What Do Data on Millions of U.S. Workers Say About Life Cycle Income Risk?[1]

Fatih Guvenen
Minnesota and NBER

Fatih Karahan
New York Fed

Serdar Ozkan
Federal Reserve Board

Jae Song
SSA

Workshop on the Estimation of Economic Models of Earnings Dynamics
June 13, 2014

[1] The findings and conclusions expressed are solely those of the authors and do not represent the views of Federal Reserve Board, Federal Reserve Bank of New York or SSA.

Earnings Dynamics: Open Questions

1. How big are earnings shocks?
2. How persistent are they?
   - Do positive and negative shocks have similar persistence?
3. How do the properties of shocks vary over the life cycle?
   - e.g., standard deviation, skewness, kurtosis, etc.
4. Are shocks log normally distributed? How about higher order moments?

Earnings Dynamics: Open Questions

1. How big are earnings shocks?
2. How persistent are they?
   - Do positive and negative shocks have similar persistence?
3. How do the properties of shocks vary over the life cycle?
   - e.g., standard deviation, skewness, kurtosis, etc.
4. Are shocks log normally distributed? How about higher order moments?

Earnings Dynamics: Open Questions

1. How big are earnings shocks?
2. How persistent are they?
    - Do positive and negative shocks have similar persistence?
3. How do the properties of shocks vary over the life cycle?
    - e.g., standard deviation, skewness, kurtosis, etc.
4. Are shocks log normally distributed? How about higher order moments?

Earnings Dynamics: Open Questions

1. How big are earnings shocks?
2. How persistent are they?
   - Do positive and negative shocks have similar persistence?
3. How do the properties of shocks vary over the life cycle?
   - e.g., standard deviation, skewness, kurtosis, etc.
4. Are shocks log normally distributed? How about higher order moments?

This Paper

Existing work:

1. Small survey-based data sets, e.g., the PSID
    - between 500 to 2000 individuals per year
2. Employ covariance matrix estimation (CME), developed for a data-constrained environment

This paper:

1. Large and clean administrative data set
    - as many as 5,000,000 individuals per year.
2. Move beyond CME and target economically significant moments.

This Paper

Existing work:

1. Small survey-based data sets, e.g., the PSID
   - between 500 to 2000 individuals per year
2. Employ covariance matrix estimation (CME), developed for a data-constrained environment

This paper:

1. Large and clean administrative data set
   - as many as 5,000,000 individuals per year.
2. Move beyond CME and target economically significant moments.

This Paper

Existing work:

1. Small survey-based data sets, e.g., the PSID
   - between 500 to 2000 individuals per year
2. Employ covariance matrix estimation (CME), developed for a data-constrained environment

This paper:

1. Large and clean administrative data set
   - as many as 5,000,000 individuals per year.
2. Move beyond CME and target economically significant moments.

This Paper

Existing work:

1. Small survey-based data sets, e.g., the PSID
   - between 500 to 2000 individuals per year
2. Employ covariance matrix estimation (CME), developed for a data-constrained environment

This paper:

1. Large and clean administrative data set
   - as many as 5,000,000 individuals per year.
2. Move beyond CME and target economically significant moments.

Data: SSA Master Earnings File

- Representative sample of US males covering 34 years: 1978 to 2011
- Salary and wage workers (from W2 forms)
- Individuals aged 25–60
- Key Advantages:
  - Very large sample size (200+ million observations)
  - No survey response error
  - No sample attrition
  - No top-coding

Data: SSA Master Earnings File

- Representative sample of US males covering 34 years: 1978 to 2011
- Salary and wage workers (from W2 forms)
- Individuals aged 25–60
- Key Advantages:
  - Very large sample size (200+ million observations)
  - No survey response error
  - No sample attrition
  - No top-coding

Data: SSA Master Earnings File

- Representative sample of US males covering 34 years: 1978 to 2011
- Salary and wage workers (from W2 forms)
- Individuals aged 25–60
- Key Advantages:
  - Very large sample size (200+ million observations)
  - No survey response error
  - No sample attrition
  - No top-coding

Data: SSA Master Earnings File

- Representative sample of US males covering 34 years: 1978 to 2011
- Salary and wage workers (from W2 forms)
- Individuals aged 25–60
- Key Advantages:
  - Very large sample size (200+ million observations)
  - No survey response error
  - No sample attrition
  - No top-coding

Moving Beyond the Covariance Matrix

- CME method was developed for a severely data-constrained environment.
  - Not a constraint here.
- One problem with CME is that selecting among rejected models is very hard:
  - moments that are missed do not have clear economic interpretations.
- Also ignores higher order moments, which we find to be very important.

Moving Beyond the Covariance Matrix

- CME method was developed for a severely data-constrained environment.
  - Not a constraint here.
- One problem with CME is that selecting among rejected models is very hard:
  - moments that are missed do not have clear economic interpretations.
- Also ignores higher order moments, which we find to be very important.

Moving Beyond the Covariance Matrix

- CME method was developed for a severely data-constrained environment.
  - Not a constraint here.
- One problem with CME is that selecting among rejected models is very hard:
  - moments that are missed do not have clear economic interpretations.
- Also ignores higher order moments, which we find to be very important.

This Paper

Uses a unique, confidential, and very large administrative dataset to:

1. Document new empirical facts on life cycle earnings dynamics
2. Estimate lifecycle labor income risk
   - by matching economically important moments (as opposed to the "covariance matrix of income residuals")
3. Provide a reliable "user's guide" for earnings process specifications.

This Paper

Uses a unique, confidential, and very large administrative dataset to:

1. Document new empirical facts on life cycle earnings dynamics
2. Estimate lifecycle labor income risk
   - by matching economically important moments (as opposed to the "covariance matrix of income residuals")
3. Provide a reliable "user's guide" for earnings process specifications.

This Paper

Uses a unique, confidential, and very large administrative dataset to:

1. Document new empirical facts on life cycle earnings dynamics
2. Estimate lifecycle labor income risk
   - by matching economically important moments (as opposed to the “covariance matrix of income residuals”)
3. Provide a reliable “user’s guide” for earnings process specifications.

This Paper

Uses a unique, confidential, and very large administrative dataset to:

1. Document new empirical facts on life cycle earnings dynamics
2. Estimate lifecycle labor income risk
    - by matching economically important moments (as opposed to the "covariance matrix of income residuals")
3. Provide a reliable "user's guide" for earnings process specifications.

NEW EMPIRICAL FACTS

Four Sets of Empirical Facts

1. Average income growth over the life cycle
2. Cross-sectional moments of earnings growth
3. Short- and long-run dynamics of income growth
4. Scarring Effects of Long-Term Unemployment
5. Distribution of Lifetime Income (skip today)

Four Sets of Empirical Facts

1. Average income growth over the life cycle
2. Cross-sectional moments of earnings growth
3. Short- and long-run dynamics of income growth
4. Scarring Effects of Long-Term Unemployment
5. Distribution of Lifetime Income (skip today)

Four Sets of Empirical Facts

1. Average income growth over the life cycle
2. Cross-sectional moments of earnings growth
3. Short- and long-run dynamics of income growth
4. Scarring Effects of Long-Term Unemployment
5. Distribution of Lifetime Income (skip today)

