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

Fatih Guvenen
Minnesota and NBER

Fatih Karahan
New York Fed

Serdar Ozkan
Toronto

Jae Song
SSA

March 26, 2015

[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.

Three Questions

1. What does the distribution of earnings shocks look like?

Three Questions

1. What does the distribution of earnings shocks look like?
   - Is it approximately lognormal?
   - Symmetric or skewed? any excess kurtosis?

Three Questions

1. What does the distribution of earnings shocks look like?
   - Is it approximately lognormal?
   - Symmetric or skewed? any excess kurtosis?
2. How do these properties vary:
   A. over the life cycle?
   B. across income groups?

Three Questions

1. What does the distribution of earnings shocks look like?
   - Is it approximately lognormal?
   - Symmetric or skewed? any excess kurtosis?
2. How do these properties vary:
   A. over the life cycle?
   B. across income groups?
3. Dynamics of earnings:
   A. Do positive and negative shocks have similar persistence?
   B. Do large and small shocks have similar persistence?

This Paper

Existing work:

1. Small survey-based data sets, e.g., the PSID
    - between 500 to 2000 individuals per year

This paper:

This Paper

Existing work:

1. Small survey-based data sets, e.g., the PSID
    - between 500 to 2000 individuals per year

This paper:

1. Large and clean administrative data set
    - as many as 5,000,000 individuals per year.

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
   - *Notable exceptions:* Meghir and Pistaferri (2004), Browning et al (2010), and Altonji et al (2013)

This paper:

1. Large and clean administrative data set
   - as many as 5,000,000 individuals per year.

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
   - *Notable exceptions:* Meghir and Pistaferri (2004), Browning et al (2010), and Altonji et al (2013)

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.

Covariance Matrix Estimation: A Recap

1. Specify a parametric income process, e.g.:

$$y_t^i = \alpha^i + z_t^i + \varepsilon_t^i$$
$$z_t^i = \rho z_{t-1}^i + \eta_t^i$$

Covariance Matrix Estimation: A Recap

1. Specify a parametric income process, e.g.:

$$y_t^i = \alpha^i + z_t^i + \varepsilon_t^i$$
$$z_t^i = \rho z_{t-1}^i + \eta_t^i$$

2. Derive the theoretical autocovariances of income implied by this specification:

$$\text{var}_i(y_t^i) = \sigma_\alpha^2 + \text{var}_i(z_t^i) + \sigma_\varepsilon^2,$$
$$\text{var}_i(z_t^i) = \sum_{s=1}^{t} \rho^2 \sigma_\eta^2,$$
$$\text{cov}(y_t^i, y_{t+n}^i) = \sigma_\alpha^2 + \rho^n \text{var}_i(z_t^i).$$

Covariance Matrix Estimation

3. Construct the empirical counterpart of the covariance matrix:

$$\begin{bmatrix} \mathrm{var}_i(y_1^i) & & & & \\ \vdots & \mathrm{var}_i(y_2^i) & & & \\ \vdots & \cdots & \ddots & & \\ \mathrm{cov}(y_1^i, y_t^i) & \cdots & \mathrm{var}_i(y_t^i) & & \\ \vdots & & & \ddots & \\ \mathrm{cov}(y_1^i, y_T^i) & & & & \mathrm{var}_i(y_T^i) \end{bmatrix}$$

Covariance Matrix Estimation

3. Construct the empirical counterpart of the covariance matrix:

$$\begin{bmatrix} \mathrm{var}_i(y_1^i) & & & & \\ \vdots & \mathrm{var}_i(y_2^i) & & & \\ \vdots & \cdots & \ddots & & \\ \mathrm{cov}(y_1^i, y_t^i) & \cdots & \mathrm{var}_i(y_t^i) & & \\ \vdots & & & \ddots & \\ \mathrm{cov}(y_1^i, y_T^i) & & & & \mathrm{var}_i(y_T^i) \end{bmatrix}$$

4. Choose $(\rho, \sigma_\alpha^2, \sigma_\eta^2, \sigma_\varepsilon^2)$ to bring the theoretical covariance matrix as close to its empirical part as possible.

Moving Beyond the Covariance Matrix

- Covariances lump many disparate features of the data into one statistic.
    - Discards lots of useful information.

Moving Beyond the Covariance Matrix

- Covariances lump many disparate features of the data into one statistic.
    - Discards lots of useful information.
- Also ignores higher-order moments, which we find to be very important.

Moving Beyond the Covariance Matrix

- Covariances lump many disparate features of the data into one statistic.
  - Discards lots of useful information.
- Also ignores higher-order moments, which we find to be very important.
- With CME, selecting among rejected models is very hard:
  - moments that are missed do not have clear economic interpretations.

This Paper

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

This Paper

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

1. Document new empirical facts on life-cycle earnings dynamics

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 earnings dynamics
    - by matching economically important moments (as opposed to the "covariance matrix of income residuals")

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 earnings dynamics
    - 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

Data: SSA Master Earnings File

- Representative sample of US males covering 34 years: 1978 to 2011

Data: SSA Master Earnings File

- Representative sample of US males covering 34 years: 1978 to 2011
- Salary and wage workers (from W-2 forms)

Data: SSA Master Earnings File

- Representative sample of US males covering 34 years: 1978 to 2011
- Salary and wage workers (from W-2 forms)
- Individuals aged 25–60

Data: SSA Master Earnings File

- Representative sample of US males covering 34 years: 1978 to 2011
- Salary and wage workers (from W-2 forms)
- Individuals aged 25–60
- Key Advantages:
  - Very large sample size (200+ million observations)
  - No survey response error
  - No sample attrition
  - No top-coding

Four Sets of Empirical Facts

1. Average income growth over the life cycle

Four Sets of Empirical Facts

1. Average income growth over the life cycle

2. Cross-sectional moments of earnings growth

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

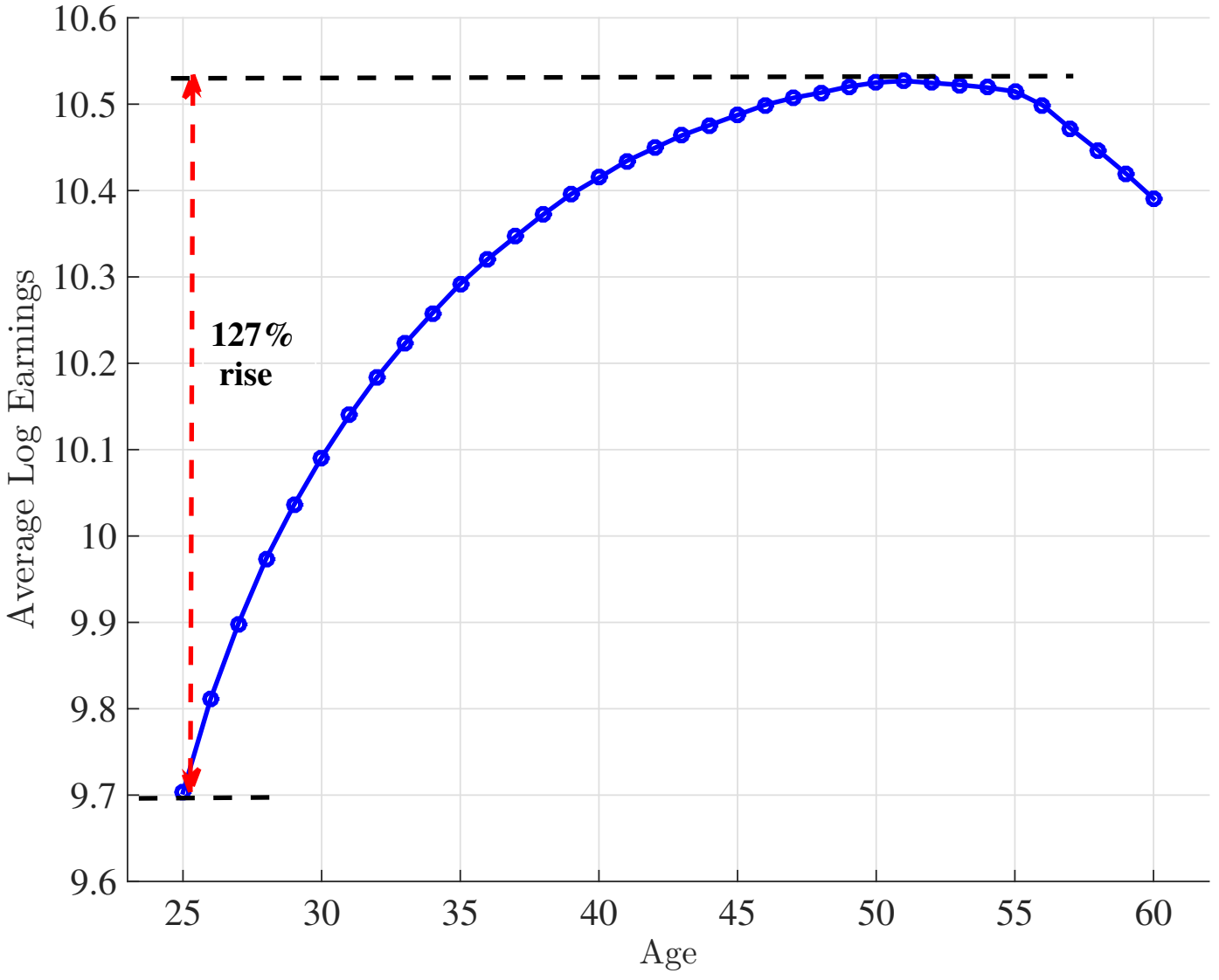
Four Sets of Empirical Facts

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)

