those with large permanent shocks behave like “bad” shock occurred?

Approach: Our filter produces two extra columns in the tax data $(\widehat{z}_{i,t|T}, \widehat{\omega}_{i,t|T})$

- Identify large declines in permanent income
- Does “economic distress” occur in response to large permanent shocks?
- Even among very high income?

Validations: In response to adverse permanent income shocks that we measure, individuals:

- File for bankruptcy Details
- Are foreclosed upon
- Have their debt charged off
- Coming soon: early 401K withdrawals, moving to lower income zipcodes, ...

How does savings respond to changes in income risk?

Goal.

- Discipline linkage between income risk and saving with 'identified elasticity'

  Nakamura, Steinsson (2018), Berger et al (2018), Braxton et al (2024)

Approach.

- Compare saving behavior of individuals with similar levels of earnings and expected growth rates, but different risk profiles

- Exploit within-income-rank, across-occupation variation in risk (CPS)

  Carroll and Samwick (1998), Boar (2022)

How does savings respond to changes in income risk?

Implementation.

(1) Capitalize dividend/interest income (SZZ 2023) to estimate wealth ($b_{i,t}$)

(2) Compute change in wealth to income

$$\Delta b_{i,t+k} = \frac{b_{i,t+k} - b_{i,t}}{Y_{i,t}}$$

(3) Estimate income process for 300+ detailed occupations (o) (e.g., $Q_{o,e,t}$)

(4) Compute savings response to risk at 5-year horizon

$$\Delta b_{i,t+5} = \beta_{Q_E} Q_{o,E,t} + \lambda_r + \lambda_o + \lambda_t + \Gamma X_{i,t} + \epsilon_{i,t}.$$

- Controls: FEs for income-rank (λ_r), 2d-occ. (λ_o), + age profile, mean growth ($B_{E,o,t}$)

Response of savings to income risk

	Dependent variable: 5-year change in wealth to income			
	(1)	(2)	(3)	(4)
Q_E	0.660***			0.645***
	(0.101)			(0.102)
Q_N		0.00292		0.00290
		(0.00296)		(0.00256)
R			0.227***	0.214***
			(0.0738)	(0.0739)
B_E				-0.221
				(0.191)
B_N				-0.00525
				(0.0146)
Year FE	Y	Y	Y	Y
Rank FE	Y	Y	Y	Y
2-Digit Occ. FE	Y	Y	Y	Y
R-squared	0.016	0.016	0.016	0.017
No. Obs	5,042,000	5,042,000	5,042,000	5,042,000

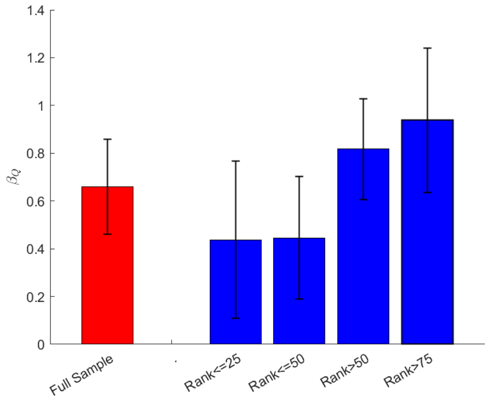
- Interpretation: $\uparrow Q_E$ by 0.1, $\uparrow$ savings ($\Delta b_{i,t+5}$) by 6.6% of prior income

Response of savings to income risk: heterogeneity

- How do high & low income individuals respond to $\uparrow$ in permanent risk?

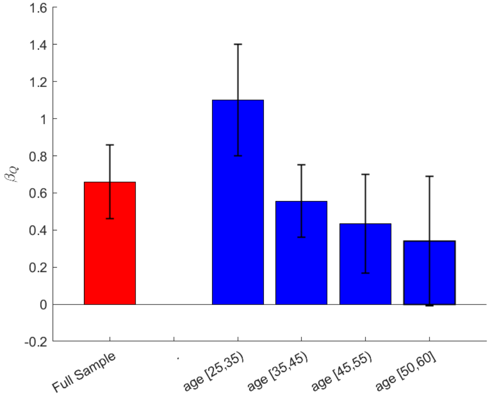
Response of savings to income risk: heterogeneity

- How do high & low income individuals respond to ↑ in permanent risk?
- High earners have larger savings response when permanent risk increases



Response of savings to income risk: heterogeneity

- How do old & young individuals respond to ↑ in permanent risk?
- Younger individuals have larger savings response when permanent risk increases



Taking stock

Main takeaways:

- Permanent income risk has increased by more than 67% for P95+ since 1970s
- Risk is real: negative permanent shocks to high income HHs implies financial distress
- Individuals increase savings when permanent risk increases
  - High income and young have larger response to $\uparrow$ in permanent risk
  - Low income and young have larger response to $\uparrow$ in transitory risk More

Next: How did the rise in permanent income risk among top earnings affect risk-free rate?

- Answer using Bewley-Huggett-Aiyagari model with our income process

Bewley-Huggett-Aiyagari model

- Heterogeneous, risk averse, finitely-lived agents with age $t \in \{1, \dots, T\}$
- *Wage only* income process for z, ω, and employment status e
- Self-insure via saving & borrowing in risk-free asset (b) at rate r_f
- CRRA utility over consumption (σ), CRRA warm-glow bequest motive (σ_b)

$$\max_{\{b_{t+1}\}} \mathbb{E}\left[\sum_{t=1}^{T} \beta^{t-1} \cdot \frac{c_t^{1-\sigma}}{1-\sigma} + \beta^T \cdot \frac{b_{T+1}^{1-\sigma_b}}{1-\sigma_b} \right].$$

subject to shock processes, the natural borrowing limit, and the budget constraint,

$$c_t + b_{t+1} \leq b_t \left(1 + r_f\right) + \underbrace{\lambda w_t \left(z_t, \omega_t, e_t\right)^{1-\nu}}_{\text{HSV after-tax earnings}}$$

Bellman Equations

Mapping Model to Data

Initial steady state: consistent with 1970s labor and asset markets

- Labor income shocks from 1970s & $r_f = 3.5\%$ (Laubach & Williams ('03))

Preferences: Jointly calibrate $(\beta, \sigma, \sigma_b)$ to match

(1) Share of wealth held by p99+ in 1970s *SCF+* Non-targeted: other wealth moments

(2) Elasticity savings to permanent risk (β_Q, pooled) Non-targeted: other savings heterogeneity

(3) Elasticity savings to permanent risk among old (β_Q, 50 to 60yo)

Variable	Value	Target	Model	Data
$\beta^{1/5}$	0.901	Top 1 Percent Wealth Share, 1970s	20.5	23.0
σ	1.411	Savings Regression Coefficient (Q_E)	0.431	0.660
σ_b	1.158	Savings Regression Coefficient (Q_E), Age 50-60	0.175	0.340

Interest rate pressures from rising income risk

Experiments.

- Change profile of risk from 1970s value to 2010s value
- Adjust age profile of earnings to neutralize first-moment effects
- Set r_f to match level of assets in positive net supply from 1970s economy

Interest rate pressures from rising income risk

Experiments.

- Change profile of risk from 1970s value to 2010s value
- Adjust age profile of earnings to neutralize first-moment effects
- Set r_f to match level of assets in positive net supply from 1970s economy

Economy	(1) Baseline	(2) 2010s	(3) $\Delta_t Q_E$	(4) $\Delta_t R$
Risk-free rate	3.5%	2.9%	3.1%	3.4%
Change in r_f vs. baseline	–	-0.6p	-0.4pp	-0.1pp

Conclusion

Contributions.

(1) Has permanent income risk increased in U.S.? Yes, especially for high income.

(2) Has rising income risk contributed to falling risk-free rate? Yes, led to 0.6pp decline.

Conclusion

Contributions.

(1) Has permanent income risk increased in U.S.? Yes, especially for high income.

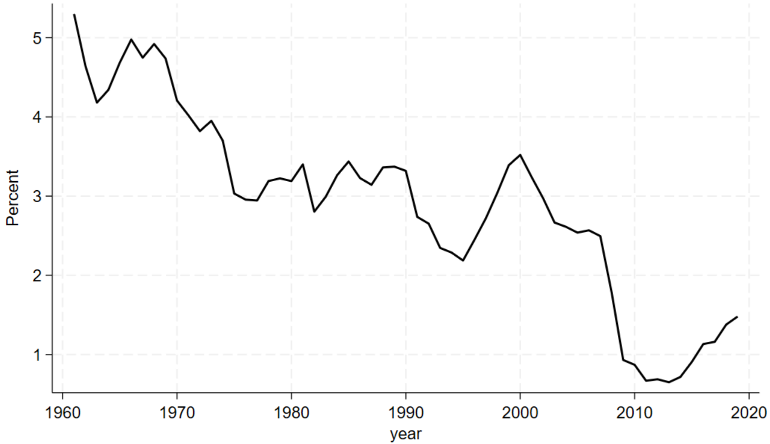
(2) Has rising income risk contributed to falling risk-free rate? Yes, led to 0.6pp decline.

In progress...

- Use linked CPS and credit reports to establish that income losses at the top are associated with adverse events (risk)
- Use additional IVs (e.g., firm risk) to establish that risk is associated with savings
- Discipline model mechanisms better with heterogeneity analysis
- Particle filter, transition dynamics, business income separate

Extra slides

Decline in risk-free rate



Source: Laubach & Williams (2003) – output=potential, inflation stable – KF with $r_t^* = c * trendgrowth_t + z_t$. z_t is AR(1), trend growth and potential output are random walks. Back

Decline in risk-free rate

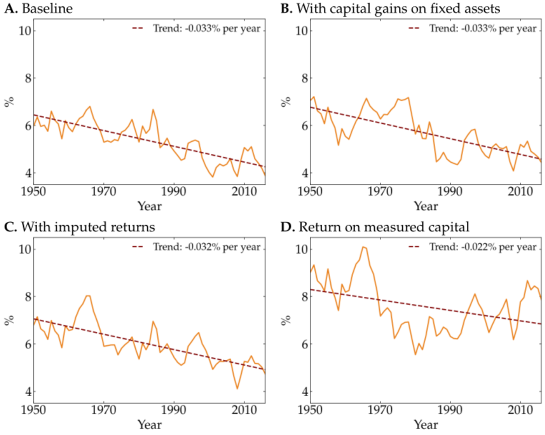
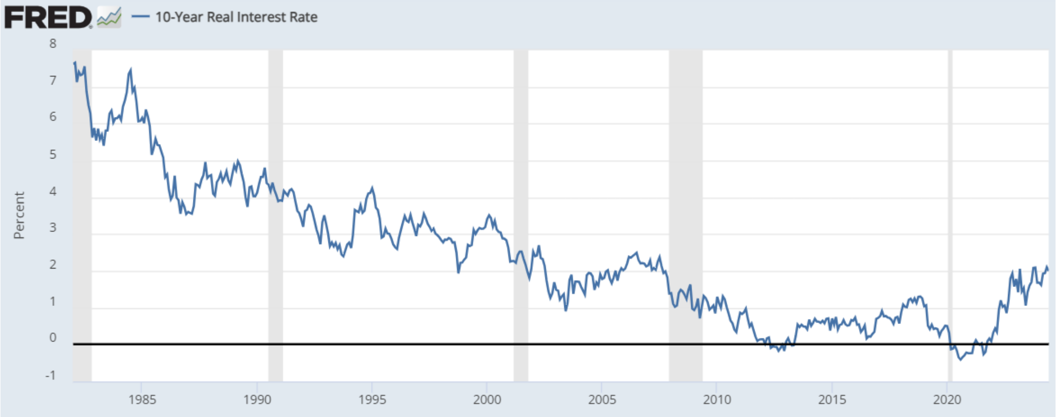


Figure A.1: Alternative ways of constructing the total return on wealth in the US

Notes: Panel A gives our baseline series for the total return on wealth in the US, as described in the text. Panel B adds capital gains on fixed assets, as measured in the fixed assets accounts. Panel C imputes an additional return on unmeasured wealth $W_t - K_t - B_t - NFA_t$ equal to trend growth. Panel D takes our baseline capital income series and divides it by capital measured in the fixed assets accounts.