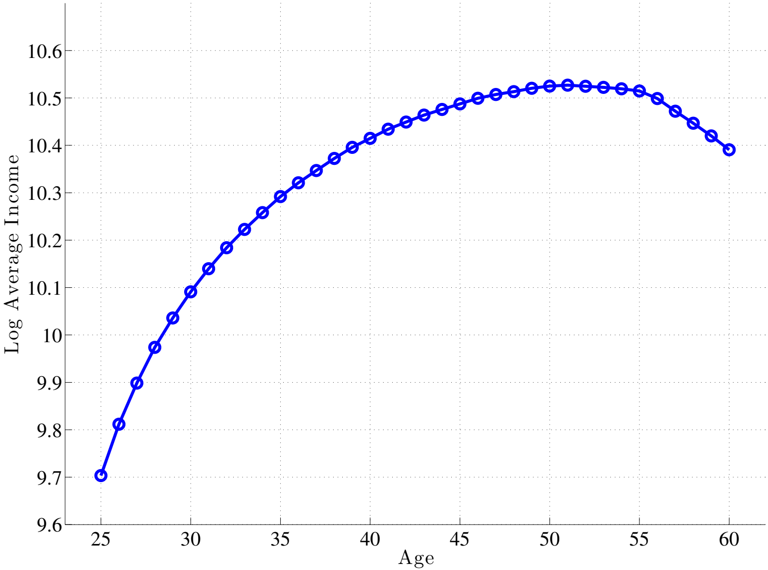
Four Sets of Empirical Facts

1. Average income growth over the life cycle
2. Cross-sectional moments of earnings growth
3. Short- and long-run dynamics of income growth
4. Scarring Effects of Long-Term Unemployment
5. Distribution of Lifetime Income (skip today)