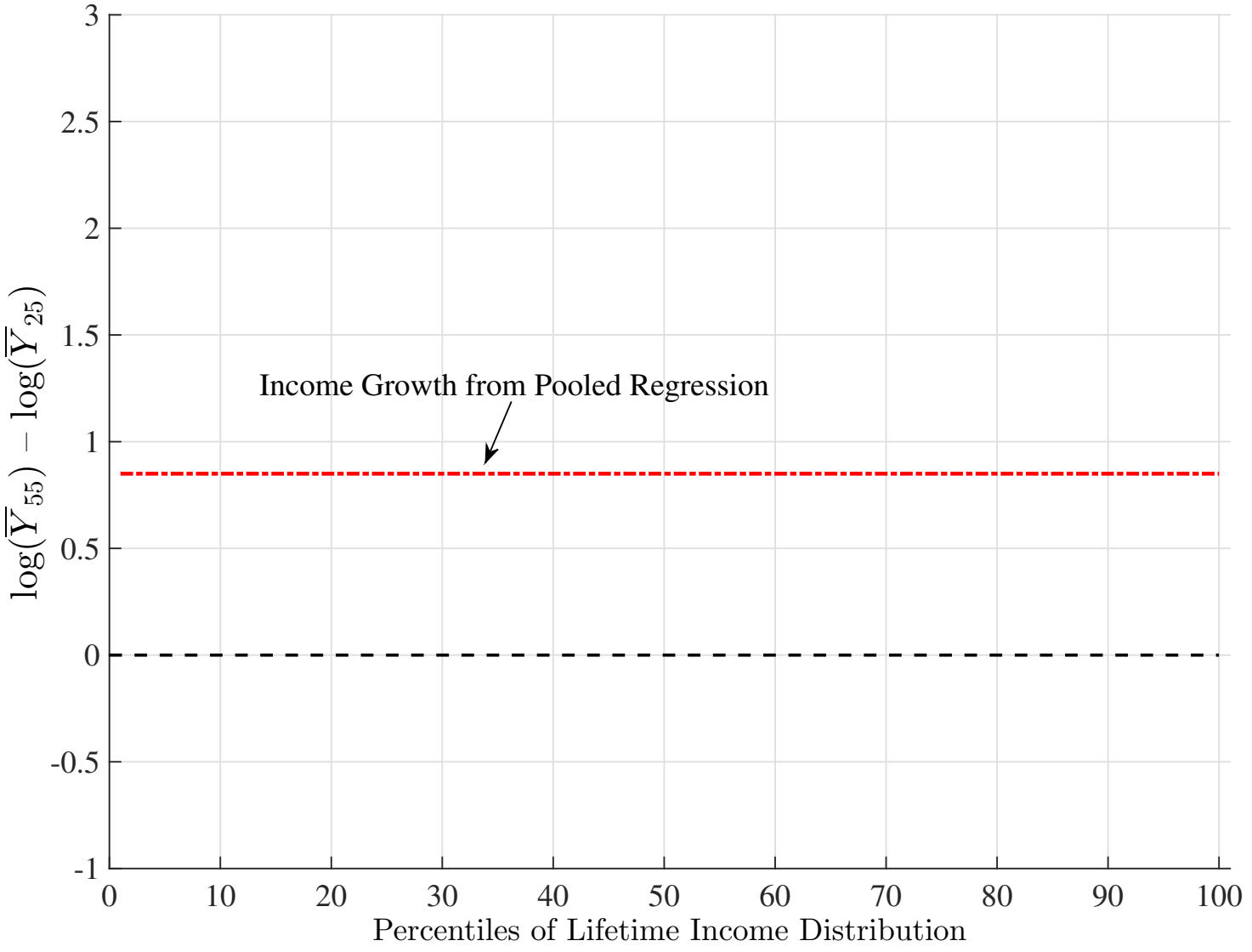
Four Sets of Empirical Facts

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

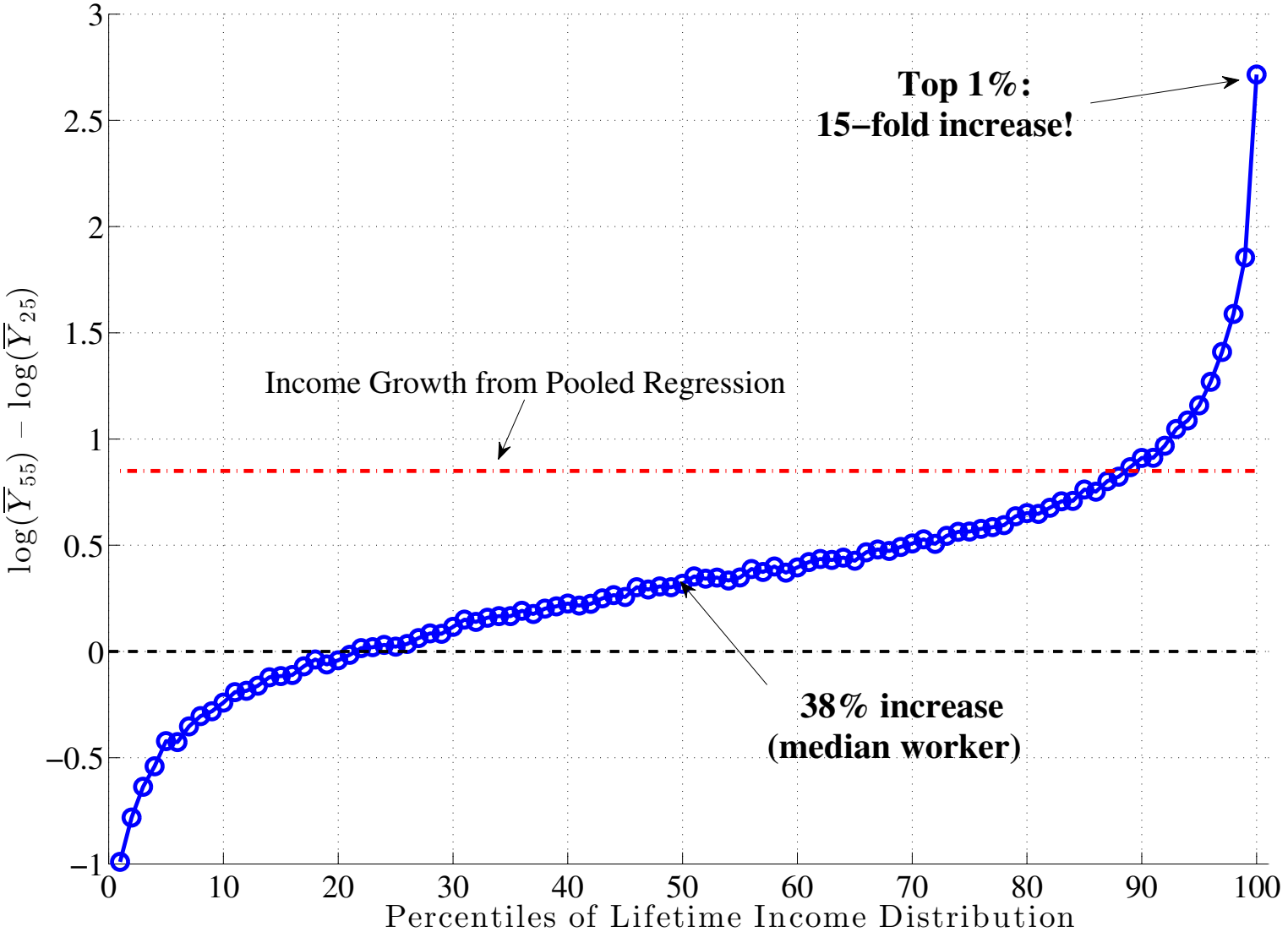
I. Age Profile of Labor Income



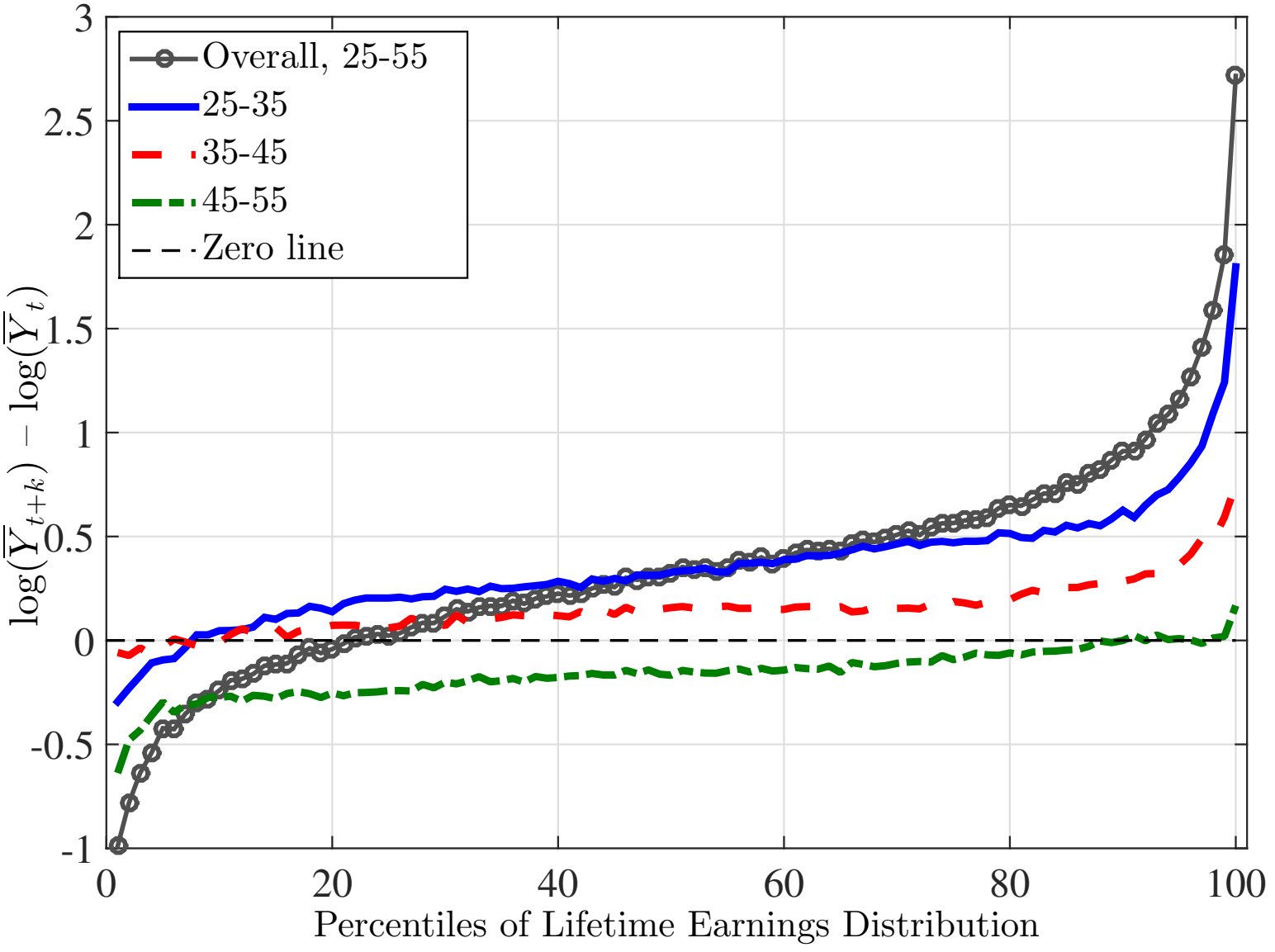
I. Income Growth Over Life Cycle



I. Income Growth Over Life Cycle



I. Income Growth Over Life Cycle

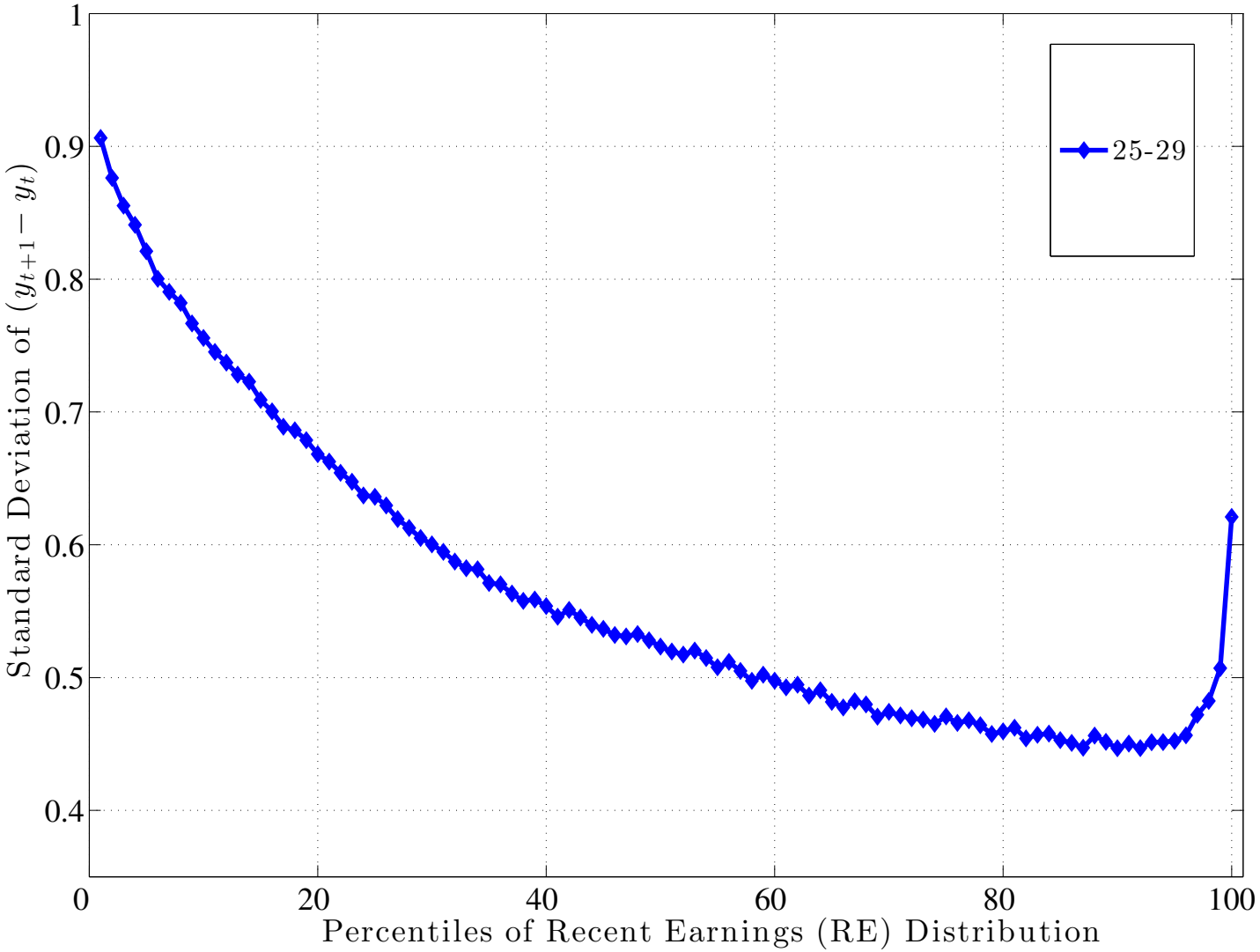