Source: Auclert, Malmberg, Martenet, Rognlie (2021). Back

Decline in risk-free rate



Source: The Federal Reserve Bank of Cleveland estimates the expected rate of inflation over the next 30 years along with the inflation risk premium, the real risk premium, and the real interest rate. Their estimates are calculated with a model that uses Treasury yields, inflation data, inflation swaps, and survey-based measures of inflation expectations. Back

Literature

- Financial and demographic factors: Favilukis (2013), Farhi and Gourio (2018), Caballero et al. (2008), Caballero et al. (2017), Auclert et al. (2021)
  - Generation 1: macro calibration, single hypothesis (Caballero et al. (2008))
  - Generation 2: model inversion on macro data, multiple hypotheses (Favilukis (2013), Farhi and Gourio (2018))
  - Generation 3: identified empirical elasticity [Our contribution]
- Rising income inequality: Favilukis (2013), Kaymak and Poschke (2016), Auclert and Rognlie (2018), Aladangady et al. (2021), Mian et al. (2021a), and Mian et al. (2021b)
  - **Closest:** Auclert, Rognlie (2018) use jump-drift (Schmidt 2022), equally scale-up the variance of permanent and transitory income – explains one-fifth of r_f decline
  - Our contribution: Estimate permanent rise (and across distribution) using novel method, identified elasticity
- "Arms race" for skewness/kurtosis in income processes: Guvenen et al. (2014), Arellano et al. (2017), De Nardi et al. (2020), and Guvenen et al. (2021)
  - Tractable/scalable (with EM and Kalman), can handle zeros, yields panel of shocks [Our contribution]

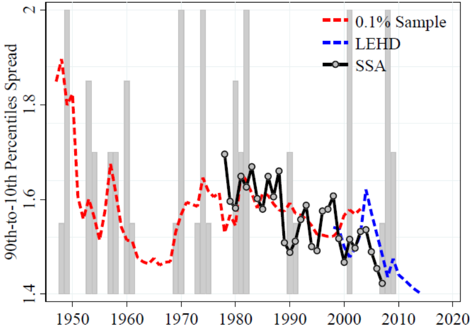
Back

Literature

- CPS-DER has falling variance of 1y and 5y wage changes (Braxton et al 2020)
- Bloom et al 2017/Salgado et al 2023 use wage
- Key differences(1) family income, (2) AGI, and (3) since 1970s income risk may have increased in their sample too

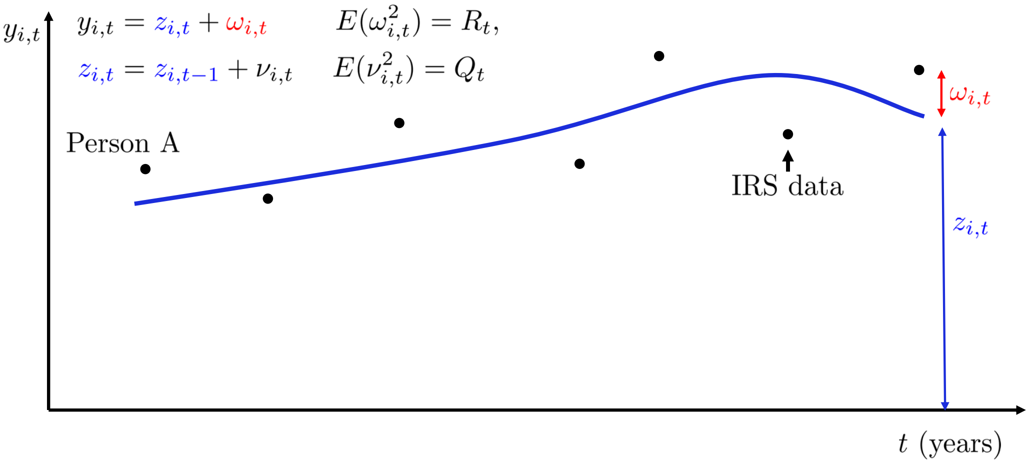
Figure A.1 – Dispersion in Five-year Earnings Growth Declines

(a) Men



Source: Salgado et al 2023. Back

Method

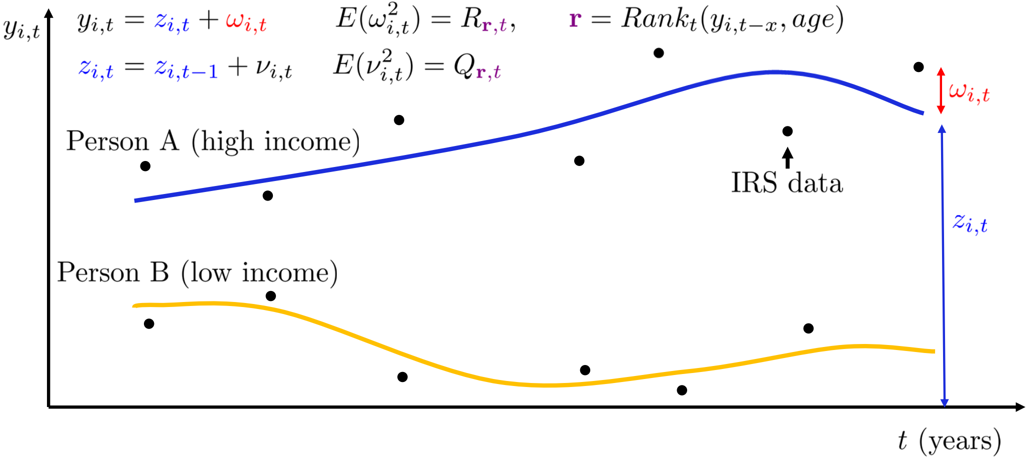


$y_{i,t}$
$y_{i,t} = z_{i,t} + \omega_{i,t}$ $E(\omega_{i,t}^2) = R_t,$
$z_{i,t} = z_{i,t-1} + \nu_{i,t}$ $E(\nu_{i,t}^2) = Q_t$
Person A
IRS data
$\omega_{i,t}$
$z_{i,t}$
t (years)

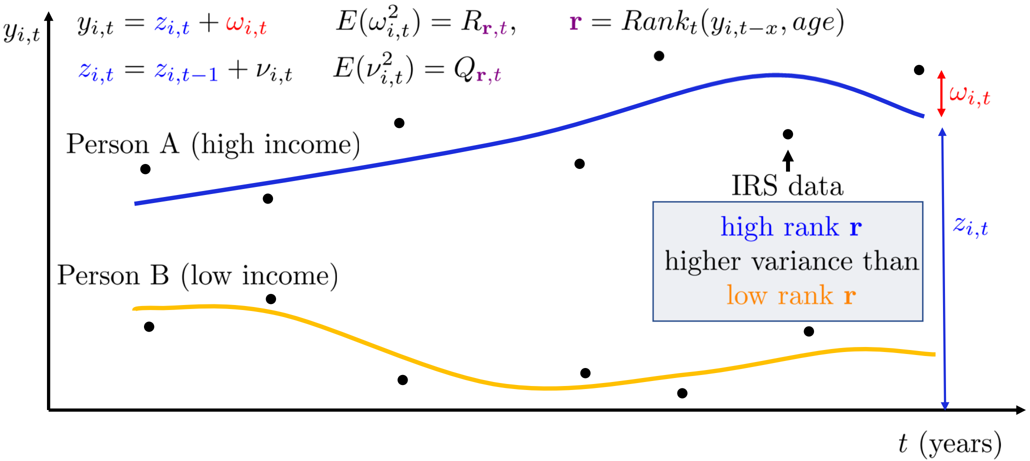
Method: Innovation 1, Shocks by Lagged Income Rank

$$y_{i,t} = z_{i,t} + \omega_{i,t} \qquad E(\omega_{i,t}^2) = R_{\mathbf{r},t}, \qquad \mathbf{r} = Rank_t(y_{i,t-x}, age)$$

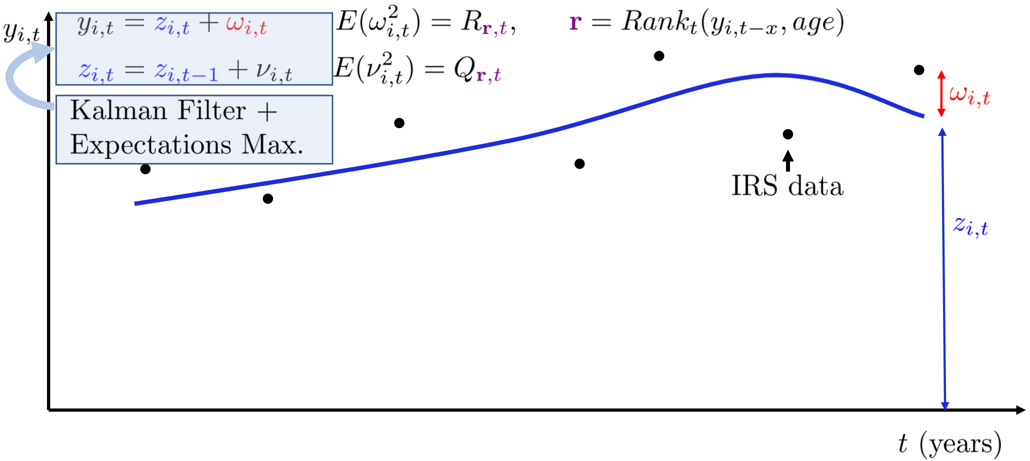
$$z_{i,t} = z_{i,t-1} + \nu_{i,t} \qquad E(\nu_{i,t}^2) = Q_{\mathbf{r},t}$$