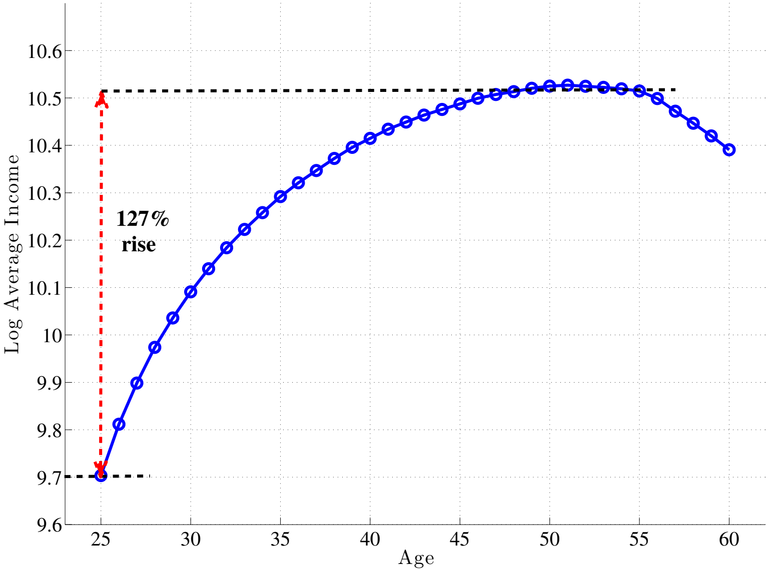
Four Sets of Empirical Facts

1. Average Income growth over the life cycle
2.
3.
4.

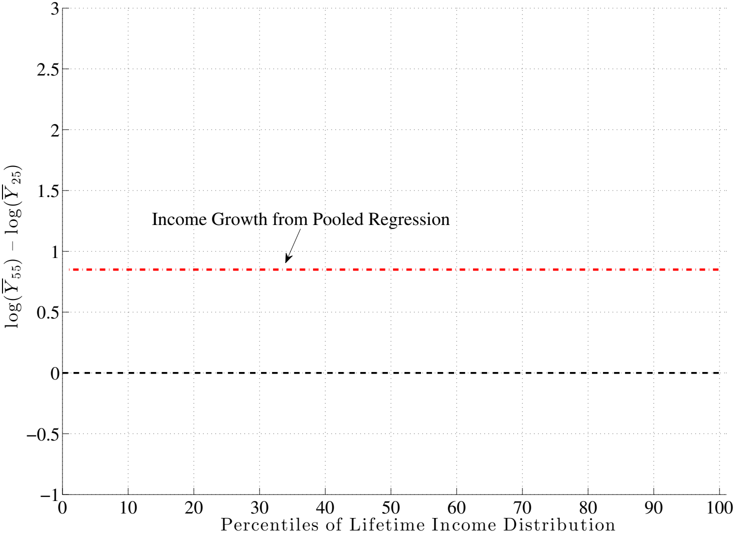
I. Age Profile of Labor Income



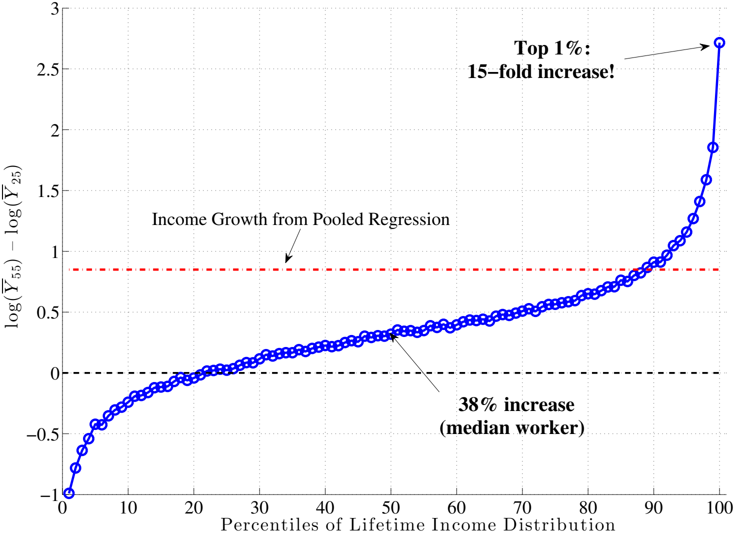
I. Age Profile of Labor Income



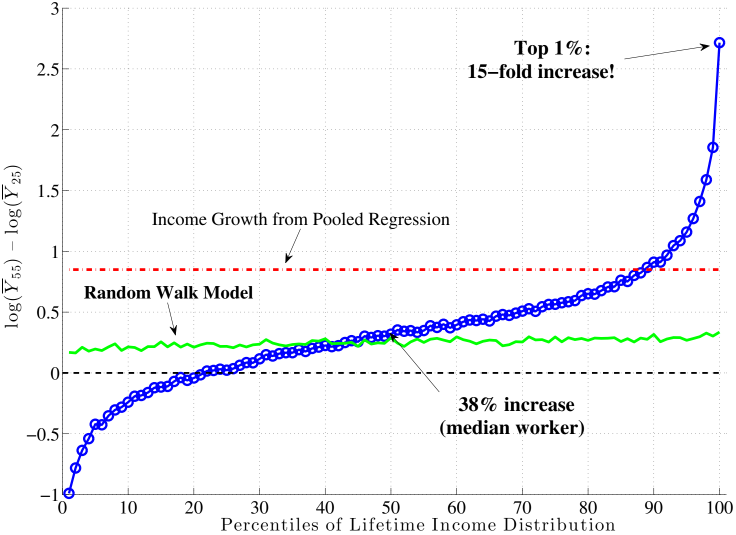
I. Income Growth Over Life Cycle



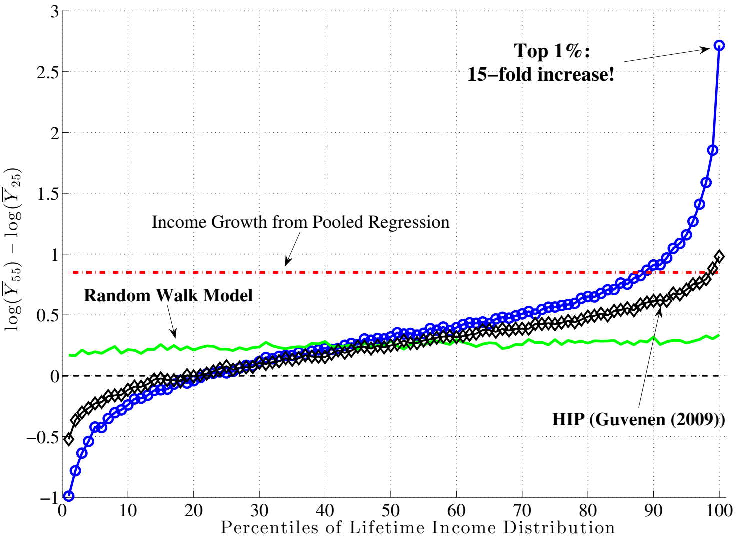
I. Income Growth Over Life Cycle



I. Income Growth Over Life Cycle



I. Income Growth Over Life Cycle



Four Sets of Empirical Facts

1. Income growth over the life cycle

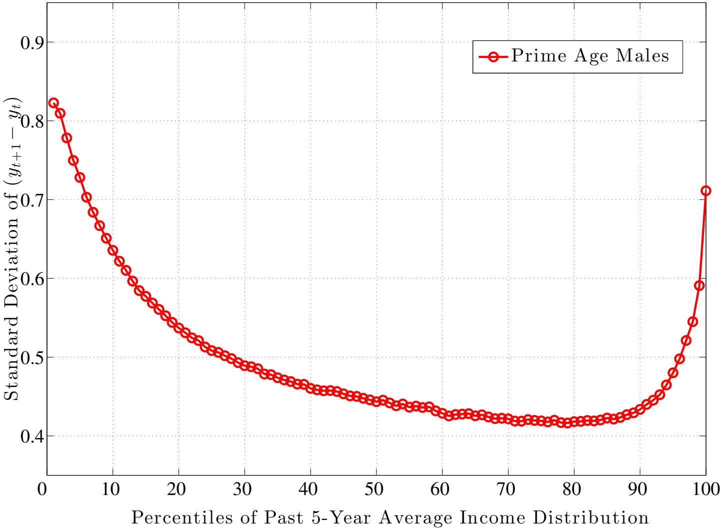
2. Cross-sectional moments of earnings growth: $y_{t+k} - y_t$

3.

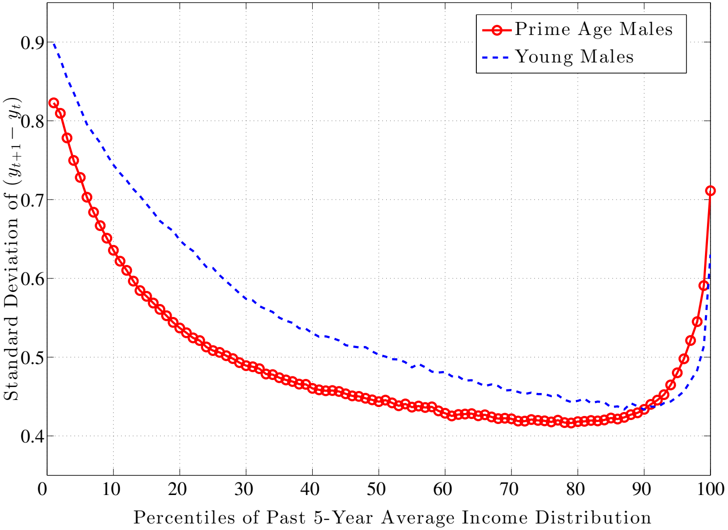
4.

Standard Deviation and Skewness

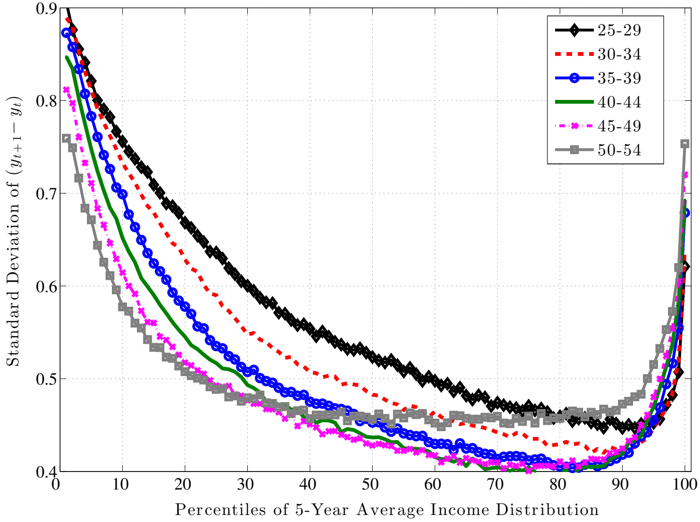
II.a Standard Deviation of $y_{t+1} - y_t$



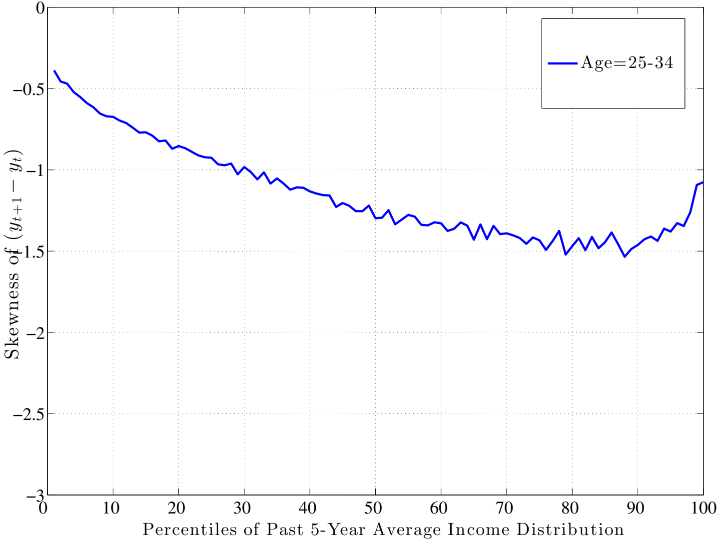
II.A Standard Deviation of $y_{t+1} - y_t$



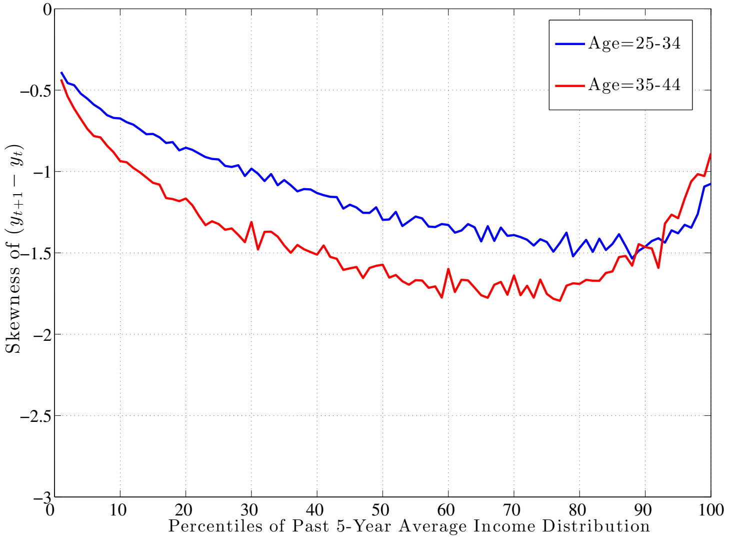
II.a Standard Deviation of $y_{t+1} - y_t$