Four Sets of Empirical Facts

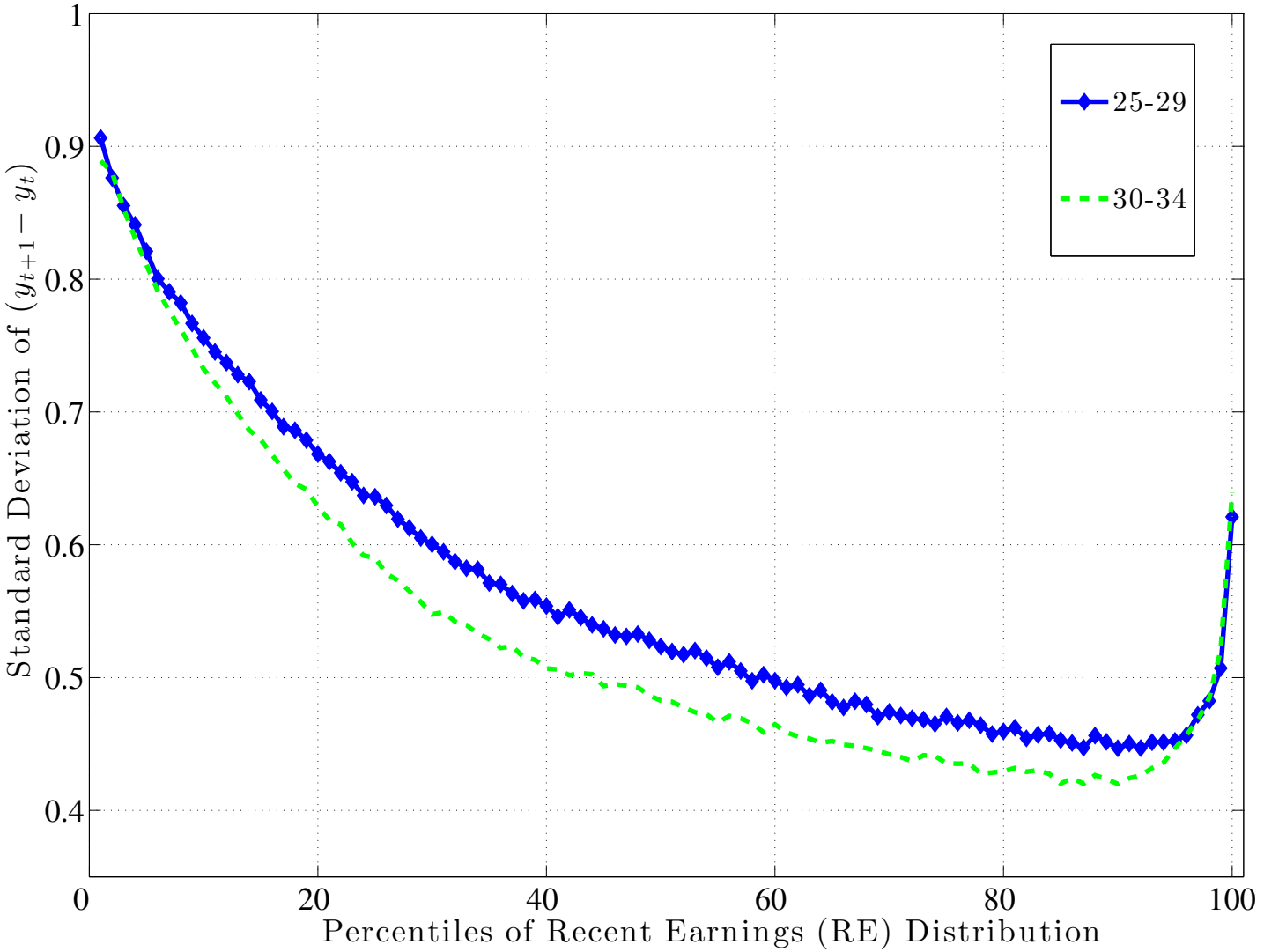
1. Income growth over the life cycle
2. Cross-sectional moments of earnings growth: $y_{t+k} - y_t$
3.
4.

Standard Deviation

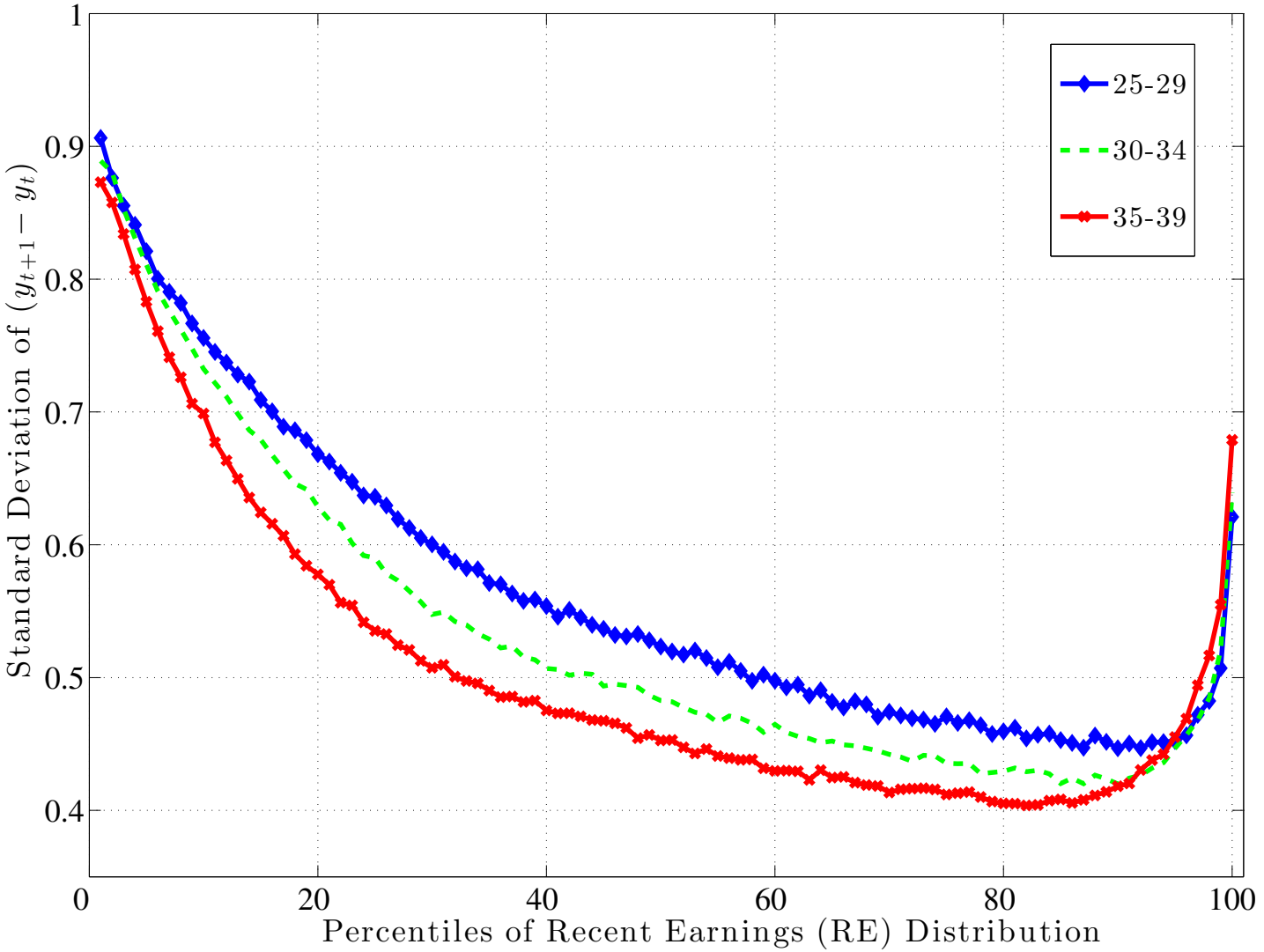
II.A Standard Deviation of $y_{t+1} - y_t$



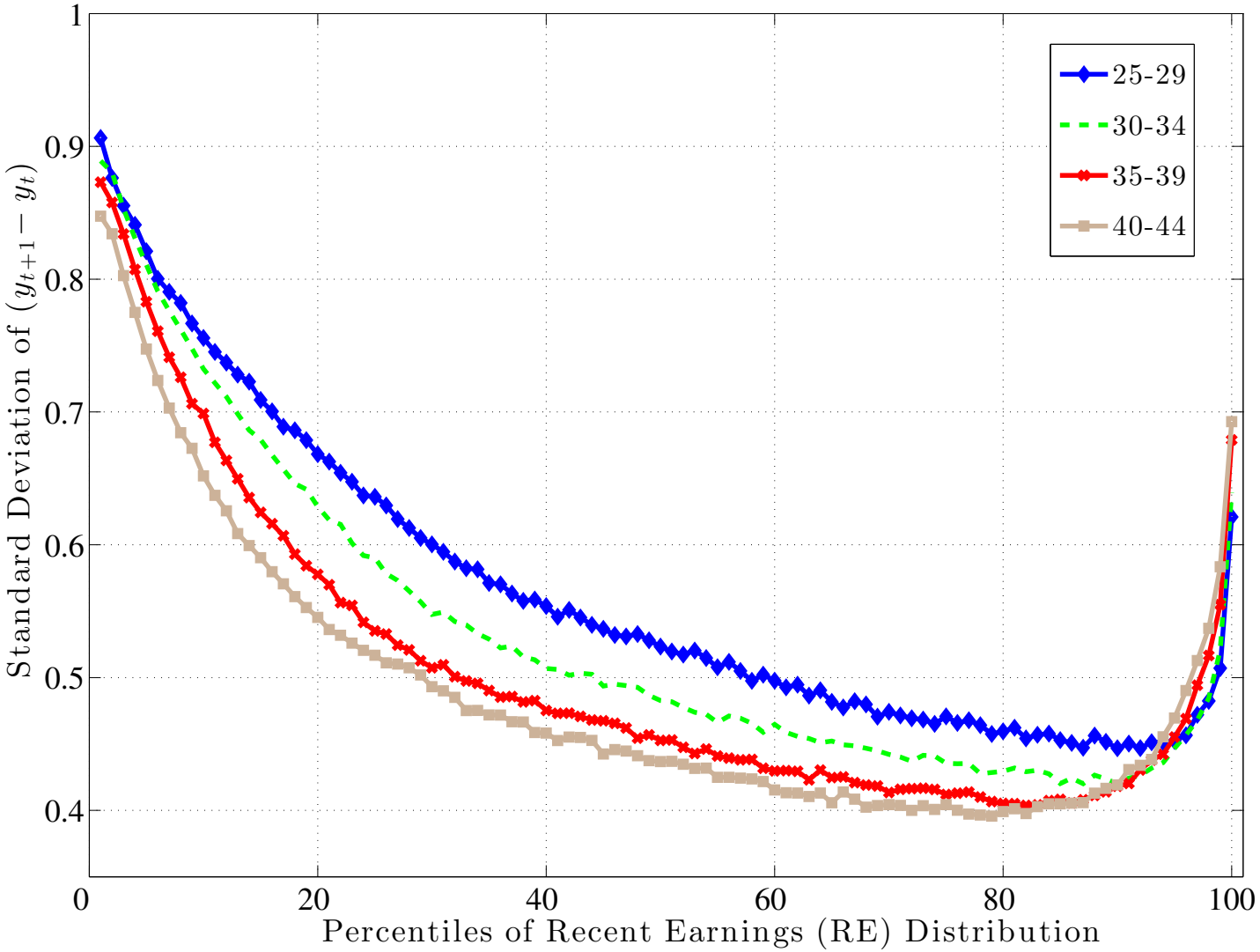
II.A Standard Deviation of $y_{t+1} - y_t$