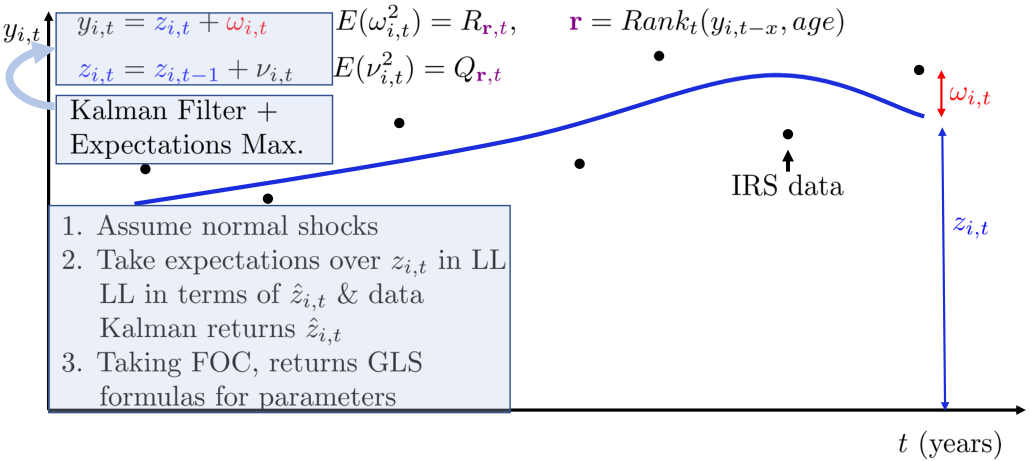
Method: Innovation 1, Shocks by Lagged Income Rank



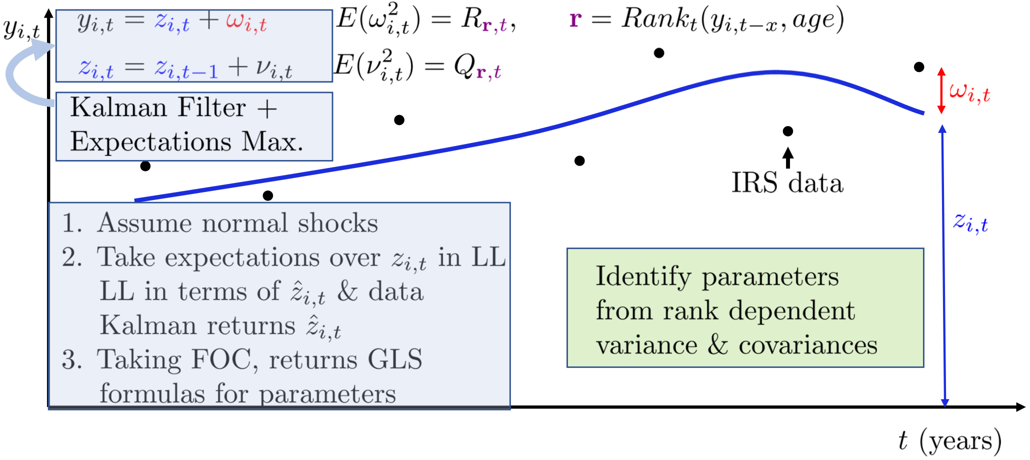
Method: Innovation 1, Shocks by Lagged Income Rank



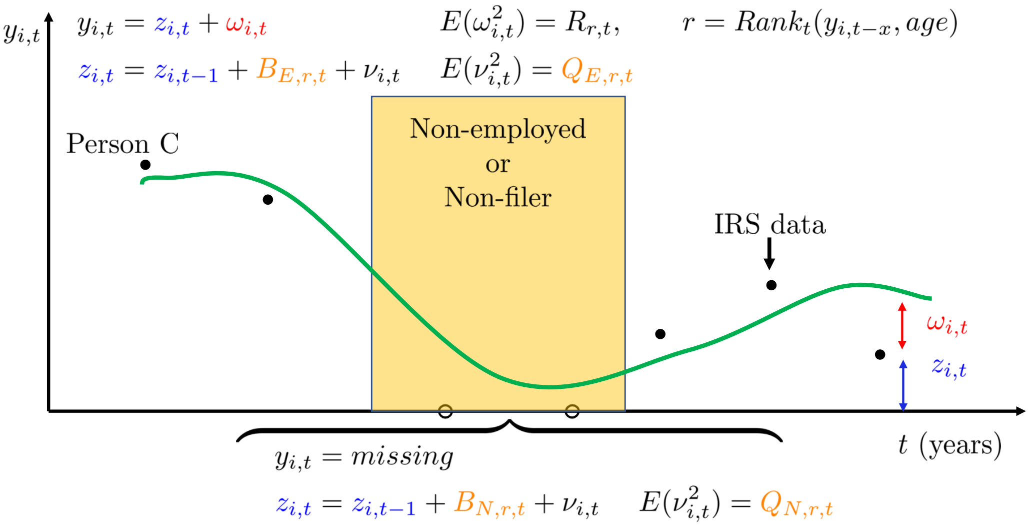
Method: Estimation



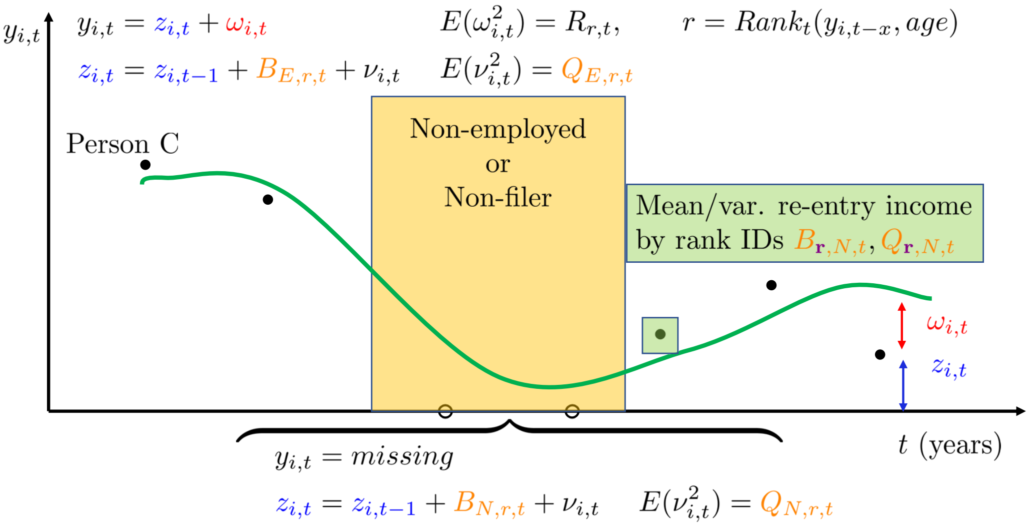
Method: Estimation



Method: Innovation 2, Non-employment/Non-filing



Method: Innovation 2, Non-employment/Non-filing



Overview of algorithm (ignoring filing status)

Given panel of residual log earnings $\{y_{i,t}\}$, we apply these steps:

1. Guess parameters $\{Q_{r,E,t}, R_{r,t}, \ldots\}$
2. Kalman filter yields permanent state forecast $\widehat{z}_{i,t|t-1}$ and its uncertainty $\widehat{M}_{i,t|t-1}$
3. Use conditional normality of $y_{i,t}$ to build partial log-likelihood:

$$y_{i,t}|y_i^{t-1} \sim N(\widehat{z}_{i,t|t-1}, \widehat{M}_{i,t|t-1} + R_{r,t})$$

4. Update parameters
   - Option 1: Global search over permanent risk $Q_{r,E,t}$, transitory risk $R_{r,t}$,... (slow)
   - Option 2: Smooth $\widehat{z}_{i,t|T}$, integrate likelihood ("plug" in $\widehat{z}_{i,t|T}$), update using first order conditions
   - We apply **Option 2**, also called the Expectation-Maximization (EM) algorithm

Rezidualization

Following De Nardi, Fella, Paz Pardo (2022) we residualize income for those above 10k:

$$\log(Y_{i,t}) = \gamma_a + \gamma_t + \gamma_n + y_{i,t}$$

- $Y_{i,t}$ income as reported on 1040
- $y_{i,t}$ (residual) log income for household i in year t
- γ_a age fixed effects
- γ_t year fixed effects
- γ_n indicator for multiple adults in household
- Keep both spouses in data, place weight of one-half on each spouse while filing jointly

Back to data

Full Specification

- Shocks depend on within-age income rank (r) at $t-5$, decade (d) (70s, 80s...) & age (a)
  - Superscript *s=b-spline, superscript q=quadratic*
  - e.g., $B^s_E(r, d)$ is a b-spline in rank r and decade dummies d

$$y_{i,t} = \begin{cases} z_{i,t} + \omega_{i,t} & \omega_{i,t} \sim N\left(0, e^{\widehat{R}^s(r,d)+\widehat{R}^q(a)}\right) \quad \text{if } Y_{i,t} \geq 10k \\ \cdot & \text{otherwise} \end{cases}$$

$$z_{i,t} = \begin{cases} z_{i,t-5} + B^s_E(r, d) + B^q_E(a) + \nu_{i,t}, & \nu_{i,t} \sim N\left(0, e^{\widehat{Q}^s_E(r,d)+\widehat{Q}^q_E(a)}\right) \quad \text{if } Y_{i,t} \geq 10k \\ z_{i,t-5} + B^s_N(r, d) + B^q_N(a) + \nu_{i,t}, & \nu_{i,t} \sim N\left(0, e^{\widehat{Q}^s_N(r,d)+\widehat{Q}^q_N(a)}\right) \quad \text{otherwise} \end{cases}$$

$$z_{i,0} \sim N\left(0, e^{\widehat{Q}_0(d)+\widehat{Q}^q_0(a,d)}\right)$$

Back to methods Back to data

Time aggregation

Problem: Tax data is observed every five years (1969, 1974,...,2014, 2019)

Three Approaches.