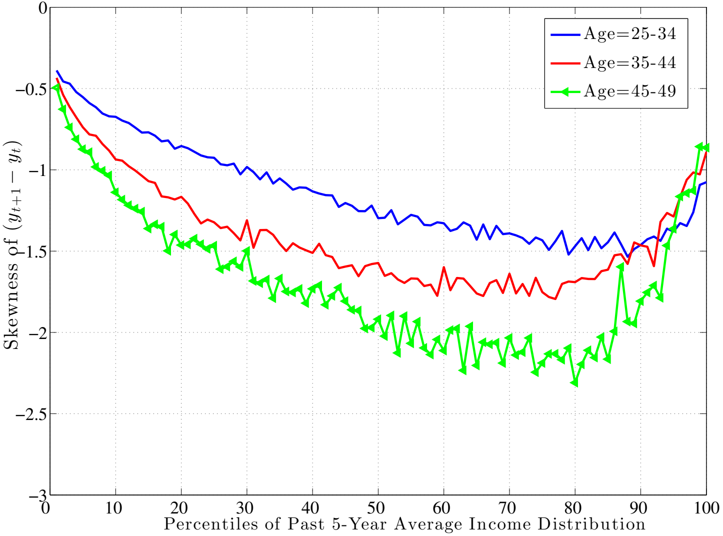
II.B Skewness of $y_{t+1} - y_t$



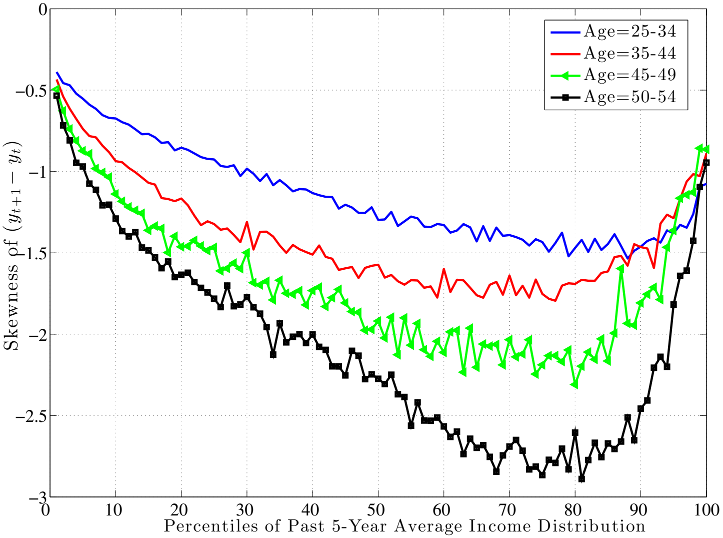
II.B Skewness of $y_{t+1} - y_t$



II.B Skewness of $y_{t+1} - y_t$

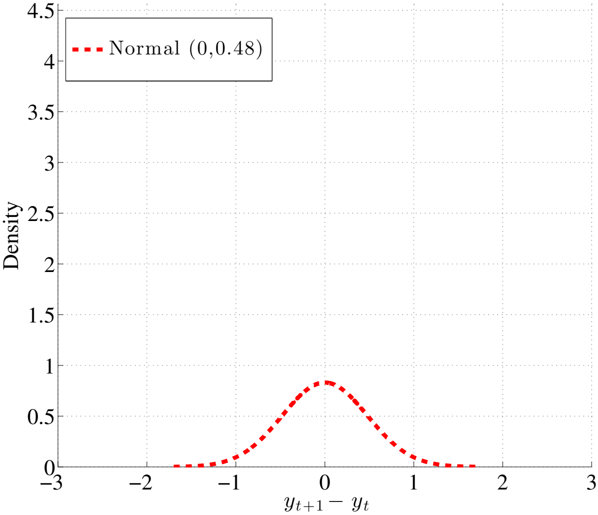


II.B Skewness of $y_{t+1} - y_t$

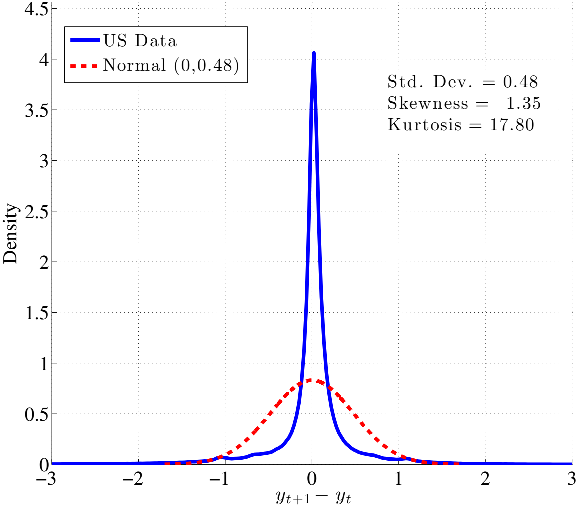


Kurtosis

II.C Histogram of $y_{t+1} - y_t$



II.C Histogram of $y_{t+1} - y_t$



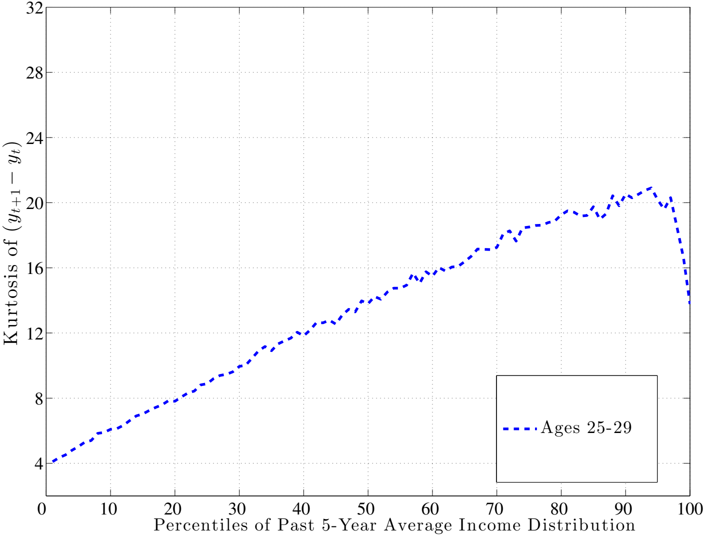
II.c Distribution of Income Changes

	Prob($\lvert y_{t+1} - y_t \rvert < x$)	
$x \downarrow$	Data	$N(0, 0.43^2)$
0.05	0.42	0.10
0.10	0.63	0.20
0.20	0.79	0.39
0.50	0.90	0.80
1.00	0.96	0.99

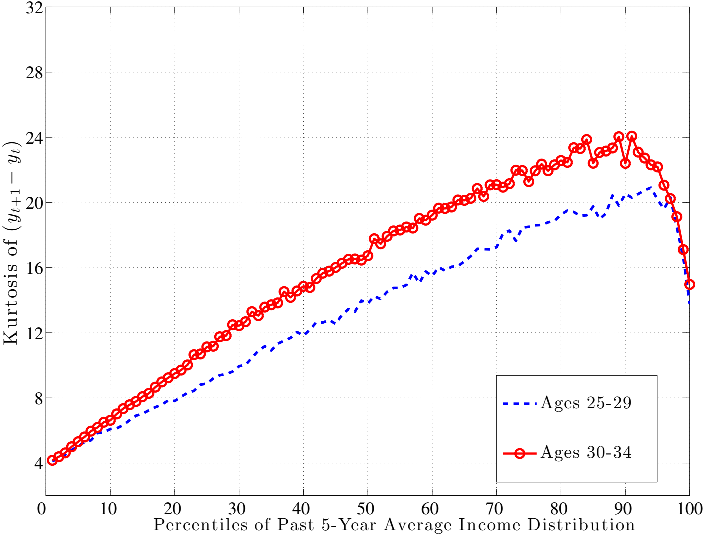
II.c Distribution of Income Changes

	$\text{Prob}(\lvert y_{t+1} - y_t \rvert < x)$	
$x \downarrow$	Data	$N(0, 0.43^2)$
0.05	0.42	0.10
0.10	0.63	0.20
0.20	0.79	0.39
0.50	0.90	0.80
1.00	0.96	0.99