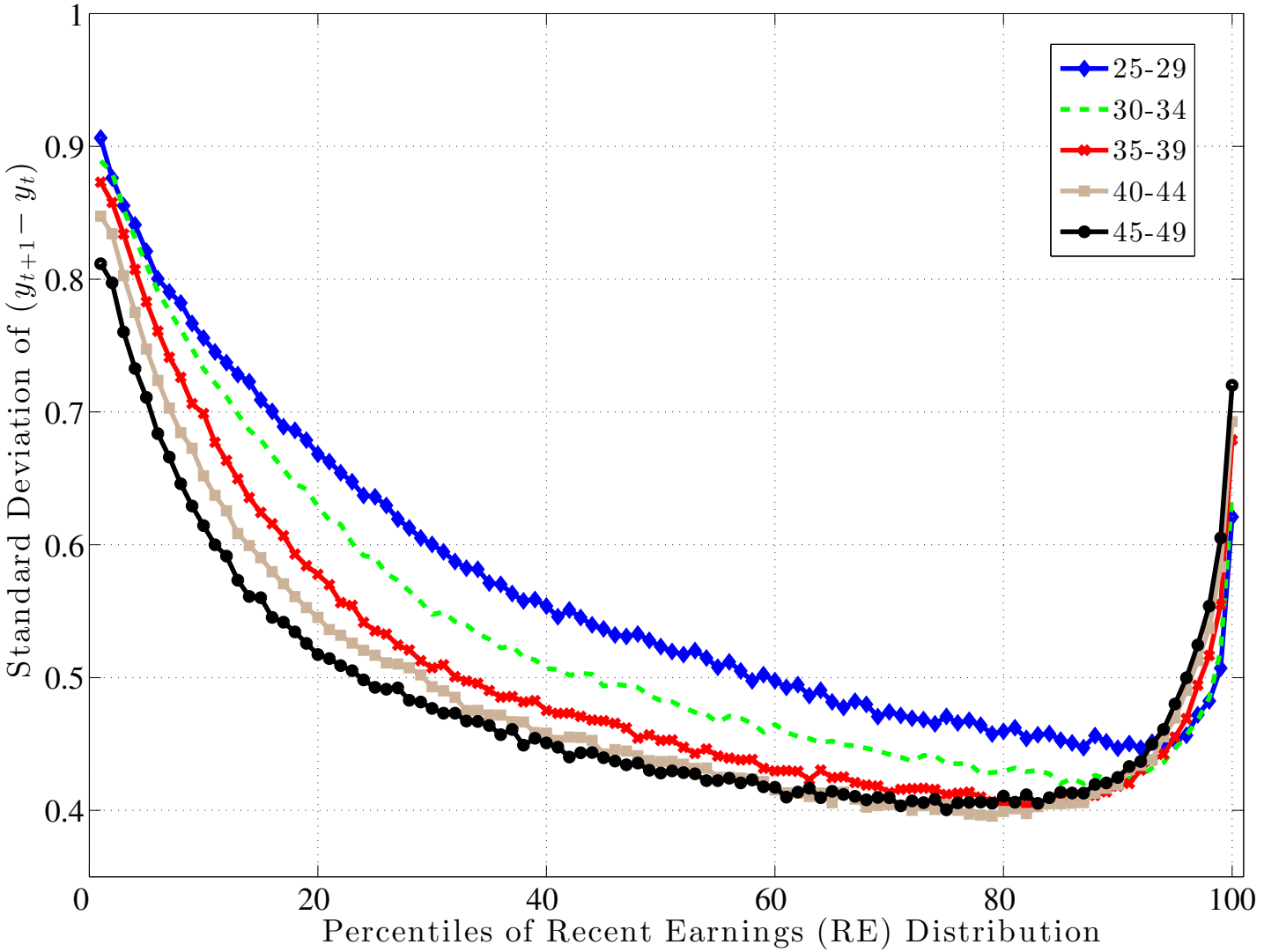
II.A Standard Deviation of $y_{t+1} - y_t$



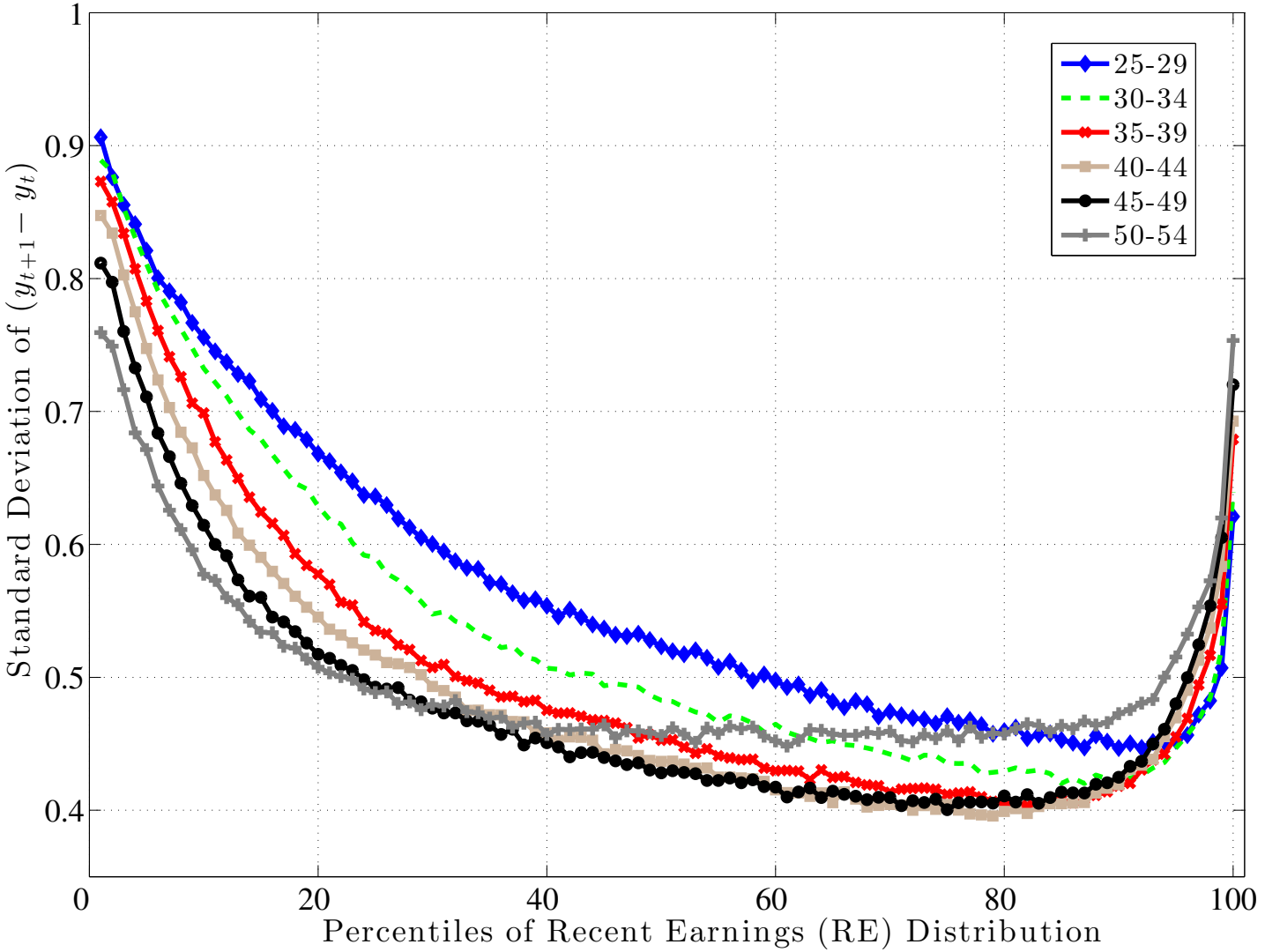
II.A Standard Deviation of $y_{t+1} - y_t$