1. Iterate income process forward (ignoring non-filers) $z_{i,t} = Fz_{i,t-1} + \omega_{i,t}$

   $z_{i,t} = z_{i,t-5} + \omega_{i,t} + \omega_{i,t-1} + \omega_{i,t-2} + \omega_{i,t-3} + \omega_{i,t-4}$

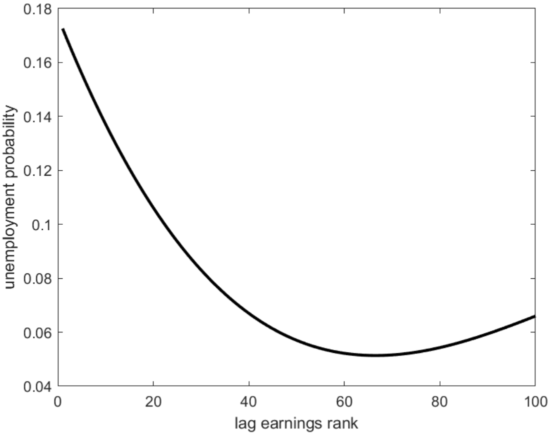
   We estimate

   $z_{i,t} = z_{i,t-5} + \widehat{\omega}_{i,t}$ where $\widehat{\omega}_{i,t} = \omega_{i,t} + \omega_{i,t-1} + \omega_{i,t-2} + \omega_{i,t-3} + \omega_{i,t-4}$
2. Monte Carlos
3. Estimation on annual data period (1996 and onwards, or SSA data)

Back to methods Back to data

AGI, CPS-Tax sample

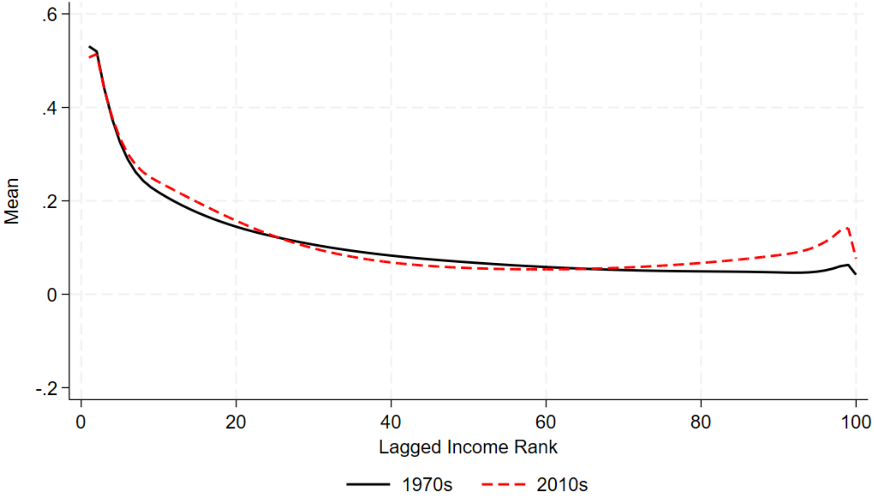
Unemployment profile is downward sloping, with slight U



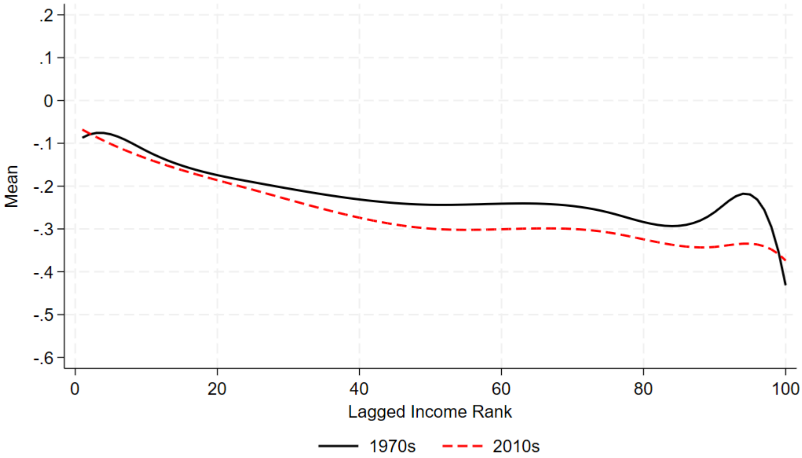
Back

AGI: B_E

Back to Q_E Back to R

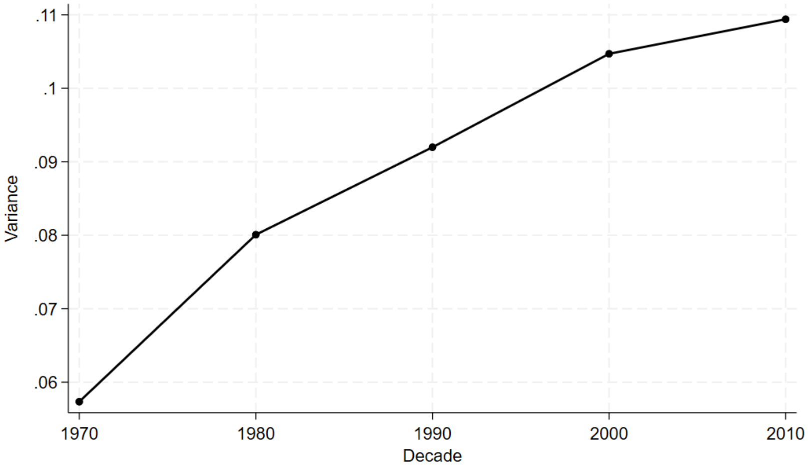


“Scarring” effect of non-employment has worsened



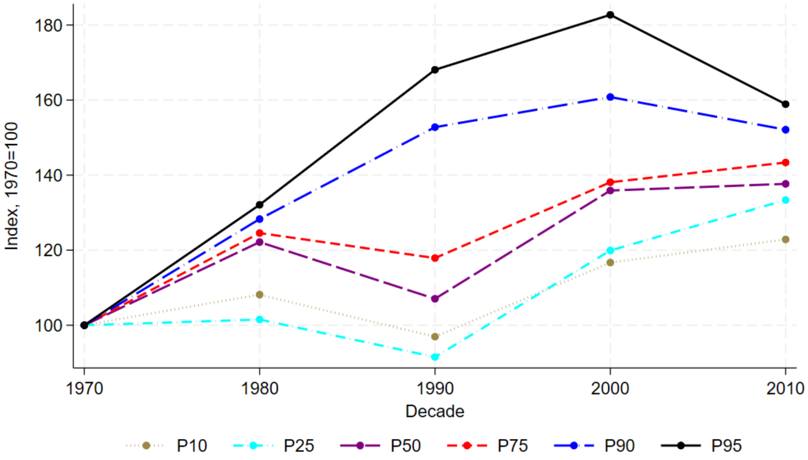
Back to taking stock

Initial dispersion in permanent income has increased



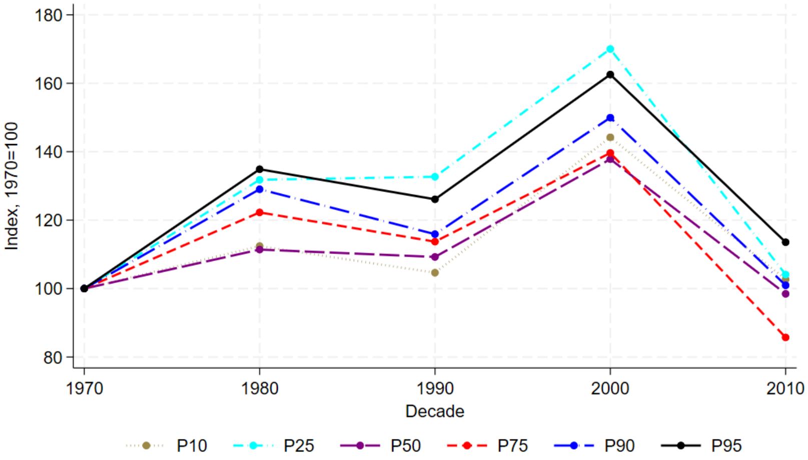
Back to taking stock

Permanent risk has increased the most among high income



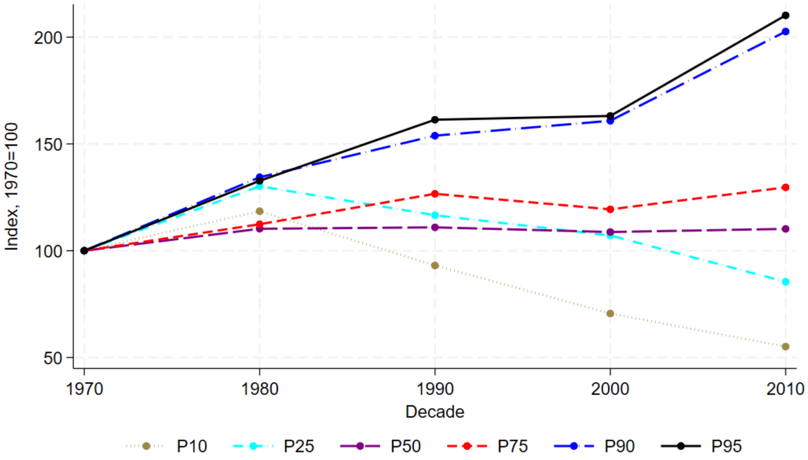
Back to taking stock

Temporary risk has increased at the top of the distribution



Back to taking stock

Permanent risk of non-emp/non-filers inc. the most among high incomes



Back to taking stock

Is it risk?

Empirical Approach: Estimate income process on linked LEHD & TransUnion Credit Reports Back

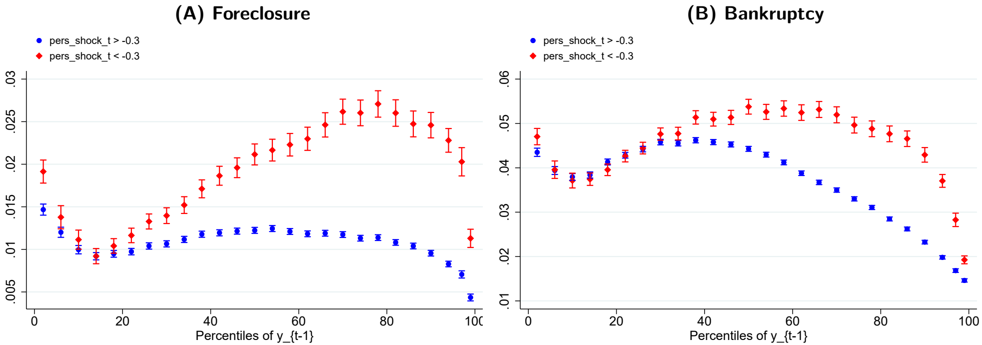
- Split sample into 26 quantiles of prior earnings $Y_{q,i,t}$– top quantile is p98 and above
- Estimate default in next 3 years $D_{i,t}$ as a function of tail shocks (permanent loss$\geq$30%)

$$D_{i,t} = \alpha + \sum_{q=2}^{26} \beta_j Y_{q,i,t} + \sum_{q=1}^{26} \gamma_q Y_{q,i,t} \times \mathbb{I}\{\Delta \widehat{z}_{i,t|T} \leq -.30\} + \Gamma X_{i,t} + \epsilon_{i,t}$$

Is it risk?