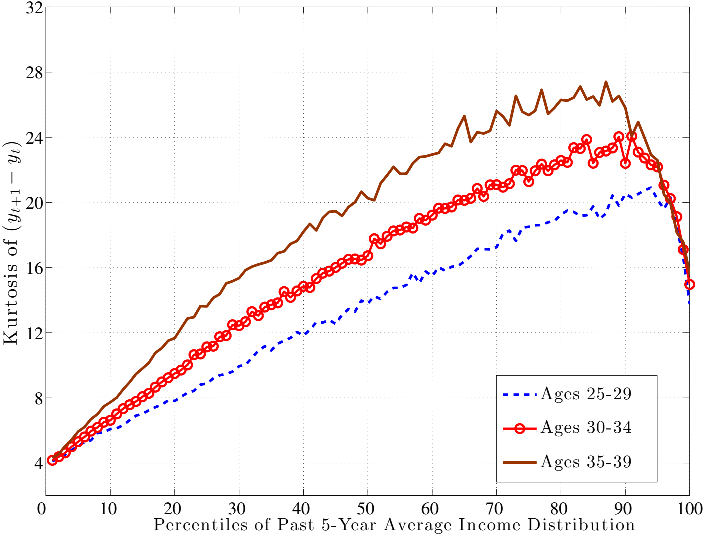
II.C Kurtosis of $y_{t+1} - y_t$



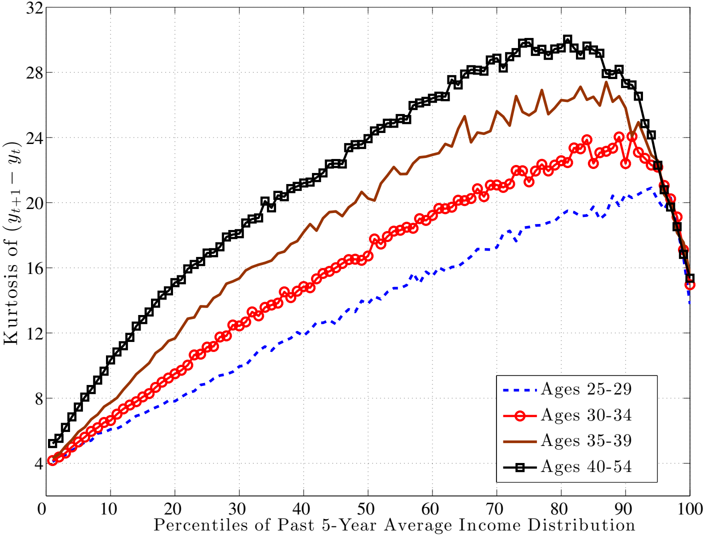
II.c Kurtosis of $y_{t+1} - y_t$



II.c Kurtosis of $y_{t+1} - y_t$



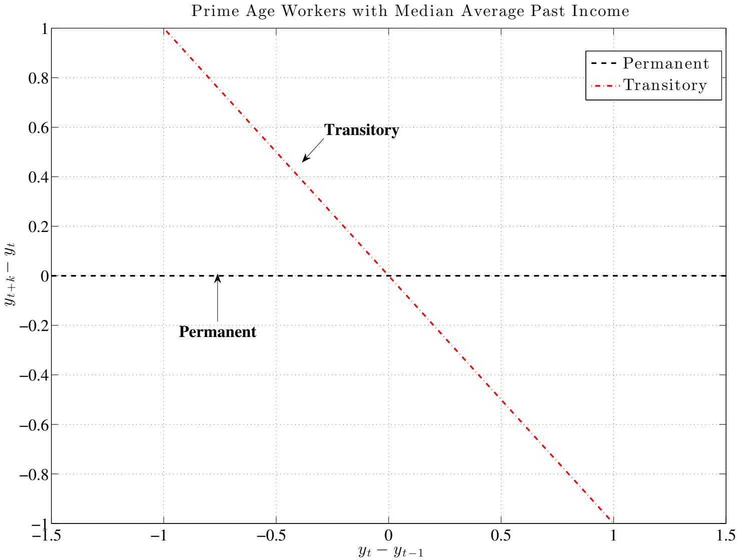
II.c Kurtosis of $y_{t+1} - y_t$



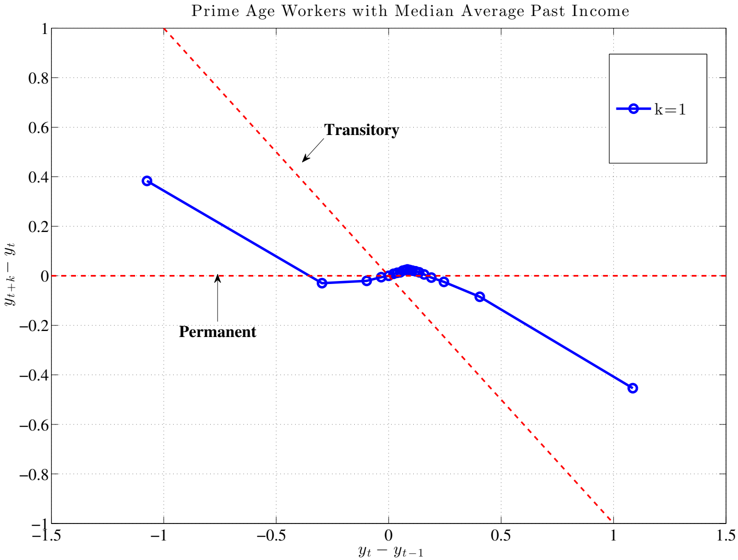
Four Sets of Empirical Facts

1. Average income growth over the life cycle
2. Cross-sectional moments of earnings growth: $y_{t+k} - y_t$
3. Short- and long-run dynamics of income growth
4.