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

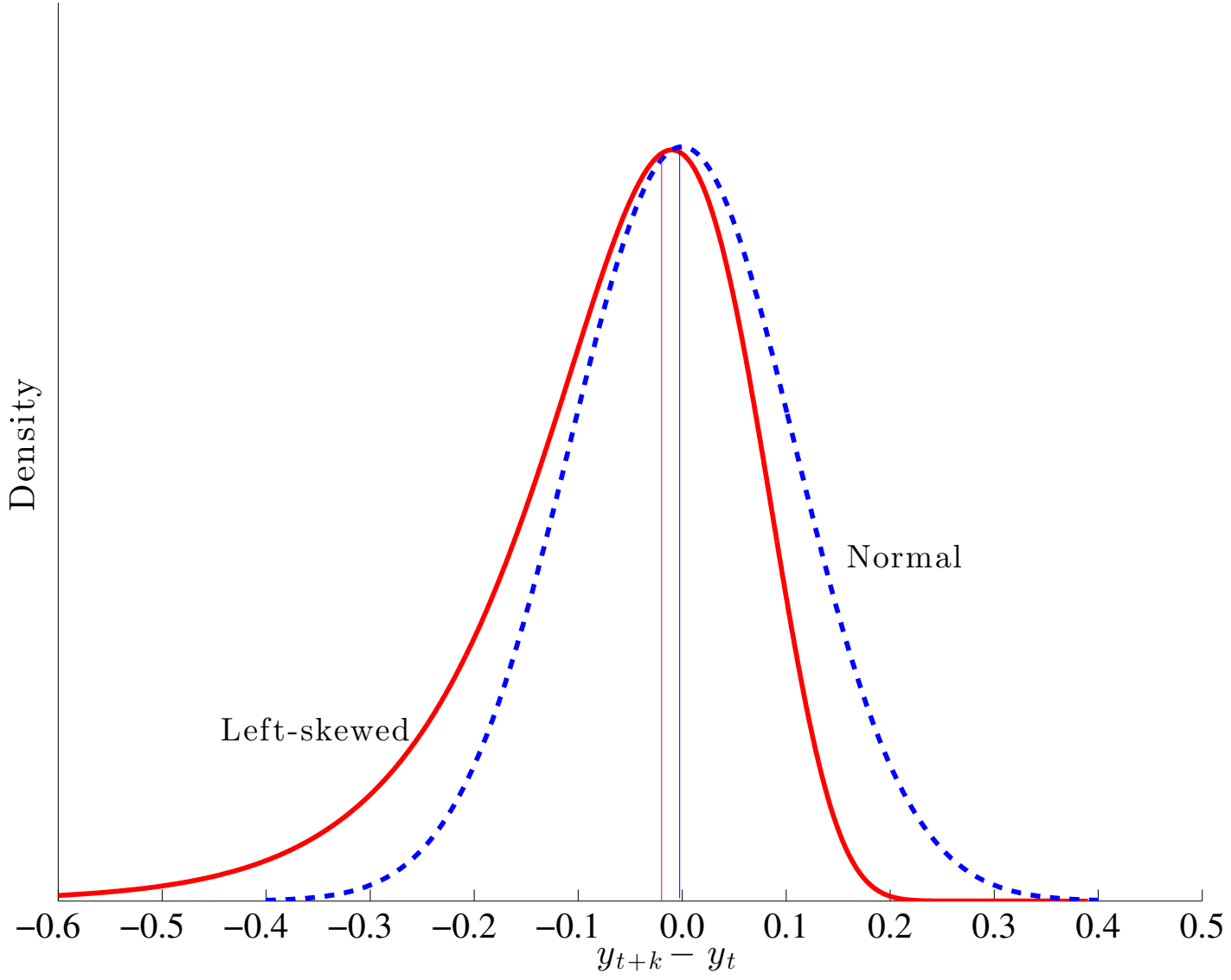


II.A Standard Deviation of $y_{t+1} - y_t$

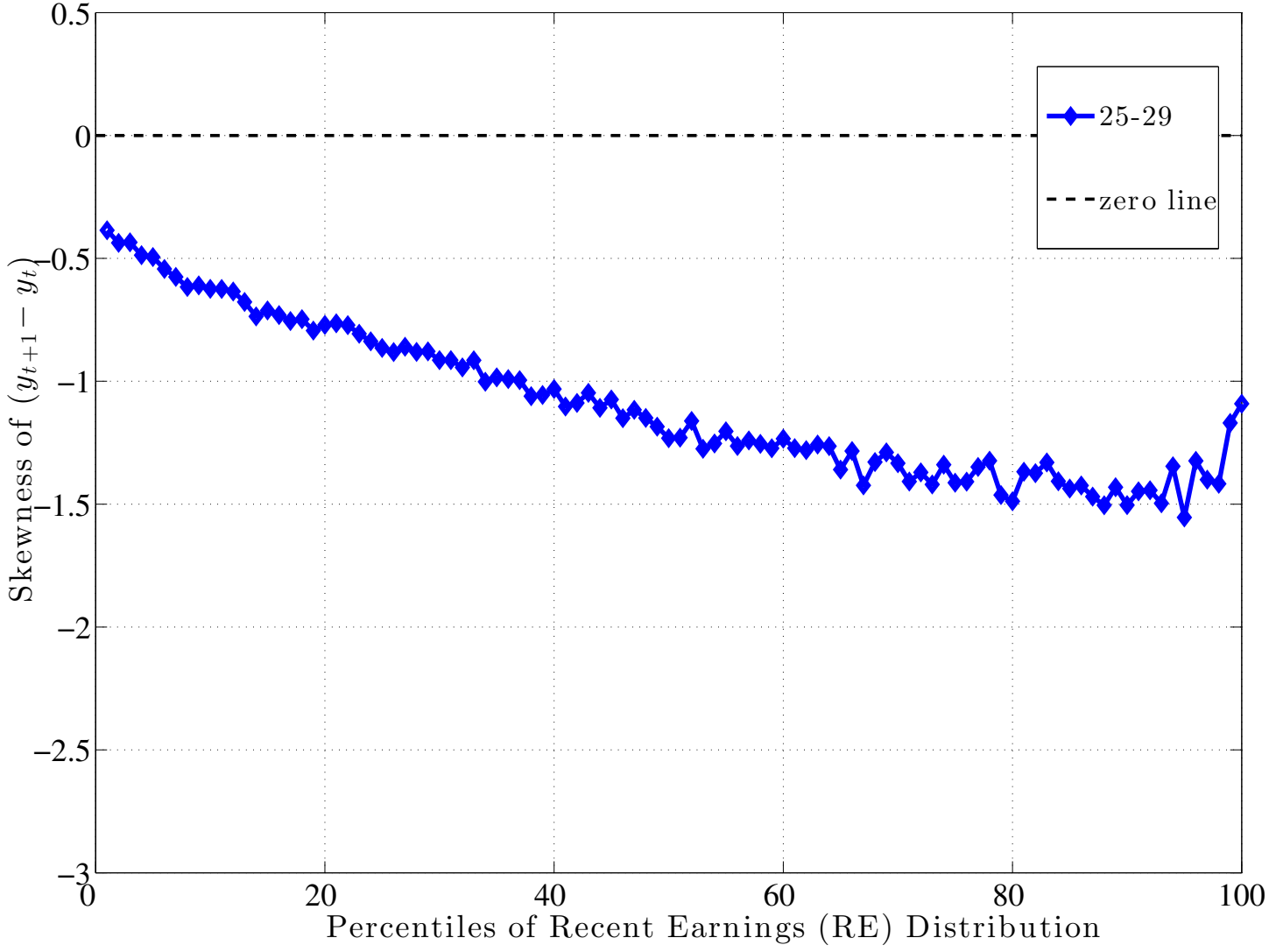


Skewness

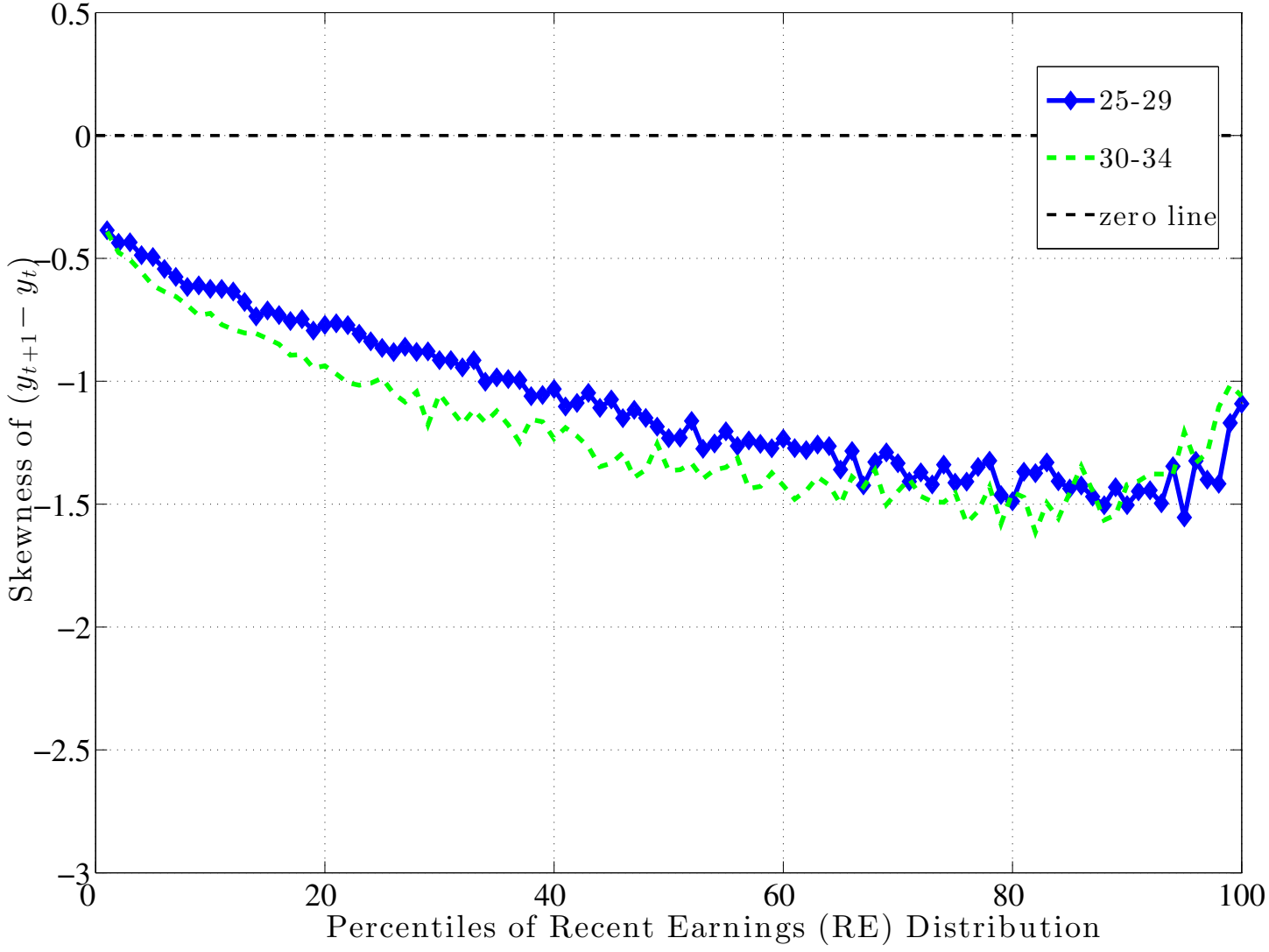
Left-Skewness



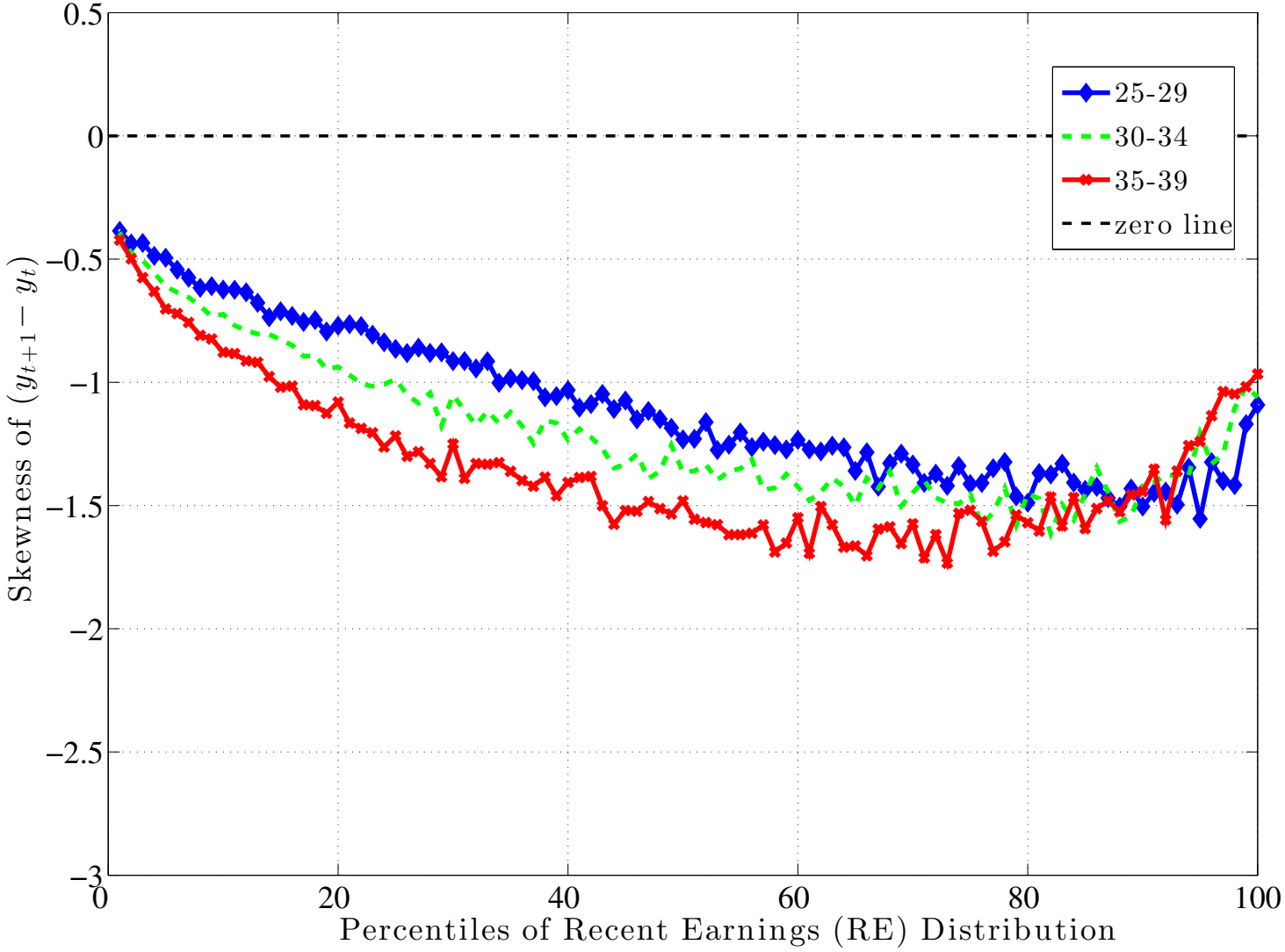
II.B Skewness of $y_{t+1} - y_t$