Results: In response to a 30% decline in permanent earnings:

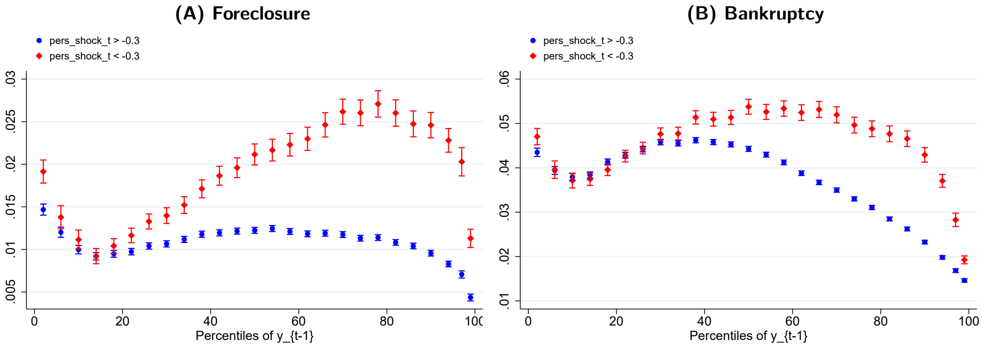
- Foreclosure rate increases More
- Bankruptcy rate increases More



Is it risk?

Results: In response to a 30% decline in permanent earnings:

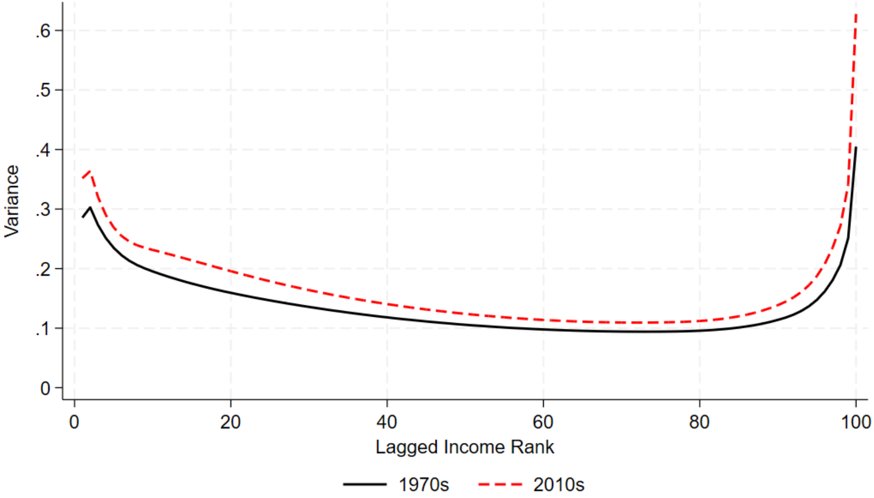
- Foreclosure rate increases More
- Bankruptcy rate increases More



Next: Do households save when permanent income risk rises?

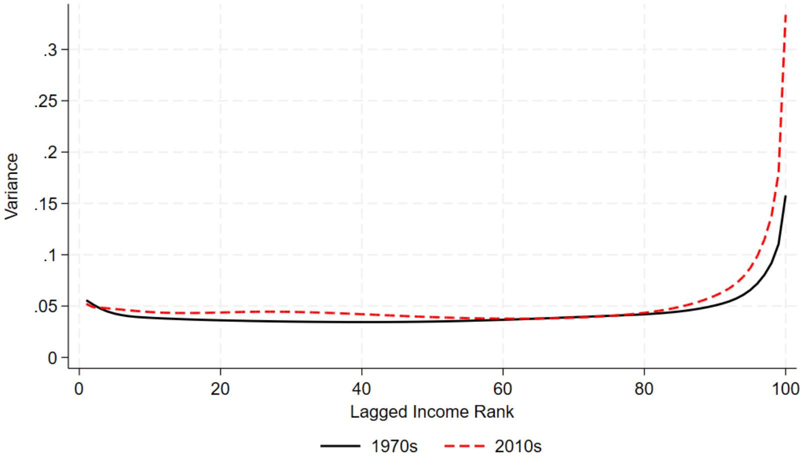
AGI, 5% random sample

AGI: Q_E for 5% random sample



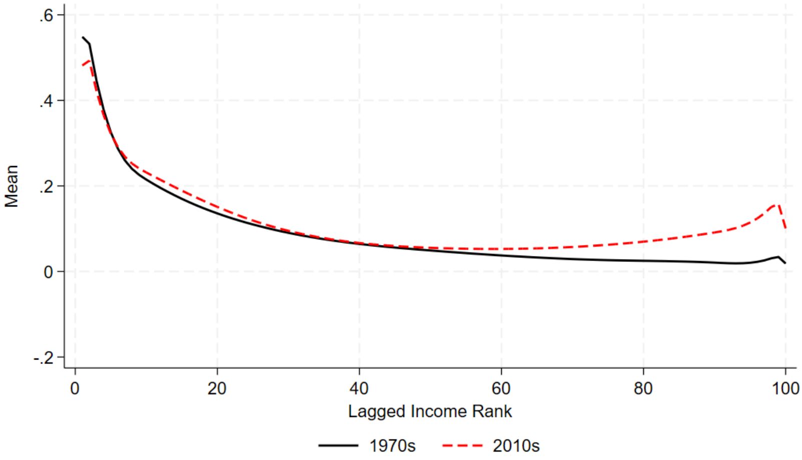
Back to data | Back to Q_E | Back to taking stock