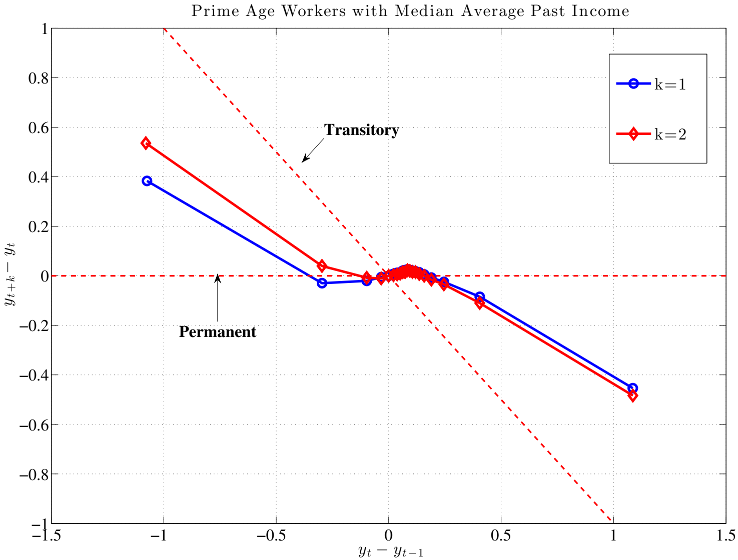
Impulse Response Functions



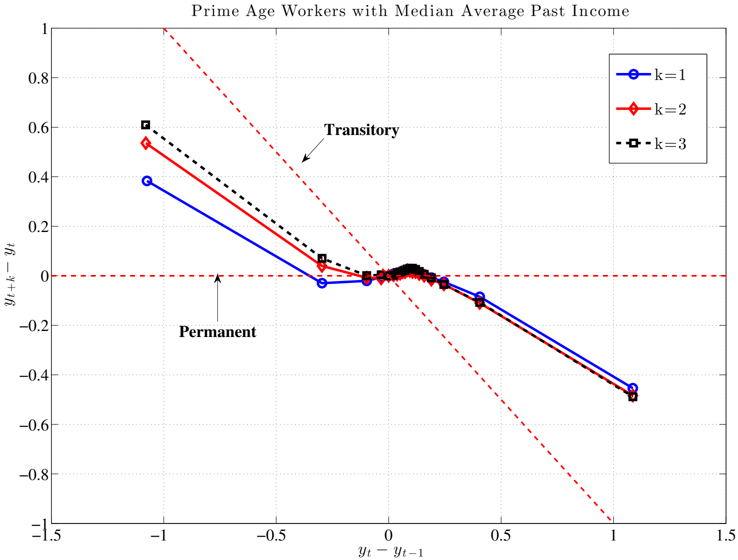
Impulse Response Functions



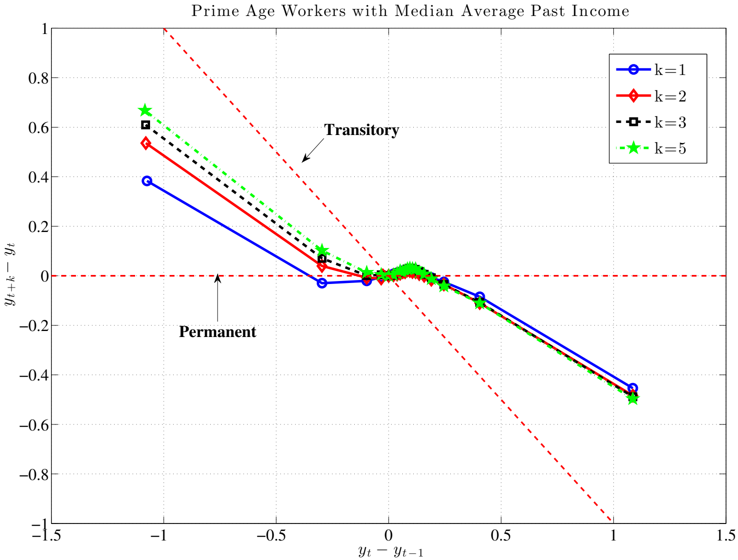
Impulse Response Functions



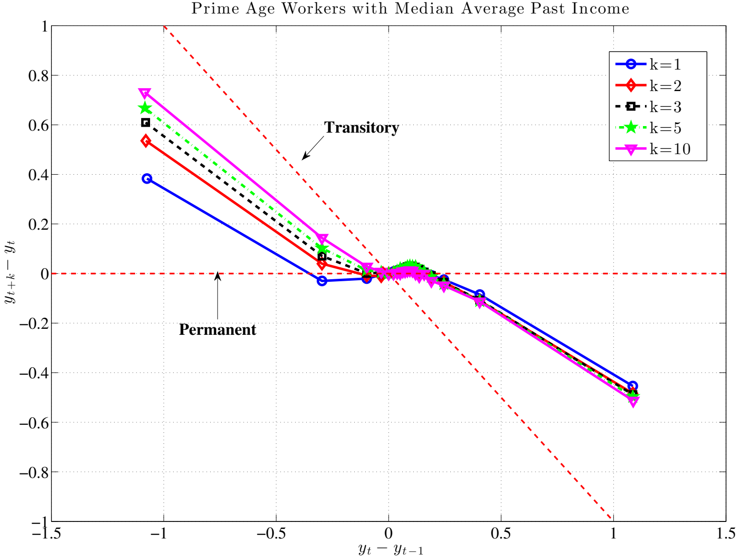
Impulse Response Functions



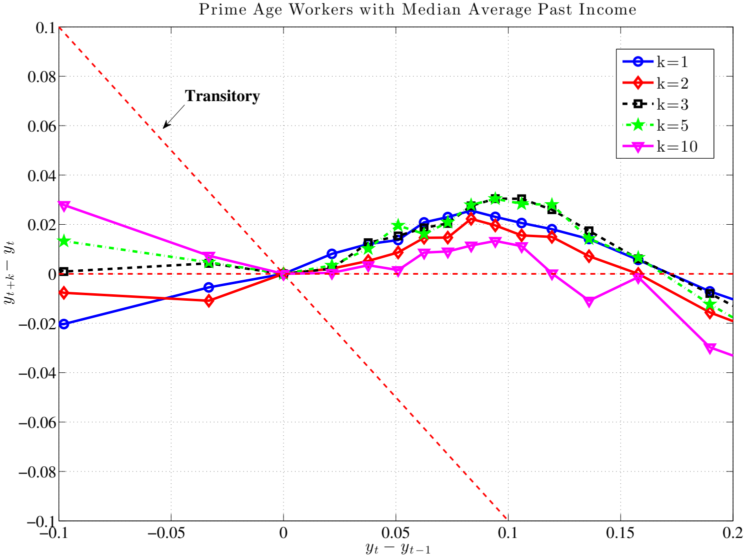
Impulse Response Functions



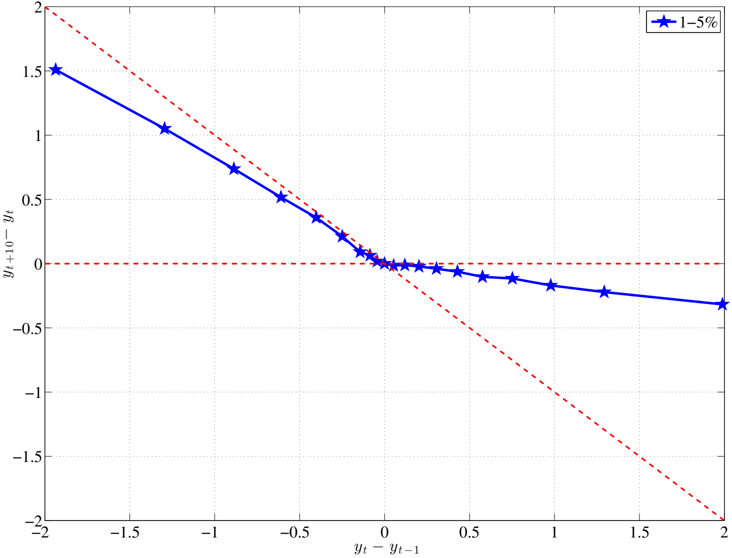
Impulse Response Functions



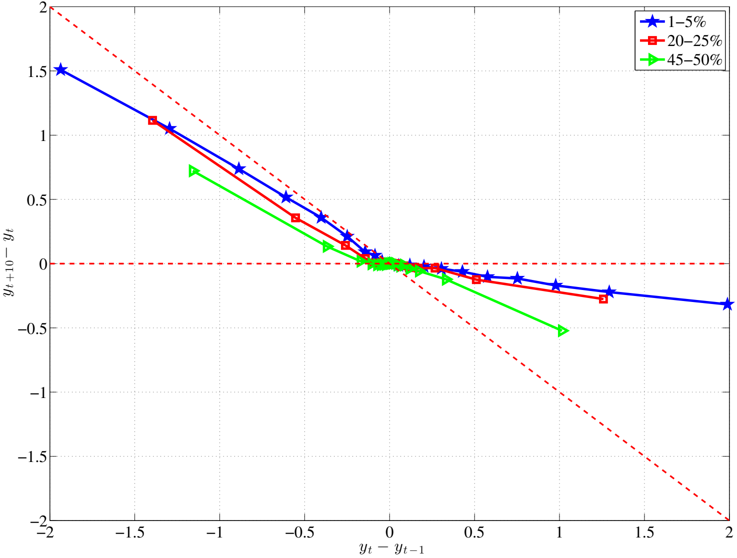
Impulse Response Functions