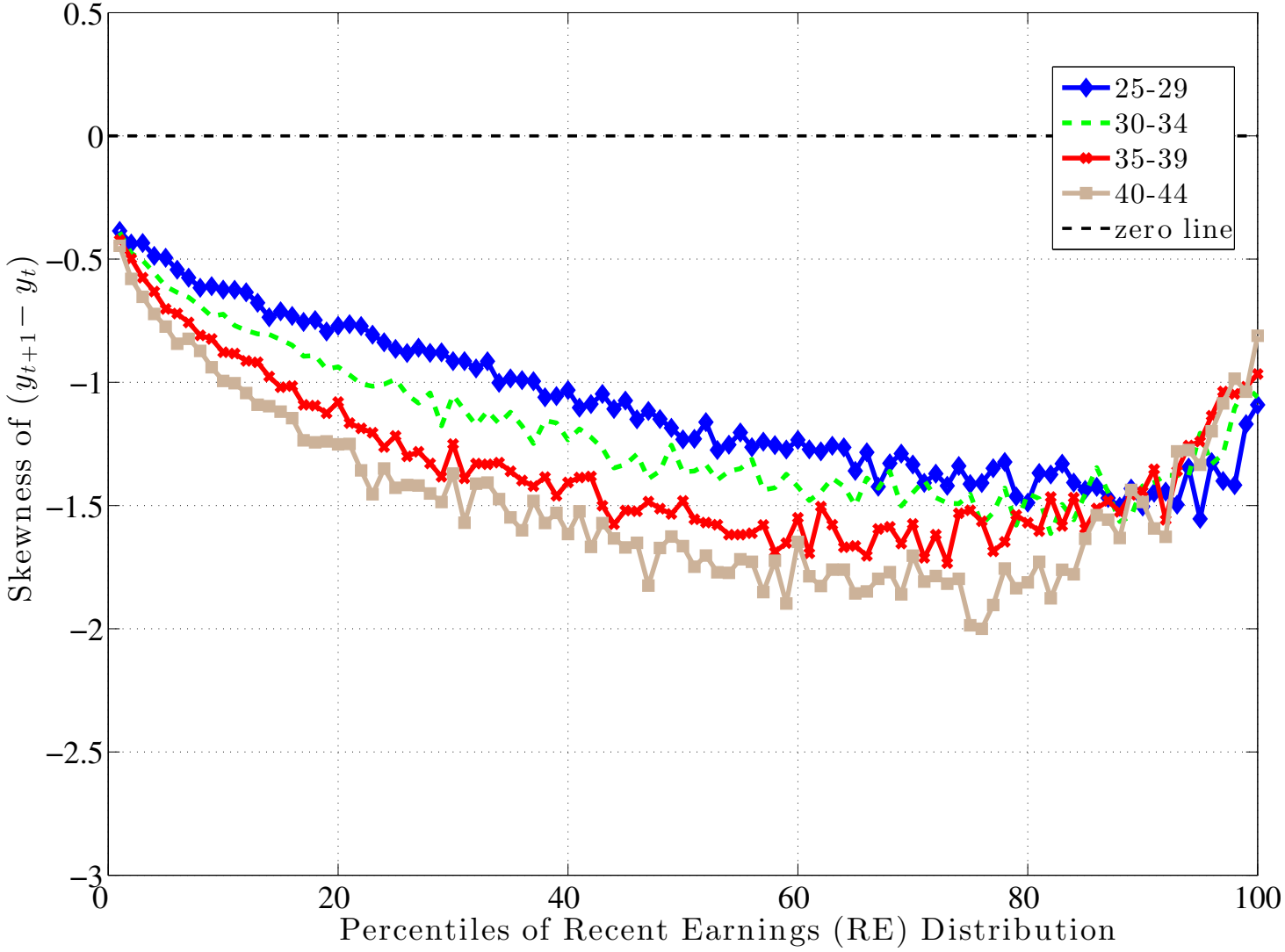
II.B Skewness of $y_{t+1} - y_t$



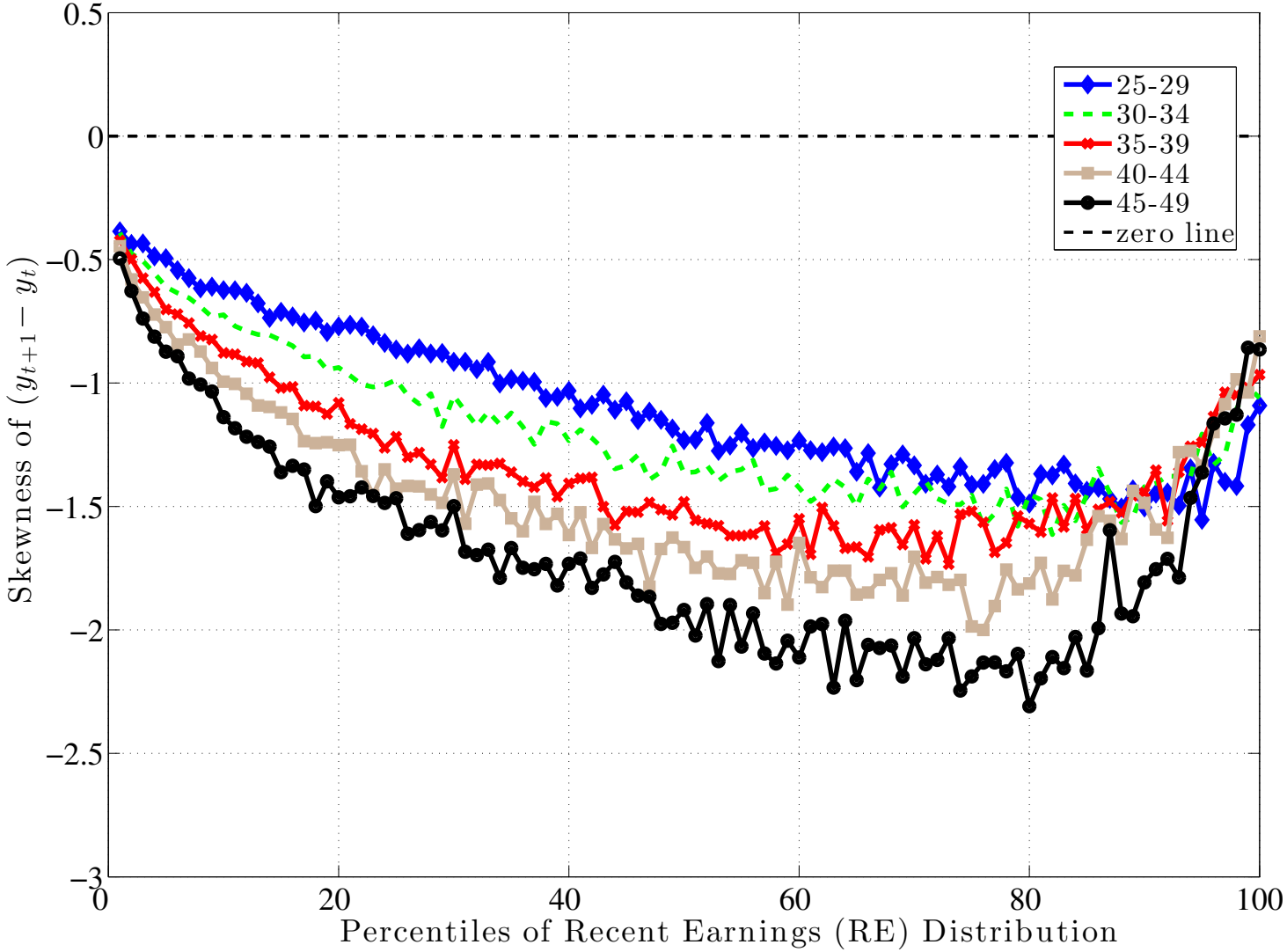
II.B Skewness of $y_{t+1} - y_t$



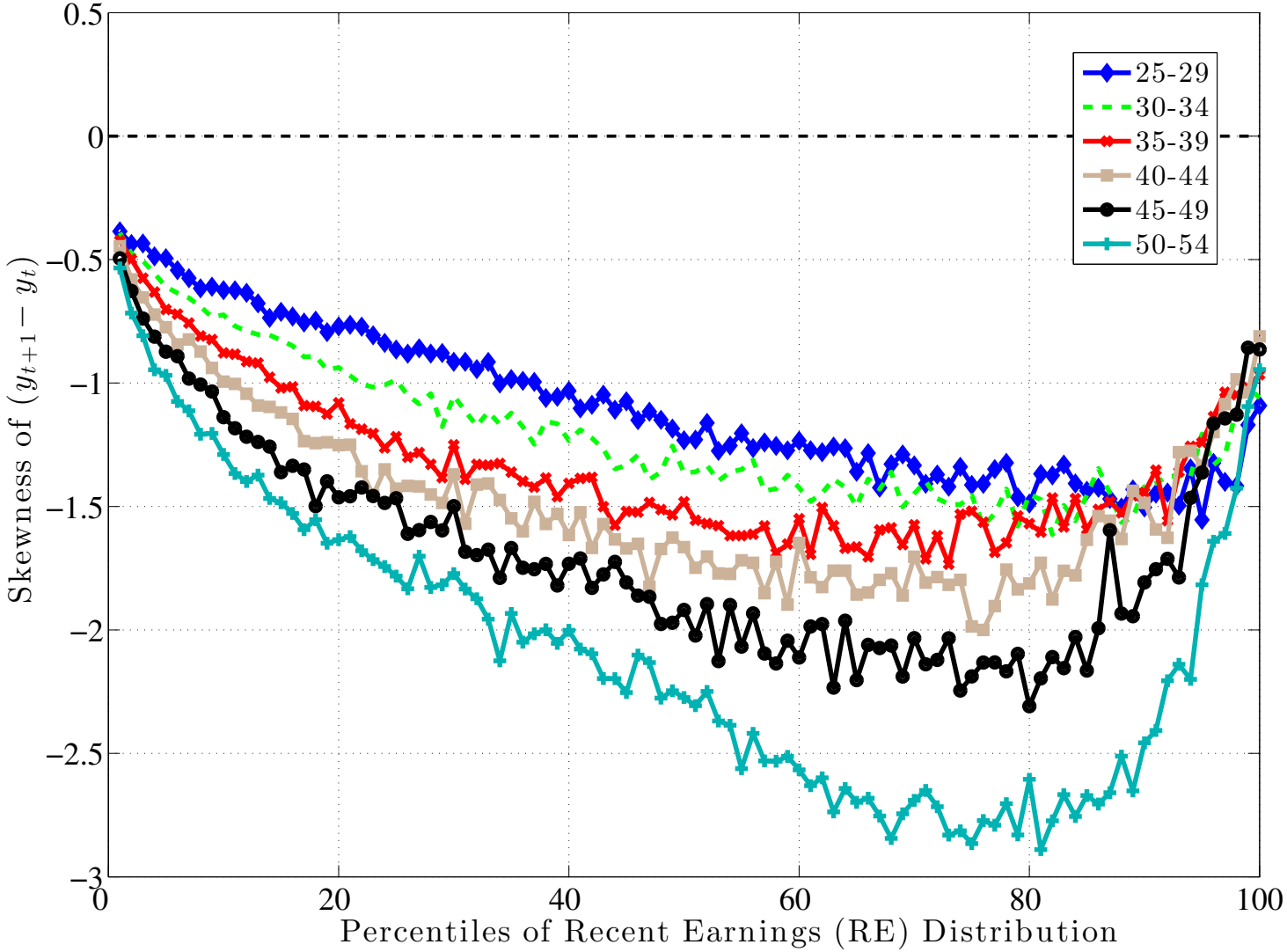
II.B Skewness of $y_{t+1} - y_t$



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

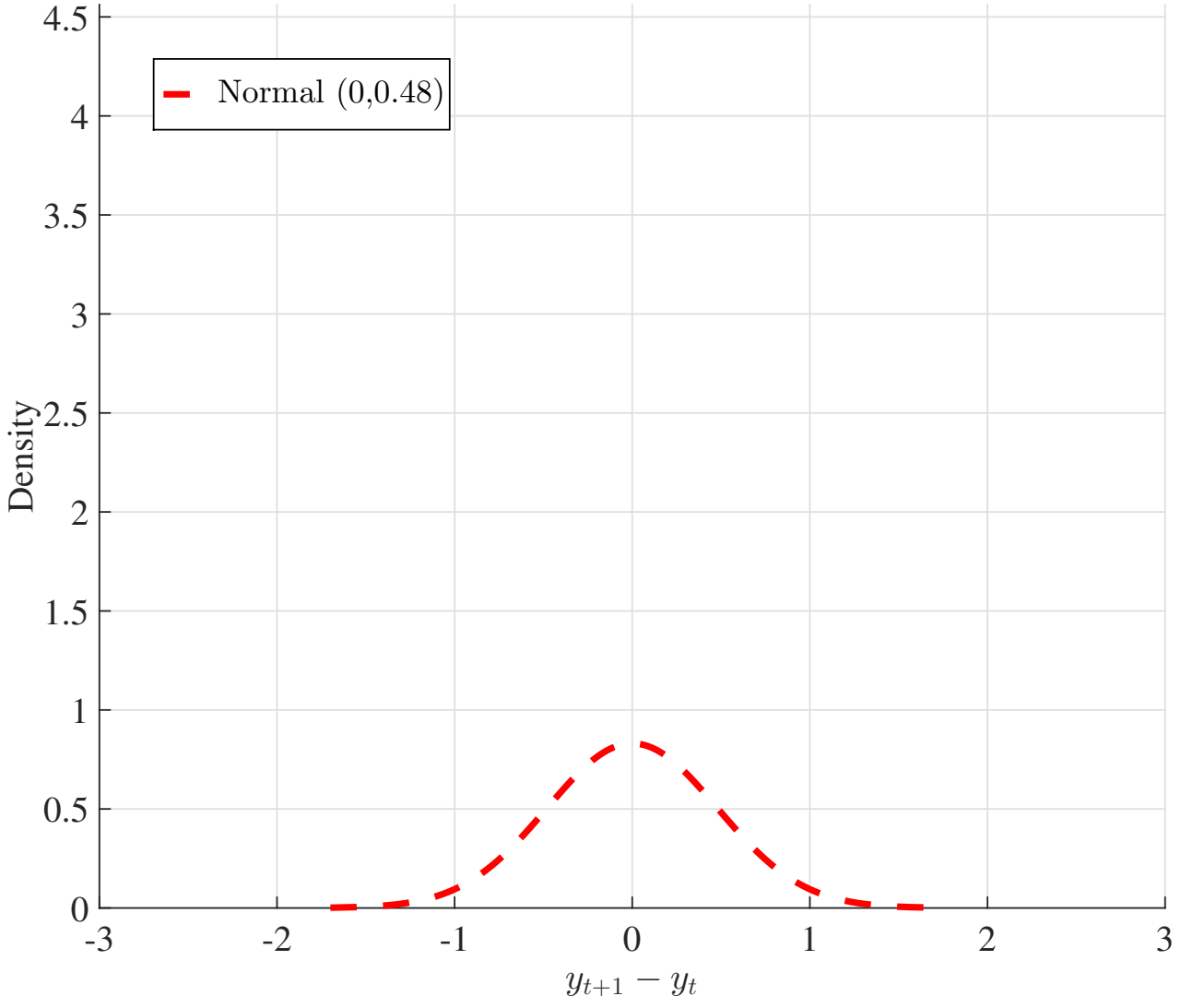