AGI: *R* for 5% random sample



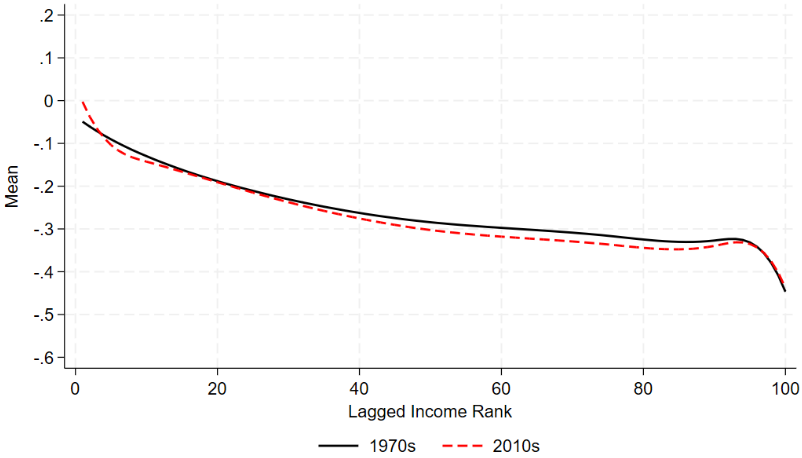
Back to *R* | Back to taking stock

AGI: B_E for 5% random sample



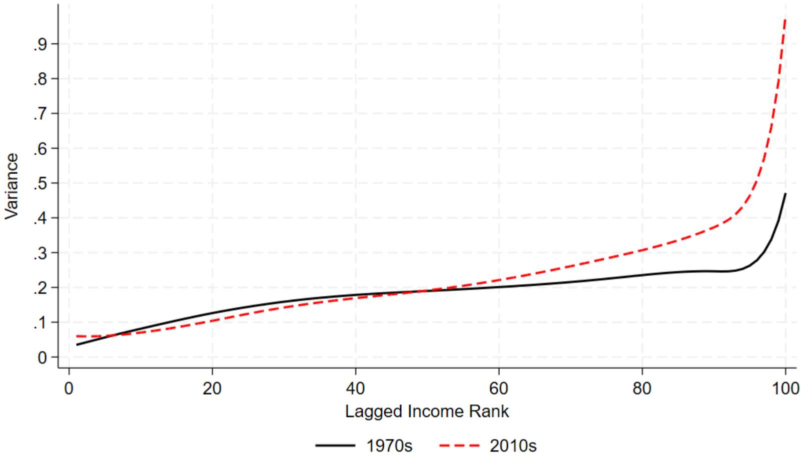
Back to Q_E Back to R Back

AGI: B_U for 5% random sample



Back to Q_E Back to R Back

AGI: Q_U for 5% random sample

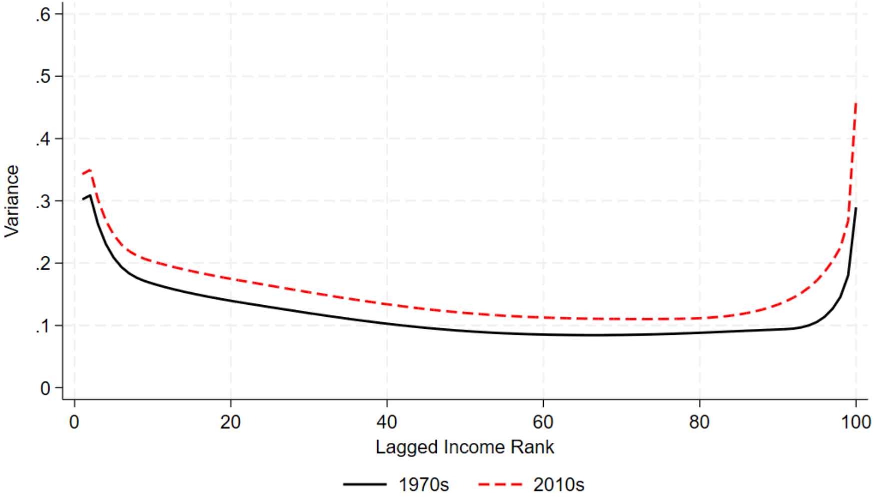


Back to Q_E Back to R Back

WSI, CPS-Tax sample

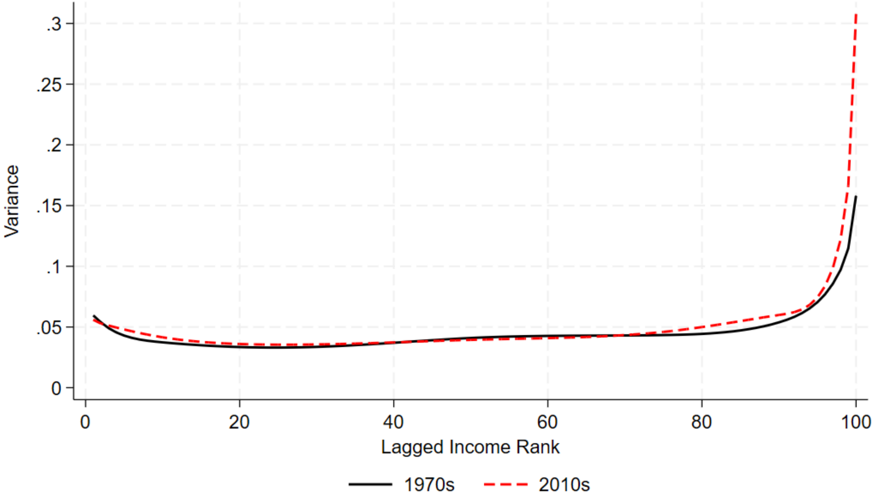
WSI: Q_E

B_E B_U Q_U Back



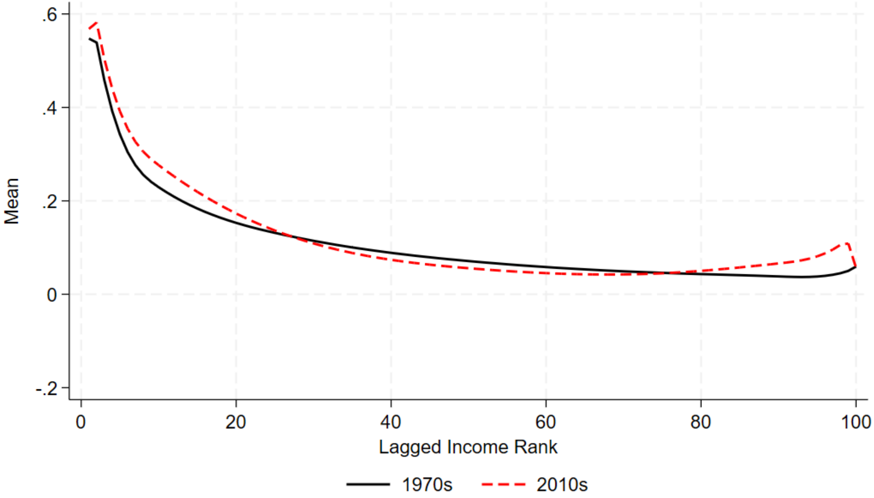
WSI: R

B_E B_U Q_U Back



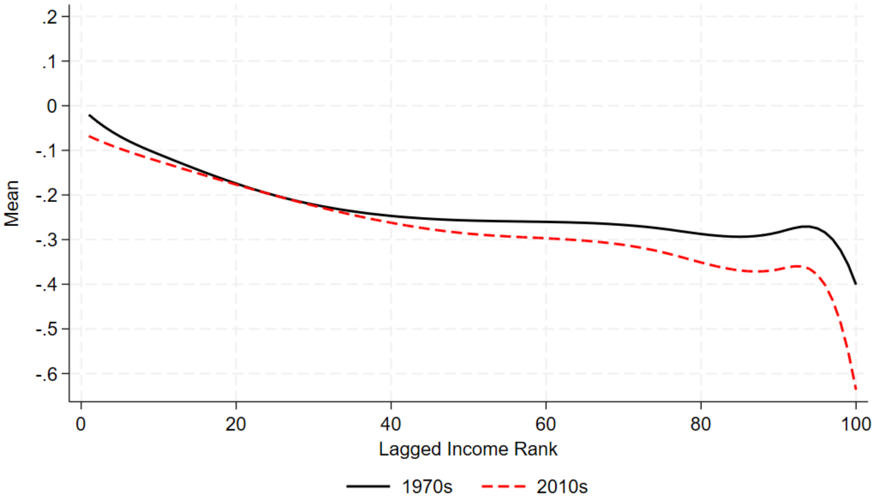
WSI: B_E

Back to Q_E Back to R Back



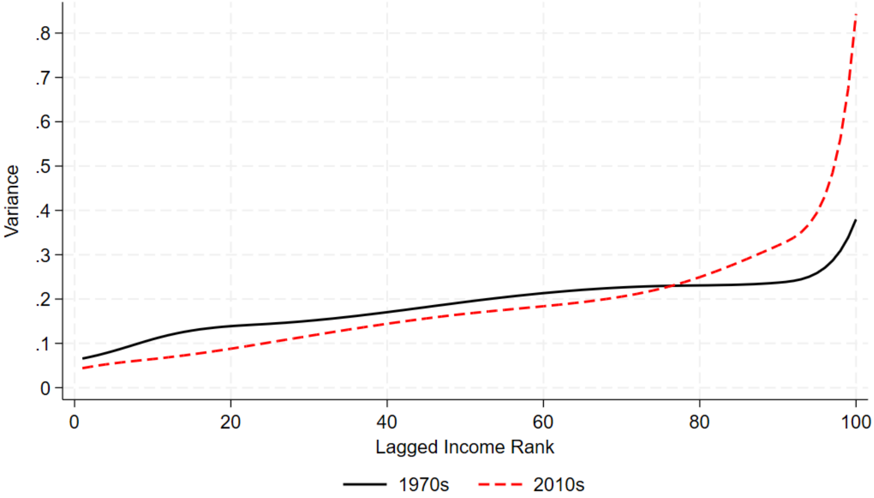
WSI: B_U

Back to Q_E | Back to R | Back



WSI: Q_U

Back to Q_E Back to R Back



Occupation regressions

Occupations

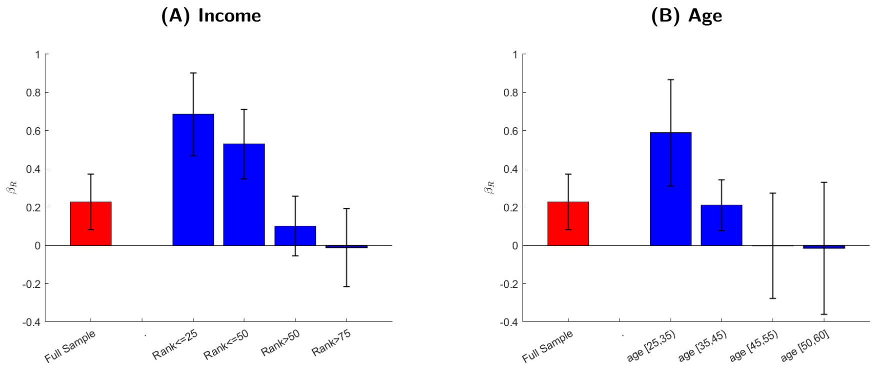
- Identify a household's occupation as their first occupation in March CPS (25+yos)
- Classify occupations using Autor and Dorn (2013) time consistent occupation codes

Occ1990dd Code	Occupation Groups and Occupation Titles	Census 1950 Codes	Census 1960 Codes	Census 1970 Codes	Census 1980 Codes	Census 1990 Codes	Census 2000 (5% Sample) Codes	ACS 2005 Codes
	C. Service Occupations							
	C.1 Housekeeping and Cleaning Occupations							
405	Housekeepers, maids, butlers, and cleaners	700, 751-753, 764	802, 820, 821, 823, 824, 832	901, 902, 940, 941, 950, 982, 984	405, 407, 449	405, 407, 449	423	423
408	Laundry and dry cleaning workers	643, 710	674, 803	611, 630, 983	403, 747, 748	403, 747, 748	830	830
	C.2 Protective Service Occupations							
415	Supervisors of guards				x	x	x	x
417	Fire fighting, fire prevention, and fire inspection occs	x	x	x	x	x	x	x
418	Police and detectives, public service	773	853	964	6, 414, 418	6, 414, 418	371, 382, 384, 385	371, 382, 384, 385
423	Sheriffs, bailiffs, correctional institution officers	771, 782	852, 854	963, 965	423, 424	423, 424	370, 380	370, 380
425	Crossing guards	785	860	960	425	425	394	394
426	Guards and police, except public service	763	851	962	426	426	391, 392	391, 392
427	Protective service, n.e.c.				x	x	x	x

Back

Response of savings to income risk: heterogeneity

- How do individuals respond to ↑ in transitory risk?
- Lower income individuals have larger savings response when permanent risk increases
- Young individuals have larger savings response when permanent risk increases



Non-targeted moments

Savings response to transitory shocks: $\beta_R = 0.227$ (data), $\beta_R = 0.163$ (model) Back

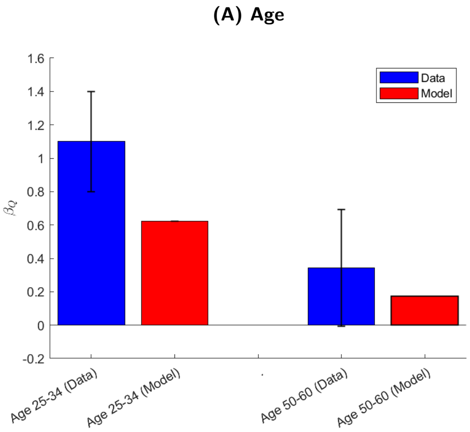
Concentration of wealth in 1970s SCF+

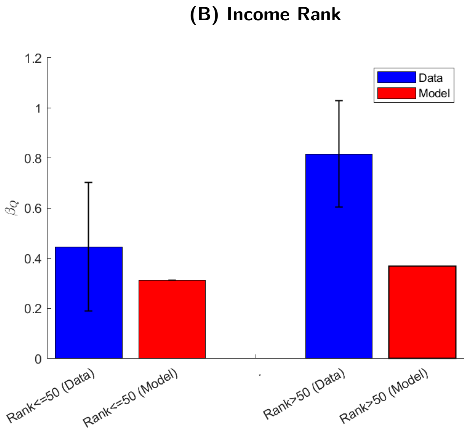
Variable	Model	Data
Top 5 Percent Wealth Share	47.5	49.4
Top 20 Percent Wealth Share	79.2	76.9
Top 40 Percent Wealth Share	97.7	93.3

Consumption and permanent income:

- Straub ('19): $\uparrow$ permanent income 1 log point, $\uparrow$ consumption 0.7 log points
- Model: $\uparrow$ permanent income 1 log point, $\uparrow$ consumption 0.74 log points

Non-targeted moments: saving response to permanent shocks





Back

Capitalizing wealth

We observe: $I_{i,t}^{\text{int}}$ taxable interest income and $I_{i,t}^{\text{div}}$ dividend income

We want to know: $b_{i,t}^{\text{int}}$ fixed asset wealth and $b_{i,t}^{\text{div}}$ C-corp wealth

Following SZZ (2023): For wealth category c, we write $I_{i,t}^{c} = r_{i,t}^{c} b_{i,t}^{c}$

- $r_{i,t}^{c}$ household i's net return on wealth category c in year t
- Given $r_{i,t}^{c}$, $b_{i,t}^{c} = \frac{1}{r_{i,t}^{c}} I_{i,t}^{c}$
- SZZ (2023) give us r_{t}^{div} constant across individuals; $r_{i,t}^{\text{int}}$ by $I_{i,t}^{\text{int}}$ percentile

Back

Model

Agent's Problem

The value function for an age t household with net worth b, permanent income z, transitory income ω in employment status e is

$$V_t(b, z, \omega, e) = \max_{b' \geq \underline{b}} u(c) + \sum_{e'} p(e'|z, \omega, r, e)\, \beta \mathbb{E}\left[V_{t+1}(b', z', \omega', e')\right] \; \forall t \leq T$$

$$V_{T+1}(b, z, \omega, e) = \phi(b)$$

subject to the budget constraint and mapping from income to ranks,

$$c + b' \leq b(1 + r_f) + \lambda w_t(z, \omega, e)^{1-\nu}, \qquad r = Rank_t(z, \omega, e)$$

$$w_t(z, \omega, e) = \begin{cases} \exp(\kappa_t + z + \omega) & \text{if } e = E \\ \gamma \exp(\kappa_t + z) & \text{if } e = N. \end{cases}$$