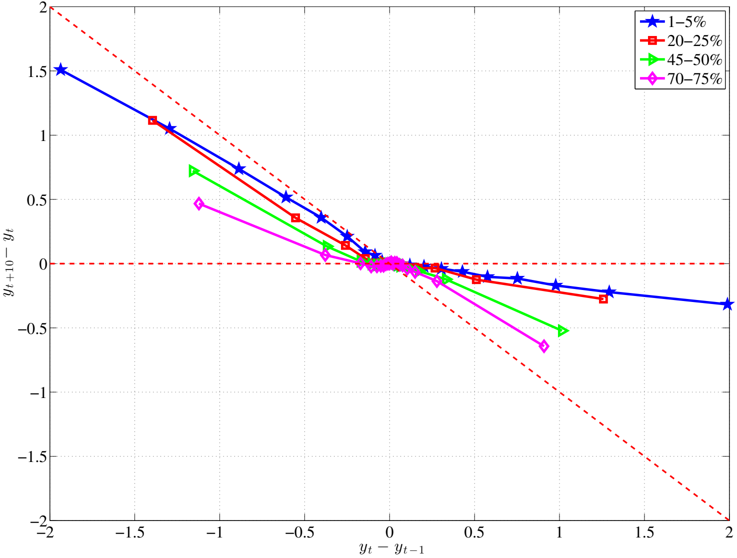
Asymmetric Mean Reversion



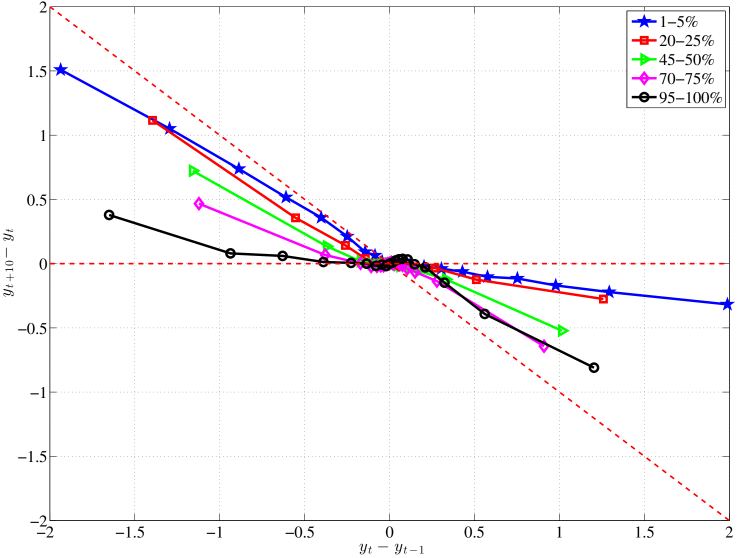
Asymmetric Mean Reversion



Asymmetric Mean Reversion



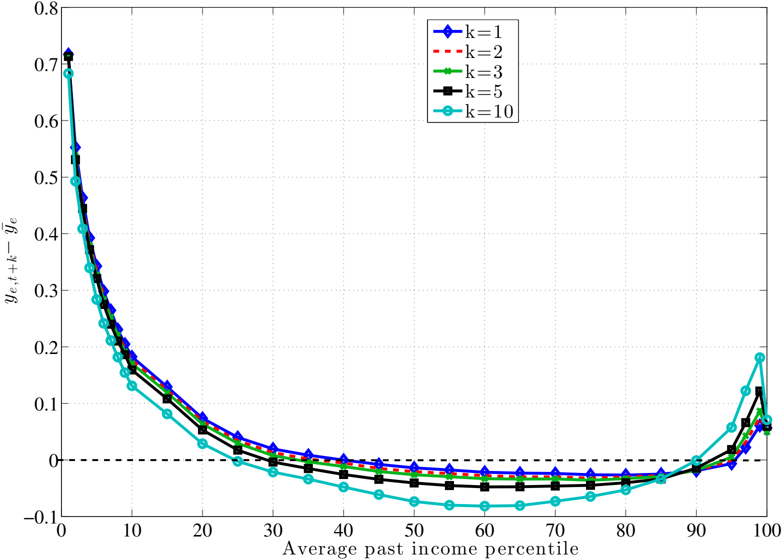
Asymmetric Mean Reversion



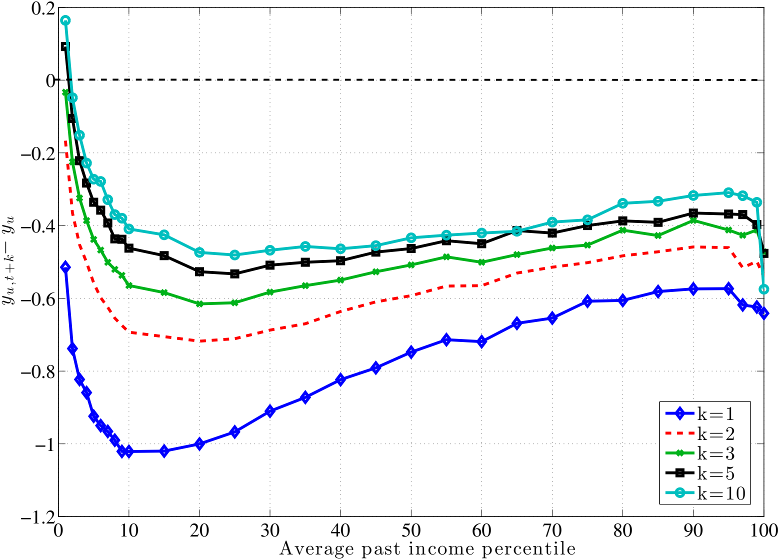
Four Sets of Empirical Facts

1. Average income growth over the life cycle
2. Cross-sectional moments of earnings growth: $y_{t+k} - y_t$
3. Short- and long-run dynamics of income growth
4. "Scarring" Effects of Long-Term Unemployment