II.B Skewness of $y_{t+1} - y_t$

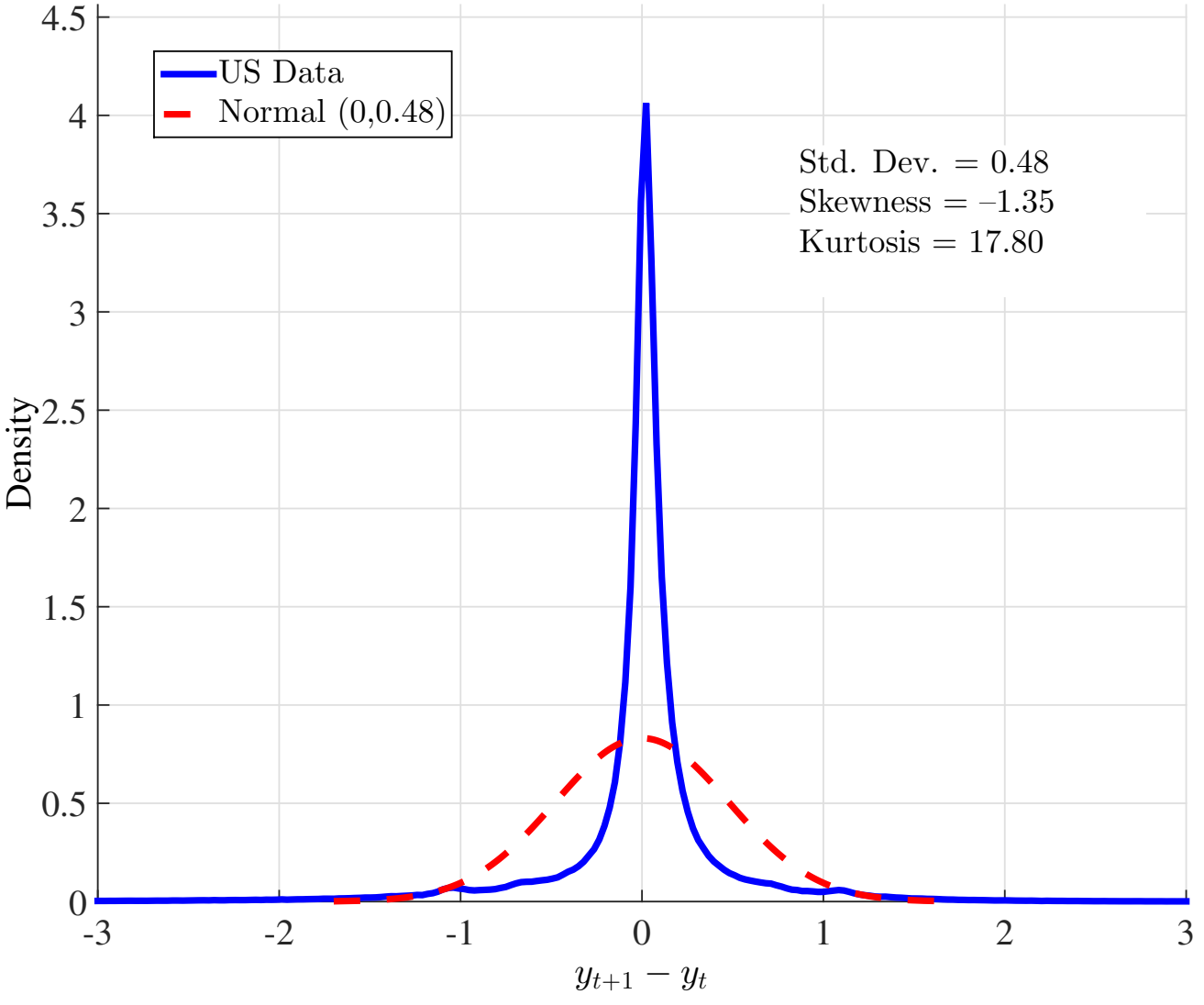


Kurtosis

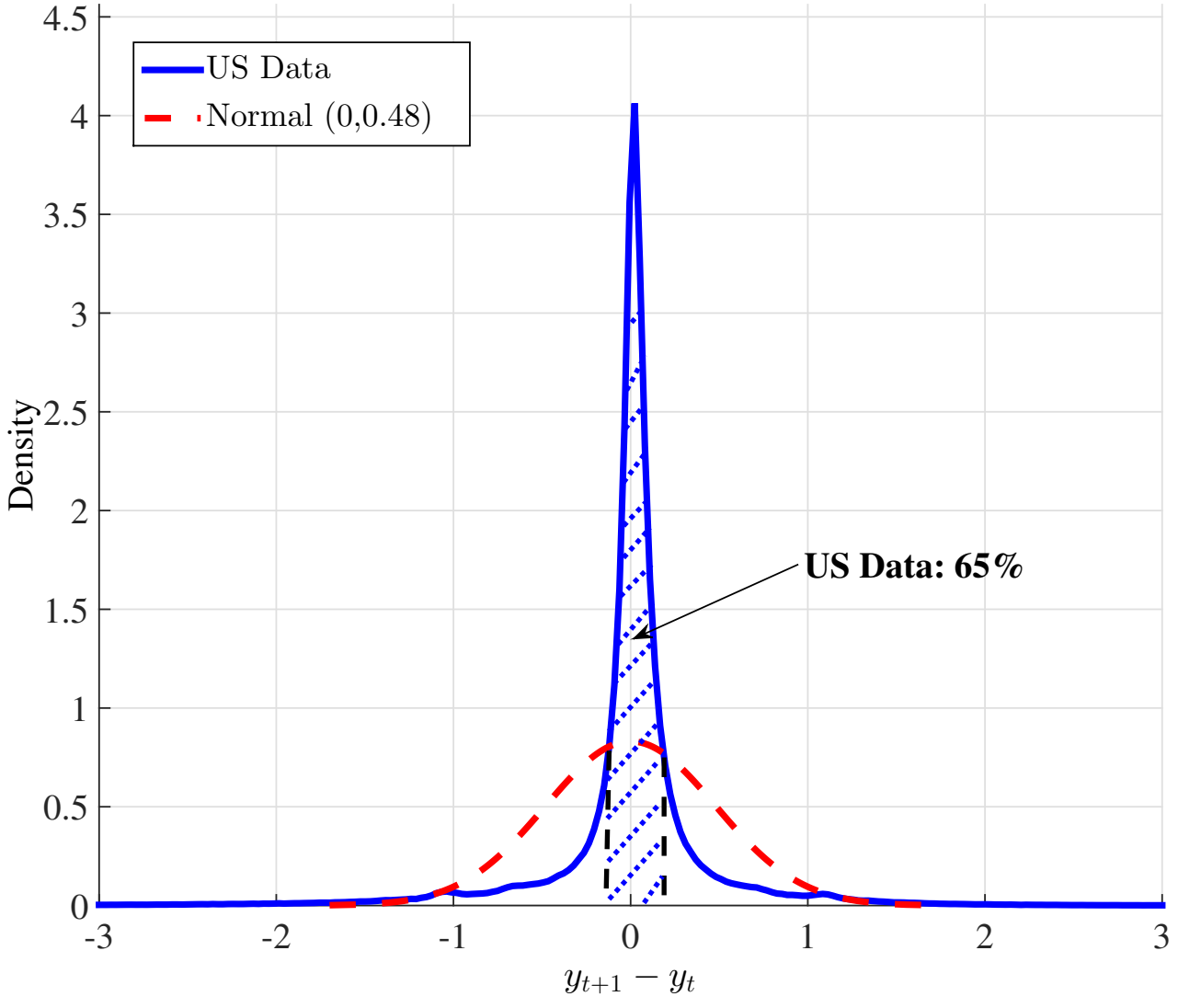
II.c Histogram of $y_{t+1} - y_t$



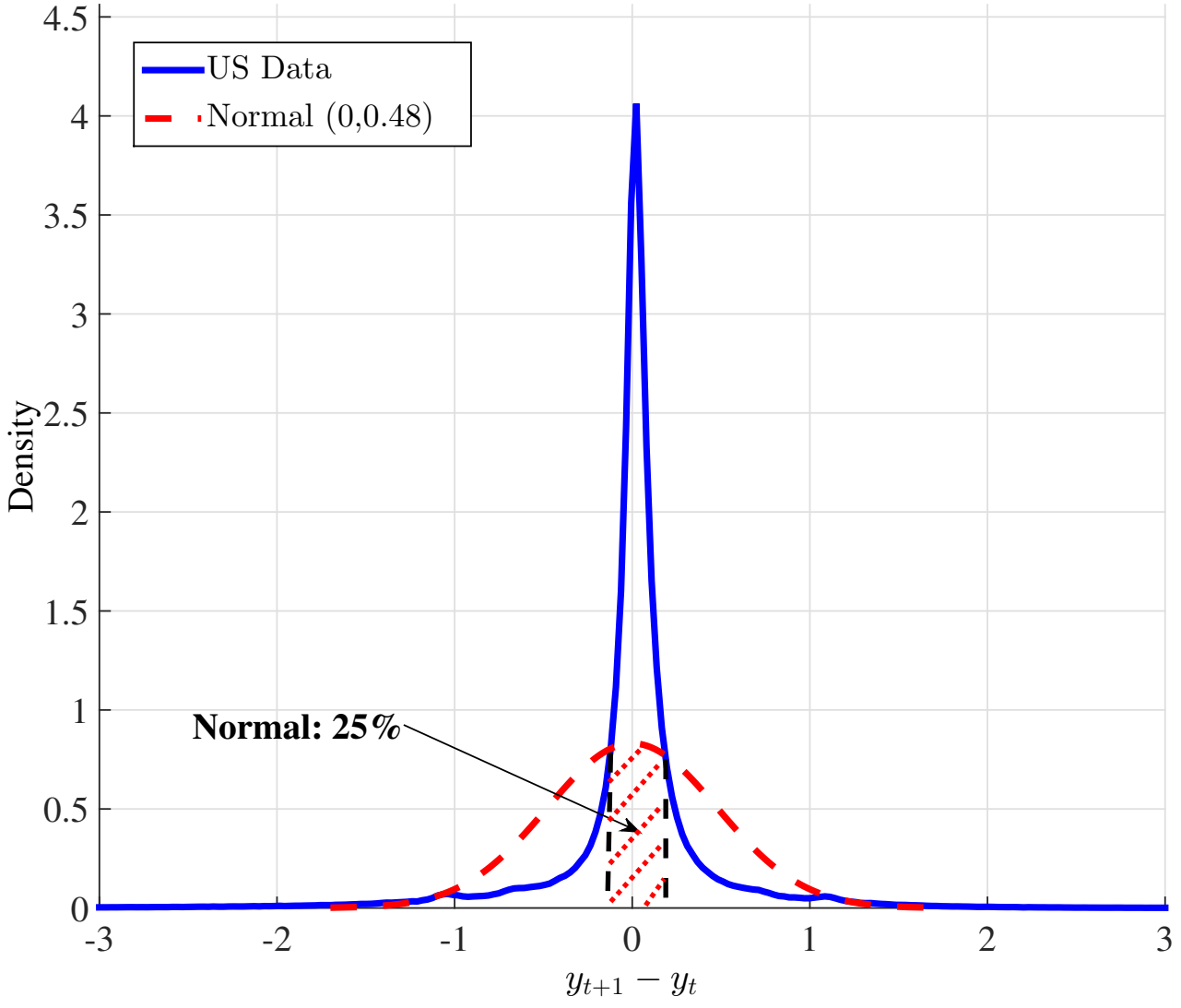
II.c Histogram of $y_{t+1} - y_t$



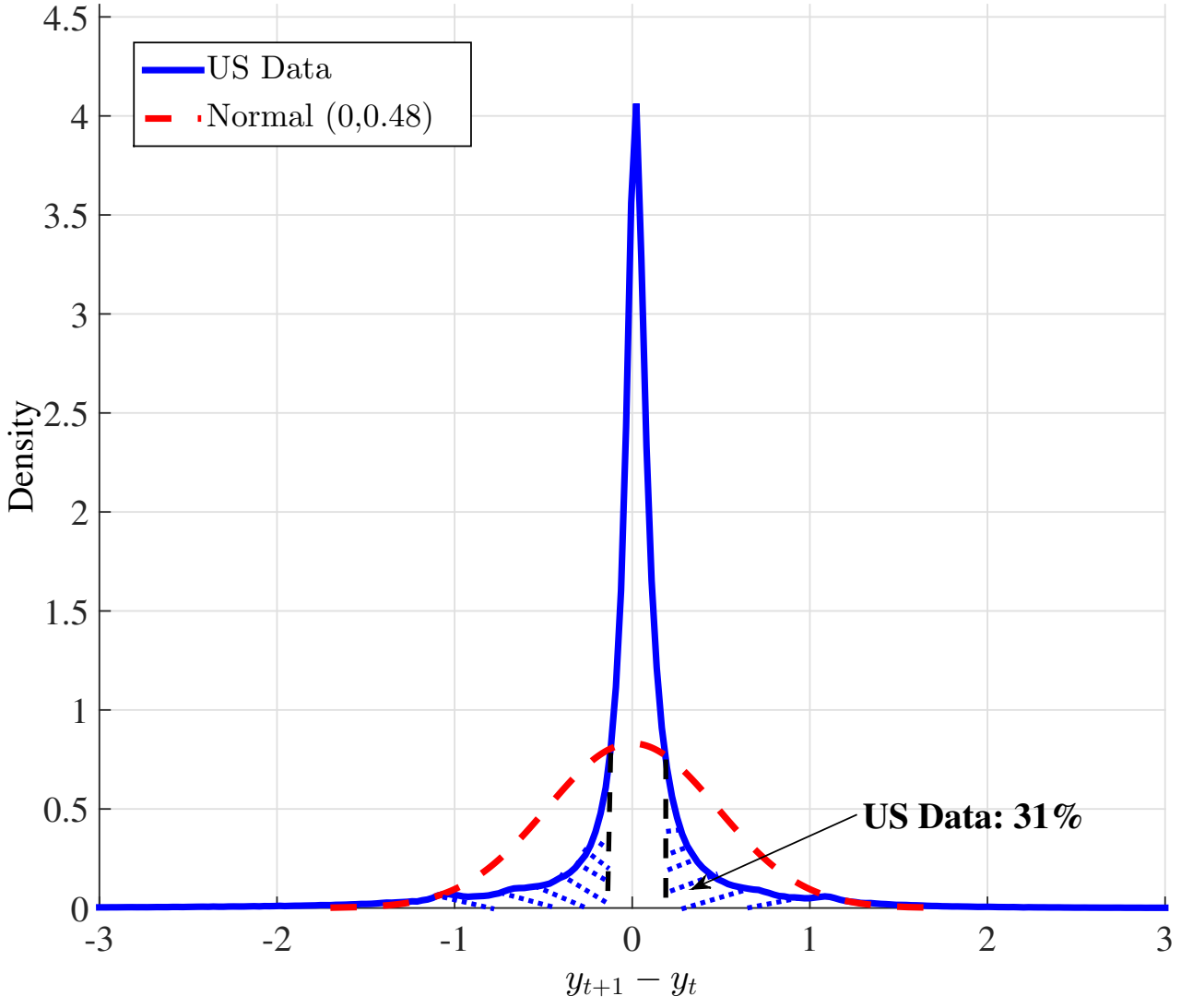
II.c Center of Histogram: [-0.12, 0.19]