Additional model parameters
Back to model overview

Model parameters

Variable	Value	Description
r_f	$(1+3.5\%)^5-1$	Risk-free interest rate
α	0.181	Progessivity of tax function
λ	5.620	Location parameter of tax function
γ	0.4	Replacement Rate UI
α_u	0.275	Job finding rate for unemployed/nonfiler
α_0	0.193	Added constant in unemployment probability for employed
α_1	$-.00484$	Linear term in unemployment probability for employed
α_2	$5.38e-05$	Square term in unemployment probability for employed
α_3	$-1.80e-07$	Cube term in unemployment probability for employed

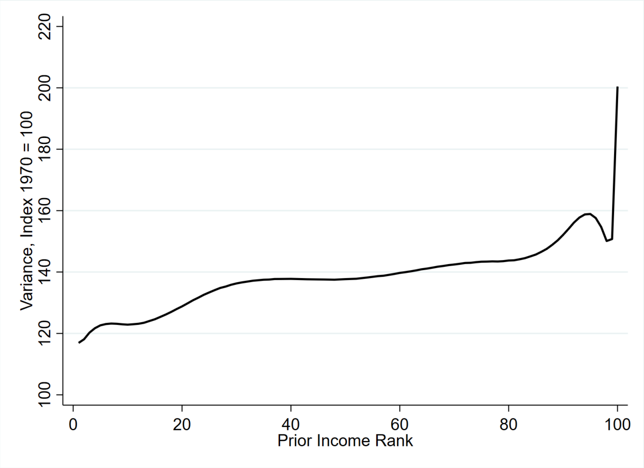
Where we impose the following law of motion for non-employment/non-filer status:

$$p(N|r,E) = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3 r^3 \tag{1}$$

Back to Bellman Equations Back to estimation

By 2010s, permanent risk has increased among top earners

B_E B_U Q_U



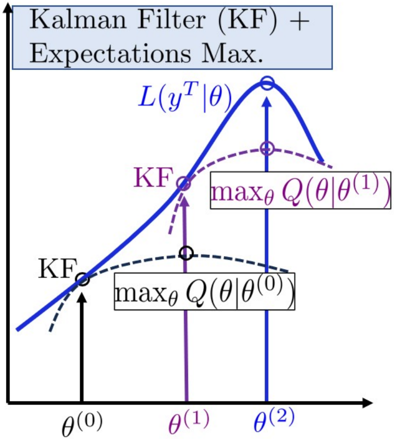
Method

EM: With latent state, hard to directly optimize likelihood

- Log likelihood $L(\theta) = \log p(y|\theta)$
- C is complete data $c = (z, y)$, $y \in C$:
  $L(\theta) = \log \int_C p(y, c|\theta) dc = \log \int_C p(c|\theta) dc$
- Multiply by 1:
  $L(\theta) = \log \int_C \frac{p(c|\theta)}{p(c|y,\theta^{(t)})} p(c|y, \theta^{(t)}) dc = \log E_{c|y,\theta^{(t)}} [\frac{p(c|\theta)}{p(c|y,\theta^{(t)})}]$
- Jensen: $L(\theta) \geq E_{c|y,\theta^t} \log[\frac{p(c|\theta)}{p(c|y,\theta^{(t)})}] = Q(\theta|\theta^{(t)}) + h(\theta^{(t)})$

I. Note when $\theta = \theta^{(t)}$ inner term is constant, no Jensen ineq.
II. Also when $Q(\theta|\theta^{(t)})$ improves, $L(\theta)$ weakly improves

- Use $f(z|y, \theta^{(t)})$ from KF (i.e. this is $E_{c|y,\theta^{(t)}}$)
- I & II let us optimize $Q(\theta|\theta^{(t)})$ More Back



EM details

Proof of property I ($L(\theta) = E_{c|y,\theta^t} \log[\frac{p(c|\theta)}{p(c|y,\theta^{(t)})}]$ when $\theta = \theta^{(t)}$):

$\log p(y|\theta) = L(\theta) \geq E_{c|y,\theta^t} \log[\frac{p(c|\theta)}{p(c|y,\theta^t)}]$

Suppose $p(c|y,\theta^t) = p(c|y,\theta)$

$\log p(y|\theta) = L(\theta) \geq E_{c|y,\theta^t} \log[\frac{p(c|\theta)}{p(c|y,\theta)}]$

$\log p(y|\theta) = L(\theta) \geq E_{c|y,\theta^t} \log[\frac{p(c|y,\theta)p(y|\theta)}{p(c|y,\theta)}]$

$\log p(y|\theta) = L(\theta) \geq E_{c|y,\theta^t} \log[p(y|\theta)]$

$p(y|\theta)$ is a constant that is not a function of z, so

$E_{c|y,\theta^t}[\log p(y|\theta)] = \log p(y|\theta) E_{z|y,\theta^t}[1] = \log p(y|\theta)$

So LHS and RHS are equal when $p(c|y,\theta^t) = p(c|y,\theta)$

QED

Back

Kalman filter & the EM Algorithm Back

Re-cast state-space model with lag. We will need the joint distribution of $z_{t|T}$ and $z_{t-1|T}$ for the EM algorithm

$$\zeta_{i,t} = \begin{bmatrix} z_{i,t} \\ z_{i,t-1} \end{bmatrix},$$

State equation.

$$\zeta_{i,t} = \begin{bmatrix} z_{i,t} \\ z_{i,t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} B_t(l_{i,t}, \{y_{i,j}\}_{j=0}^{t-1}) \\ 0 \end{bmatrix}}_{\widehat{B}_t(l_{i,t}, \{y_{i,j}\}_{j=0}^{t-1})} + \zeta_{i,t-1} + \begin{bmatrix} \nu_{i,t} \\ 0 \end{bmatrix}. \tag{2}$$

Measurement equation.

$$y_{i,t} = H(l_{i,t})\zeta_{i,t} + l_{E,i,t}\, \omega_{i,t}, \tag{3}$$

Kalman filter & the EM Algorithm ‣ Back

Forward filter.

1. Estimate the Kalman Gain :

$$K_{i,t} = M_{i,t|t-1} H^{'}(l_{i,t}) \left[H(l_{i,t}) M_{i,t|t-1} H^{'}(l_{i,t}) + R_t(l_{i,t}, \{y_{i,j}\}_{j=0}^{t-1}) \right]^{-1}.$$

2. Update the state vector:

$$\widehat{\zeta}_{i,t|t} = \widehat{\zeta}_{i,t|t-1} + K_{i,t} \left(y_{i,t} - H(l_{i,t}) \widehat{\zeta}_{i,t|t-1} \right)$$
$$\widehat{\zeta}_{i,t+1|t} = \widehat{\zeta}_{i,t|t} + \widehat{B}_t(l_{i,t+1}, \{y_{i,j}\}_{j=0}^{t}).$$

3. Update the MSE matrix:

$$\begin{aligned} M_{i,t|t} &= M_{i,t|t-1} - K_{i,t} H(l_{i,t}) M_{i,t|t-1} \\ M_{i,t+1|t} &= M_{i,t|t} + Q_t(l_{i,t+1}, \{y_{i,j}\}_{j=0}^{t}) e_1^2 e_1^{2'}. \end{aligned}$$

Kalman filter & the EM Algorithm ▸ Back

Initial conditions. We treat the first observation of the dataset as y_0, and then we set $\widehat{\zeta}_{i,1|0} = \widehat{B}_1(l_{i,1}, y_0)$, and we initialize the variance of the state vector as,

$$M_{i,1|0} = \begin{bmatrix} Q_0(y_{i,0}) + Q(l_{i,1}, y_{i,0}) & Q_0(y_{i,0}) \\ Q_0(y_{i,0}) & Q_0(y_{i,0}) \end{bmatrix}.$$

RTS Smoother Under Augmented Sequential Exogeneity. In order to apply the EM algorithm below, we must apply the Kalman smoother after running the Kalman filter. We denote the resulting posterior estimates from running the Kalman smoother step as $\{\{\widehat{\zeta}_{i,t|T}\}_{t=0}^{T}\}_{i=1}^{N}$ and $\{\{M_{i,t|T}\}_{t=0}^{T}\}_{i=1}^{N}$.

- Must assume **Augmented Sequential Exogeneity (ASQ):** (l_t, x_t) satisfy

$$f\left(l_t, x_t | y^{t-1}, z^{t-1}, l^{t-1}, x^{t-1}; \{\theta, \lambda\}\right) = f\left(l_t, x_t | y^{t-1}, l^{t-1}, x^{t-1}; \lambda\right)$$

- Where θ is income process parameters, and λ is separate parameters governing missing data
- In words: *"given income history, no new information about income process is revealed by missing data"*

Kalman filter & the EM Algorithm ▸ Back

Restrictions one could place on the dynamics of (l_t, x_t) (Gourieroux and Monfort 1995)

1. **Strict exogeneity (S):** Given **LS**, (l_t, x_t) satisfy

$$f\left(l_t, x_t | y^{t-1}, z^{t-1}, l^{t-1}, x^{t-1}; \{\theta, \lambda\}\right) = f\left(l_t, x_t | l^{t-1}, x^{t-1}; \lambda\right)$$

   Two key components of the **S** assumption. First, the law of motion for (l_t, x_t) only depends on a disjoint parameter space λ, so conveys no information about θ. Second, the law of motion for (l_t, x_t) does not depend on $\{y_1, \ldots, y_T\}$ or $\{z_1, \ldots, z_T\}$ conditional on its own lags.

2. **Sequential Exogeneity (SQ):** Given **LS**, (l_t, x_t) satisfy

$$f\left(l_t, x_t | y^{t-1}, z^{t-1}, l^{t-1}, x^{t-1}; \{\theta, \lambda\}\right) = f\left(l_t, x_t | y^{t-1}, z^{t-1}, l^{t-1}, x^{t-1}; \lambda\right)$$

   The **SQ** assumption allows for dependence of l_t and x_t on the history of observable y^{t-1}, l^{t-1}, x^{t-1}, and unobservable z^{t-1}, while maintaining that (l_t, x_t) only depends on a disjoint parameter space λ.