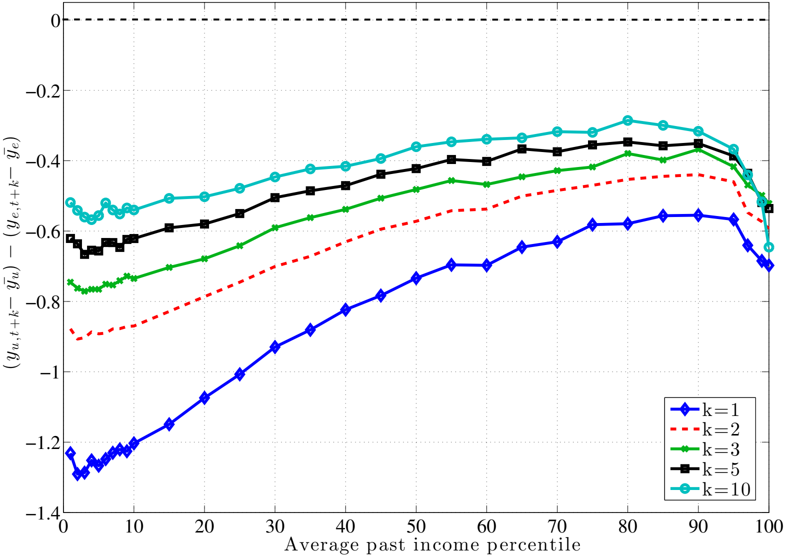
Prime-Age Workers: Employed



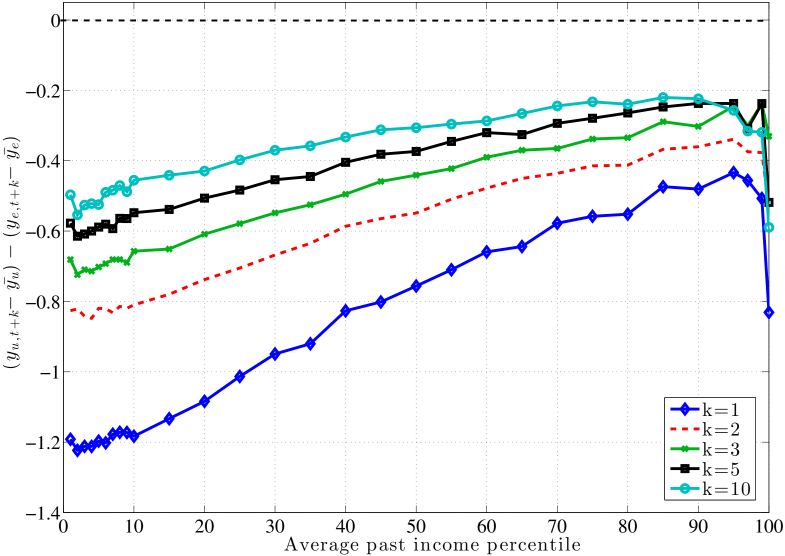
Prime-Age Workers: Unemployed



Prime-Age Workers: Diff. in Diff.



Young Workers: Diff. in Diff.



ESTIMATION

Econometric Specification

$$y_t^i = \underbrace{\left[\alpha^i + \beta^i t + \gamma^i t^2\right]}_{\text{HIP}} + \underbrace{z_{1,t}^i + z_{2,t}^i}_{\text{mixture of AR(1)s}} + \underbrace{\varepsilon_t^i}_{\text{i.i.d.}}$$

$$z_{1,t}^i = \rho_1 z_{1,t-1}^i + \eta_{1,t}^i$$
$$z_{2,t}^i = \rho_2 x_{2,t-1}^i + \eta_{2,t}^i$$

where for $j = 1, 2$:

$$\eta_{jt}^i = \begin{cases} 0 & \text{w.p.} \quad 1 - p_j \\ \sim \mathcal{N}(0, \sigma_j) & \text{w.p.} \quad p_j \end{cases}$$

and

$$\sigma_j(t, z_{t-1}) = \max\left(0, \sigma_{j,0} + a_j \times z_{t-1} + b_j \times t + c_j \times z_{t-1} \times t\right)$$

Estimation Results

Parameters	Group 1	Group 2	Group 2
Fractions	0.10	0.80	0.10
mean(α)	2.21	2.95	3.57
mean(β)$\times$100	4.31	9.44	12.27
quadratic	–0.25	–0.25	–0.25
σ_α	-0.74	0.00	0.63
$\sigma_\beta \times 100$	1.02	1.35	0.68
$\sigma_{\alpha\beta}$	–0.02	–0.41	0.21
p_1		0.11	
p_2		0.77	
ρ_1		0.25	
ρ_2		0.54	
σ_1		$1.07 + 0.65z_{t-1} + 0.32t + 0.148tz_{t-1}$	
σ_2		$0.07 - 0.15z_{t-1} - 0.15t - 0.21tz_{t-1}$	
σ_ϵ		0.03	

What to Use in Calibration?

- These estimated processes are complex and richly parameterized.
  - How to use them for calibration?
- We intend to construct Markov transition matrices that summarize these processes.
- Civale-Guvenen-Stefanides (2013) explore how to do this for processes with excess kurtosis and large skewness.
  - Results to so far quite encouraging.

What to Use in Calibration?

- These estimated processes are complex and richly parameterized.
  - How to use them for calibration?
- We intend to construct Markov transition matrices that summarize these processes.
- Civale-Guvenen-Stefanides (2013) explore how to do this for processes with excess kurtosis and large skewness.
  - Results to so far quite encouraging.

What to Use in Calibration?

- These estimated processes are complex and richly parameterized.
  - How to use them for calibration?
- We intend to construct Markov transition matrices that summarize these processes.
- Civale-Guvenen-Stefanides (2013) explore how to do this for processes with excess kurtosis and large skewness.
  - Results to so far quite encouraging.

A Structural Interpretation

- Within-job earnings changes are small.
  - Every once in a while: find a better job or lose the job.
- Job mobility declines with age and wage.
  - Kurtosis goes up with age and wage
    - Variance of income changes decline with age and wage
- Skewness: Job losses contribute to the left tail.
  - Larger left tail for older people and for high wage people.
- These insights are mostly missed in income dynamics literature.

A Structural Interpretation

- Within-job earnings changes are small.
  - Every once in a while: find a better job or lose the job.
- Job mobility declines with age and wage.
  - Kurtosis goes up with age and wage
    - Variance of income changes decline with age and wage
- Skewness: Job losses contribute to the left tail.
  - Larger left tail for older people and for high wage people.
- These insights are mostly missed in income dynamics literature.

A Structural Interpretation

- Within-job earnings changes are small.
  - Every once in a while: find a better job or lose the job.
- Job mobility declines with age and wage.
  - Kurtosis goes up with age and wage
    - Variance of income changes decline with age and wage
- Skewness: Job losses contribute to the left tail.
  - Larger left tail for older people and for high wage people.
- These insights are mostly missed in income dynamics literature.

A Structural Interpretation

- Within-job earnings changes are small.
  - Every once in a while: find a better job or lose the job.
- Job mobility declines with age and wage.
  - Kurtosis goes up with age and wage
    - Variance of income changes decline with age and wage
- Skewness: Job losses contribute to the left tail.
  - Larger left tail for older people and for high wage people.
- These insights are mostly missed in income dynamics literature.

Conclusions

- Striking new regularities and patterns in individual earnings.
- Existing specifications do not capture these salient features of the data.
- We propose a richer specification that captures many of these patterns.
- Excess kurtosis and mixture of AR(1)s can explain an important puzzle in the CME literature.
  - Ongoing research (Guvenen-Tonetti aims to show this more rigorously.)
- New benchmarks and targets for calibration.

Conclusions

- Striking new regularities and patterns in individual earnings.
- Existing specifications do not capture these salient features of the data.
- We propose a richer specification that captures many of these patterns.
- Excess kurtosis and mixture of AR(1)s can explain an important puzzle in the CME literature.
  - Ongoing research (Guvenen-Tonetti aims to show this more rigorously.)
- New benchmarks and targets for calibration.

Conclusions

- Striking new regularities and patterns in individual earnings.
- Existing specifications do not capture these salient features of the data.
- We propose a richer specification that captures many of these patterns.
- Excess kurtosis and mixture of AR(1)s can explain an important puzzle in the CME literature.
    - Ongoing research (Guvenen-Tonetti aims to show this more rigorously.)
- New benchmarks and targets for calibration.

Conclusions

- Striking new regularities and patterns in individual earnings.
- Existing specifications do not capture these salient features of the data.
- We propose a richer specification that captures many of these patterns.
- Excess kurtosis and mixture of AR(1)s can explain an important puzzle in the CME literature.
    - Ongoing research (Guvenen-Tonetti aims to show this more rigorously.)
- New benchmarks and targets for calibration.

Conclusions

- Striking new regularities and patterns in individual earnings.
- Existing specifications do not capture these salient features of the data.
- We propose a richer specification that captures many of these patterns.
- Excess kurtosis and mixture of AR(1)s can explain an important puzzle in the CME literature.
    - Ongoing research (Guvenen-Tonetti aims to show this more rigorously.)
- New benchmarks and targets for calibration.