3. **Augmented Sequential Exogeneity (ASQ) (see BHRS '24):** Given **LS**, (l_t, x_t) satisfy

$$f\left(l_t, x_t | y^{t-1}, z^{t-1}, l^{t-1}, x^{t-1}; \{\theta, \lambda\}\right) = f\left(l_t, x_t | y^{t-1}, l^{t-1}, x^{t-1}; \lambda\right)$$

   ASQ is weaker than **S** but stronger than **SQ**. It allows (l_t, x_t) to depend on observed earnings y^{t-1} and the history of l^{t-1} and x^{t-1}. However, its key restriction is that $(l_t, x_t) \perp z^{t-1} | (y^{t-1}, l^{t-1}, x^{t-1})$, so employment (and other conditioning variables x_t) don't depend on the latent states conditional on $(y^{t-1}, l^{t-1}, x^{t-1})$. Again, it also only depends on a disjoint parameter space λ.

Kalman filter & the EM Algorithm Back

To derive the EM updating formulas, we first write down the full-information log likelihood, assuming that state-variables are observed:

$$
\begin{aligned}
LL_i(\{y_{i,t}\}_{t=0}^T, \{z_{i,t}\}_{t=0}^T \mid \{I_{i,t}\}_{t=0}^T, \theta_0) = & -\frac{T+1}{2}\log(2\pi) \\
& -\frac{1}{2}\log(Q_0(y_{i,0})) - \frac{1}{2}\frac{(z_{i,0})^2}{Q_0(y_{i,0})} \\
& -\frac{1}{2}\sum_{t=1}^{T}\log(Q_t(I_{i,t}, \{y_{i,j}\}_{j=0}^{t-1})) - \frac{1}{2}\sum_{t=1}^{T}\frac{(z_{i,t} - z_{i,t-1} - B_t(I_{i,t}, \{y_{i,j}\}_{j=0}^{t-1}))^2}{Q_t(I_{i,t}, \{y_{i,j}\}_{j=0}^{t-1})} \\
& -\frac{1}{2}\sum_{t=1}^{T}\log(R_t(I_{i,t}, \{y_{i,j}\}_{j=0}^{t-1})) - \frac{1}{2}\sum_{t=1}^{T}\frac{(I_{E,i,t})(y_{i,t} - z_{i,t})^2}{R_t(I_{i,t}, \{y_{i,j}\}_{j=0}^{t-1})},
\end{aligned} \tag{4}
$$

where the parameter vector is denoted $\theta_0 = \{Q_t(\cdot), R_t(\cdot), B_t(\cdot)\}$.

E-Step. The second step is to take the expectation of the likelihood with respect to $\{z_{i,t}\}$, given $\{y_{i,t}\}_{t=0}^T$:

$$
\mathcal{L}(\theta_0) = E_{\{z_{i,t}\}_{t=0}^T}\left[LL_i(\{y_{i,t}\}_{t=0}^T, \{z_{i,t}\}_{t=0}^T \mid \{I_{i,t}\}_{t=0}^T, \theta_0) \middle| \{y_{i,t}\}_{t=0}^T\right]
$$

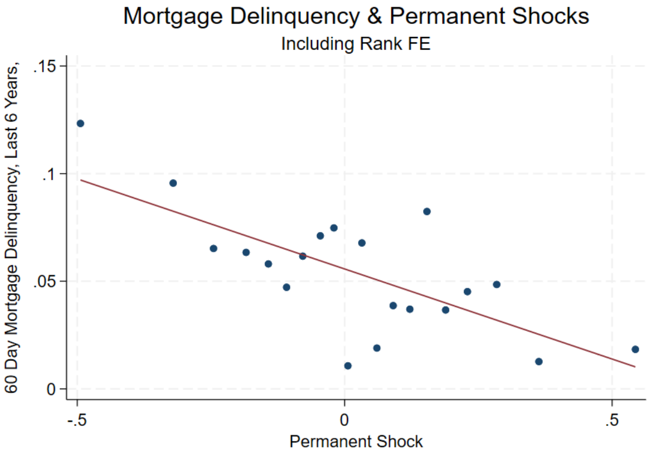
Kalman filter & the EM Alogrithm

M-step. The final step is to take first order conditions of $\mathcal{L}(\cdot)$ with respect to the income process parameter vector, θ_0. Consider the drift $B_t(\cdot)$. One may intuit from the quadratic functional form of the log likelihood that the resulting first order conditions will resemble ordinary least squares regressions formulas. The intuition is correct, except the updating formulas are generalized least squares regression formulas with weights that adjust for filtering uncertainty. The intuition comes from taking the expectation of the state equation with respect to $\{z_{i,t}\}$, given $\{y_{i,t}\}_{t=0}^{T}$, which naturally suggests that B can be updated via regression:

$$\widehat{z}_{i,t|T} = \widehat{z}_{i,t-1|T} + B_t(l_{i,t}, \{y_{i,j}\}_{j=0}^{t-1})$$

Is it risk?

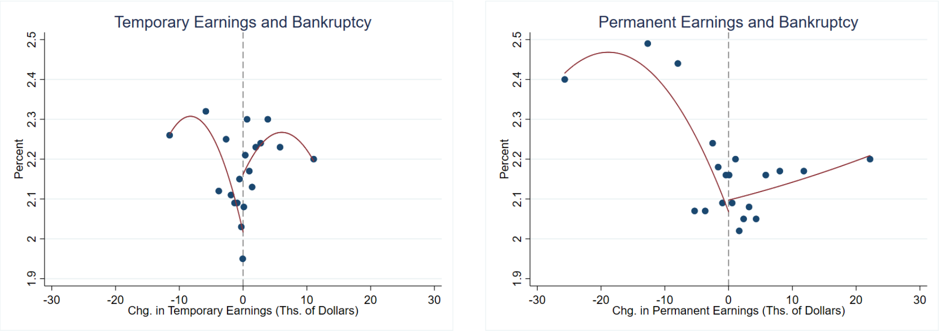
- Today: Show correlation of filtered shocks with negative events in PSID



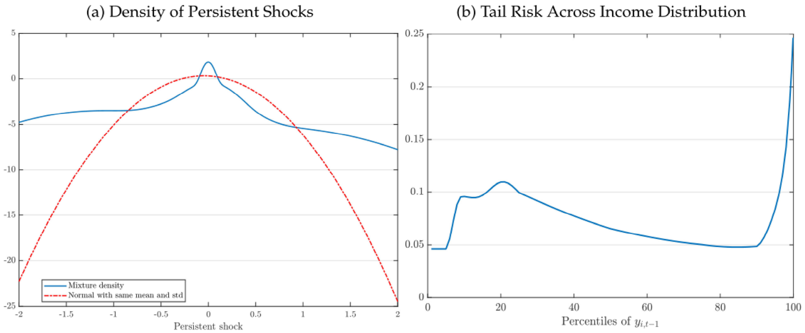
Is it risk?

- Today: Older filter with credit reports

Figure: Response of Bankruptcy to Changes in Temporary and Permanent Earnings

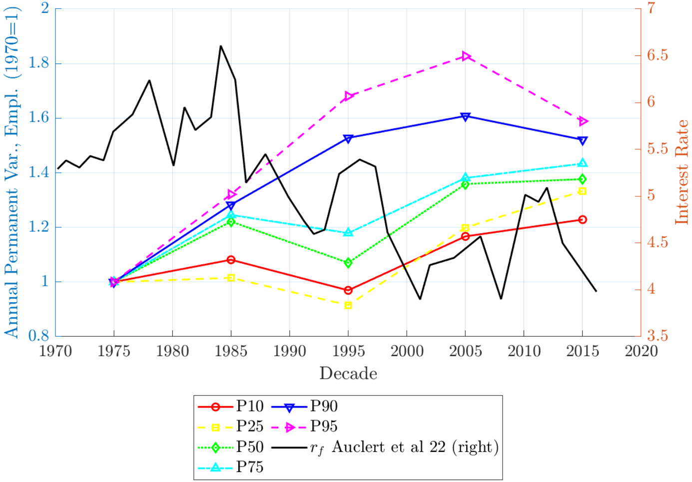


Tail risk Back

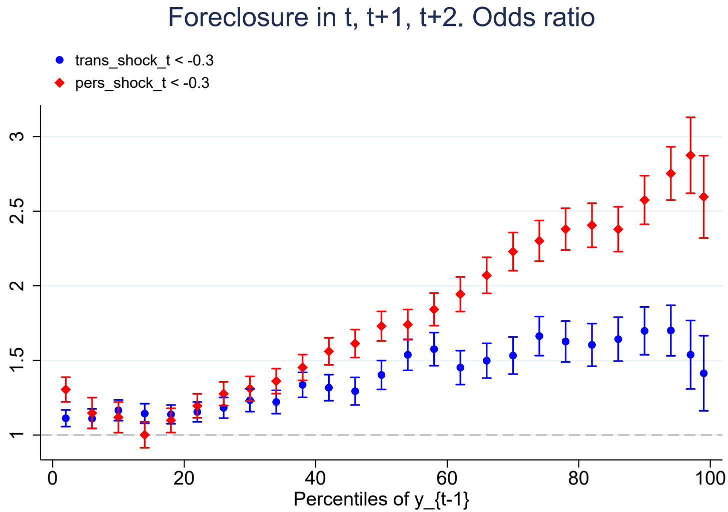


Note: Panel(a) shows the density of shocks to persistent earnings for our income process for an individual at the median of the lagged income distribution (blue, solid line) and for a normal distribution with the same mean and standard deviation (red, dashed line). The y-axis is in log-scale to highlight differences in the tails. Panel (b) shows the probability of having a 30 percent or larger decline in persistent earnings as a function of prior earnings from our individual level estimations of persistent shocks.

Permanent risk rose most for high income Back

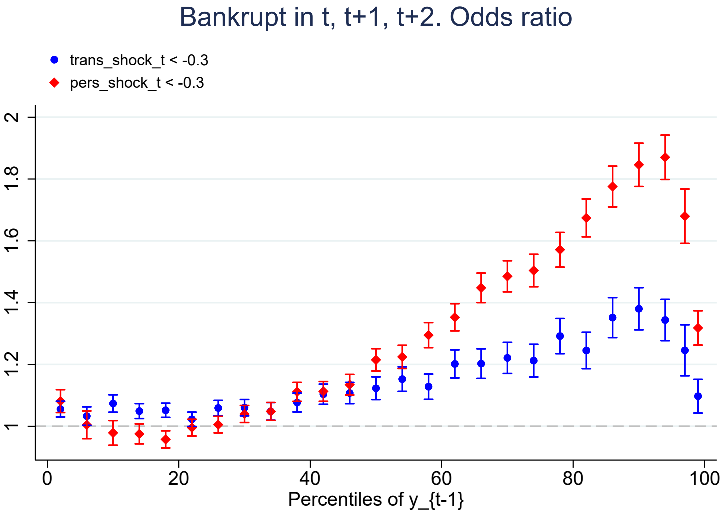


Odds ratio after permanent (red) & transitory shocks (blue)



Back

Odds ratio after permanent (red) & transitory shocks (blue)



Back