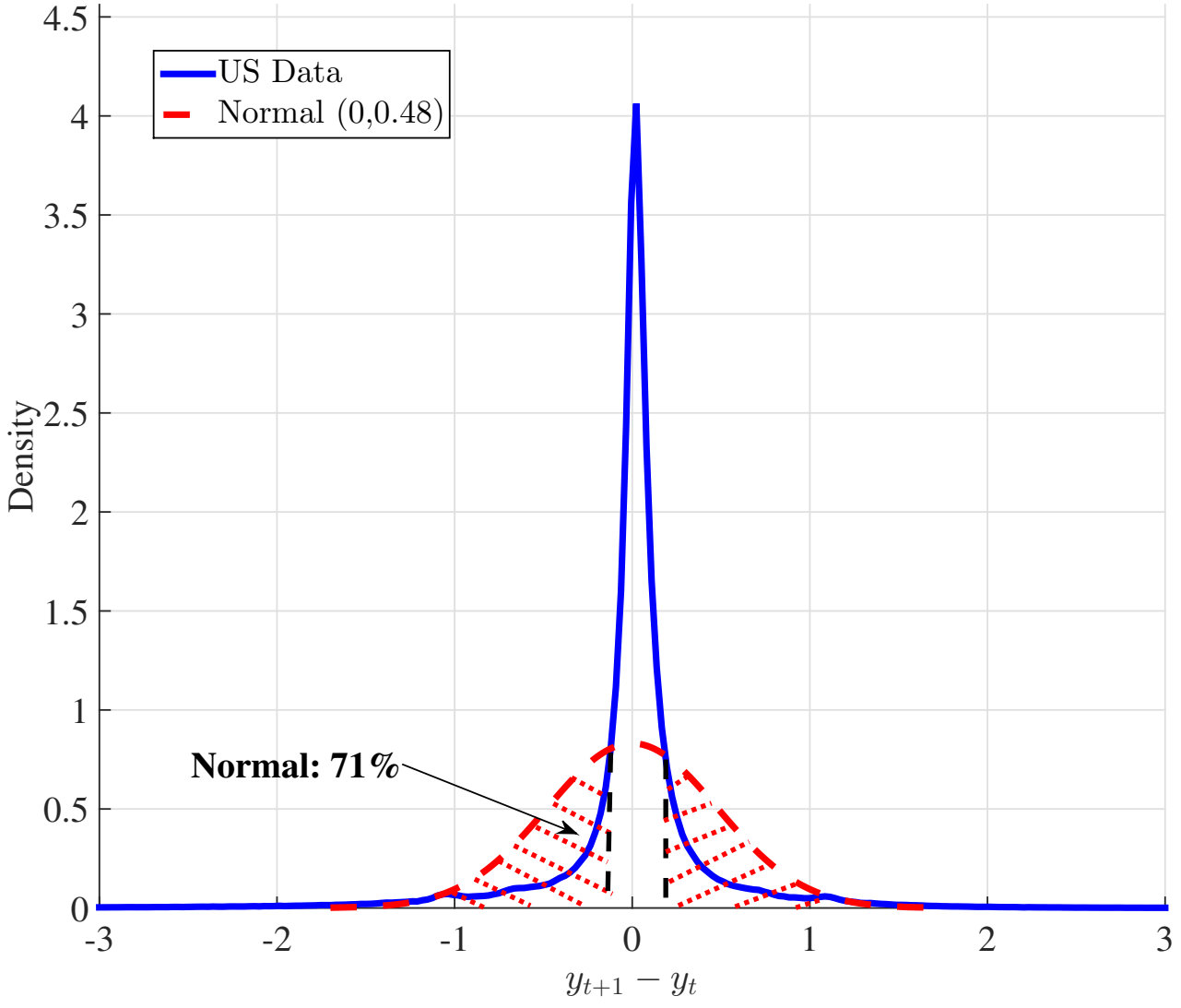
II.c Center of Histogram: [-0.12, 0.19]



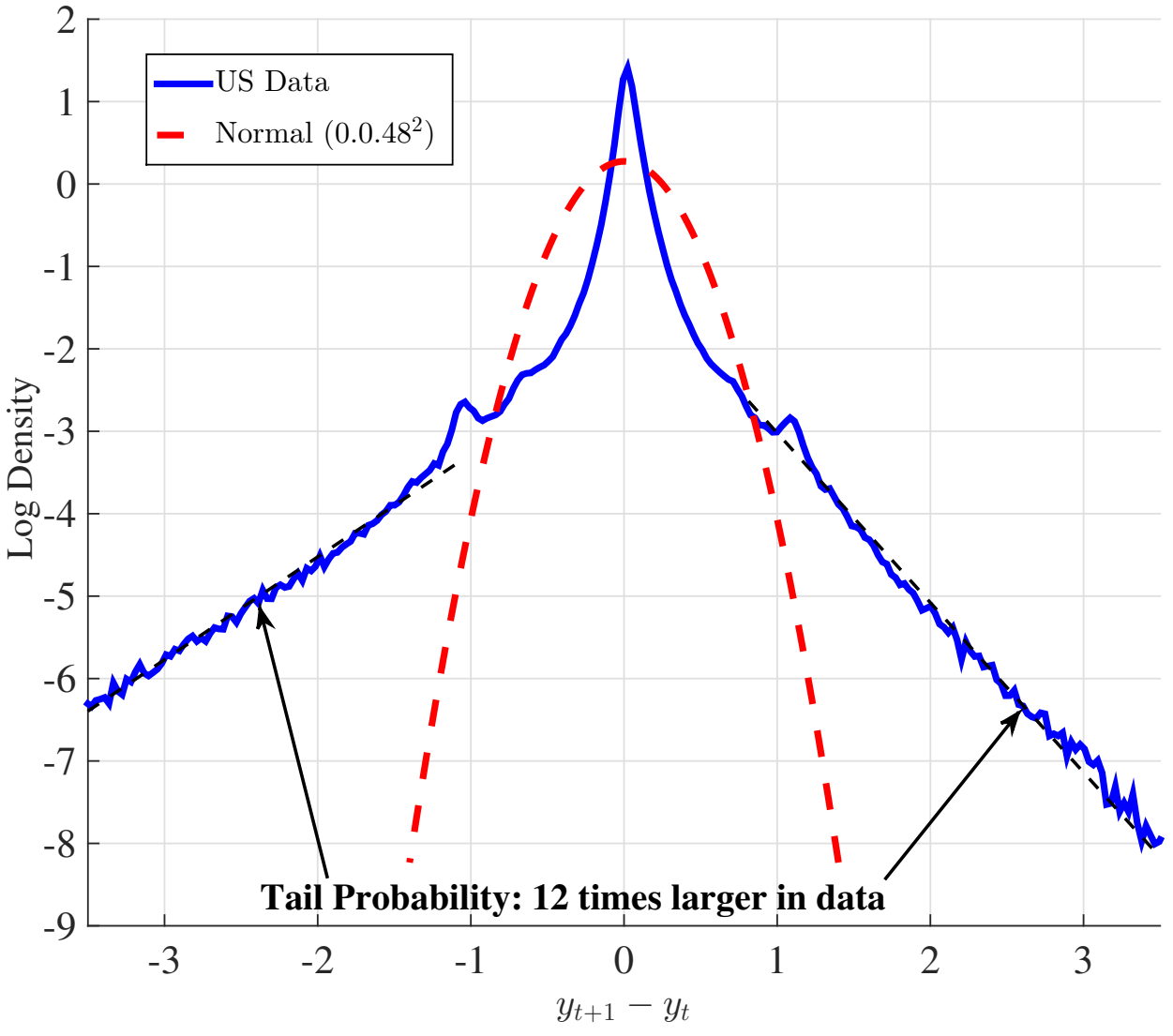
II.c Shoulders of Histogram



II.c Shoulders of Histogram



I. Tails of Histogram



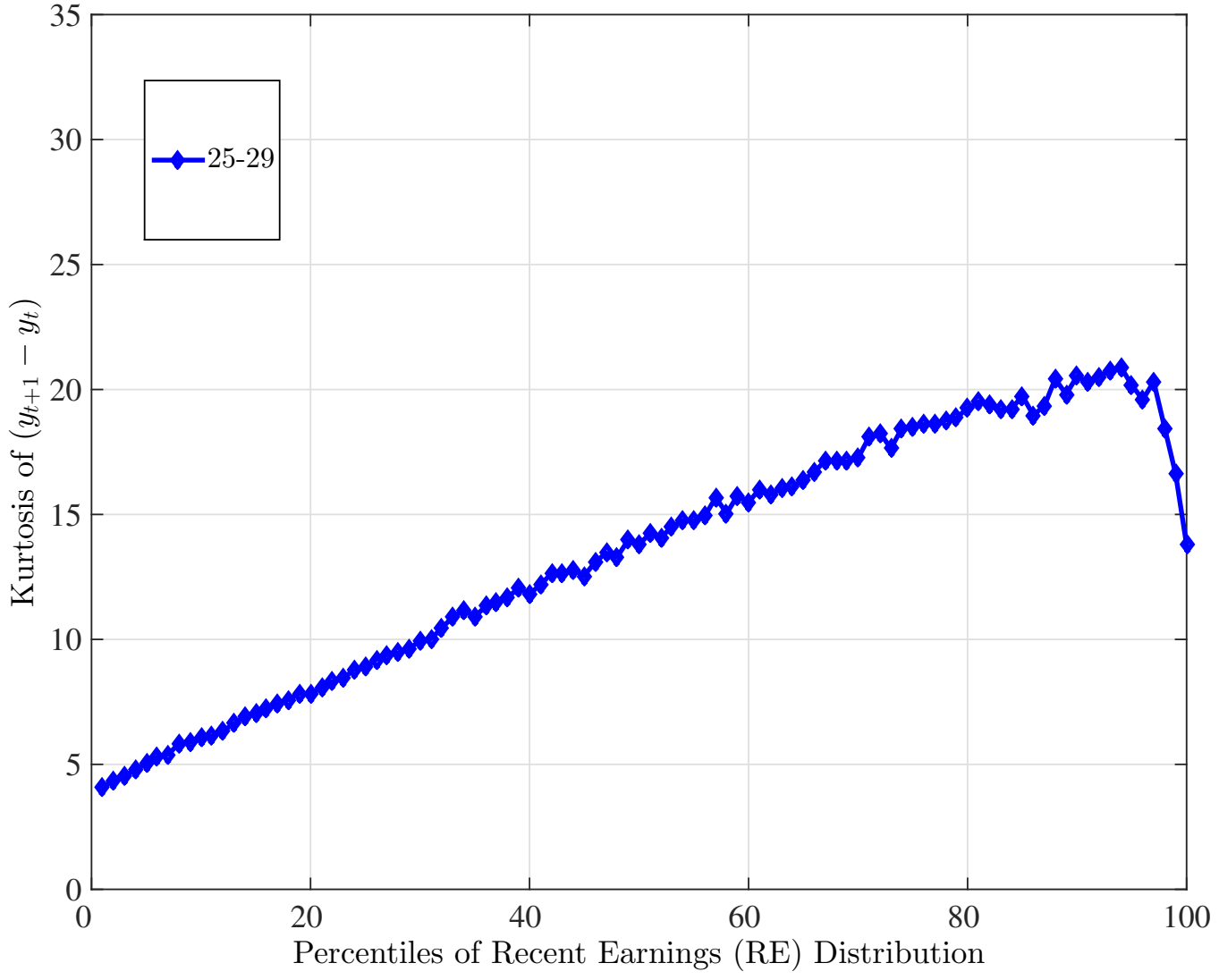
II.c Distribution of Income Changes

	Prob($\lvert y^i_{t+1} - y^i_t \rvert < x$)		
x :	Data*	$\mathcal{N}(0, 0.48^2)$	Ratio
0.05	0.35	0.08	4.38
0.10	0.54	0.16	3.38
0.20	0.71	0.32	2.23
0.50	0.86	0.70	1.22
1.00	0.94	0.96	0.98

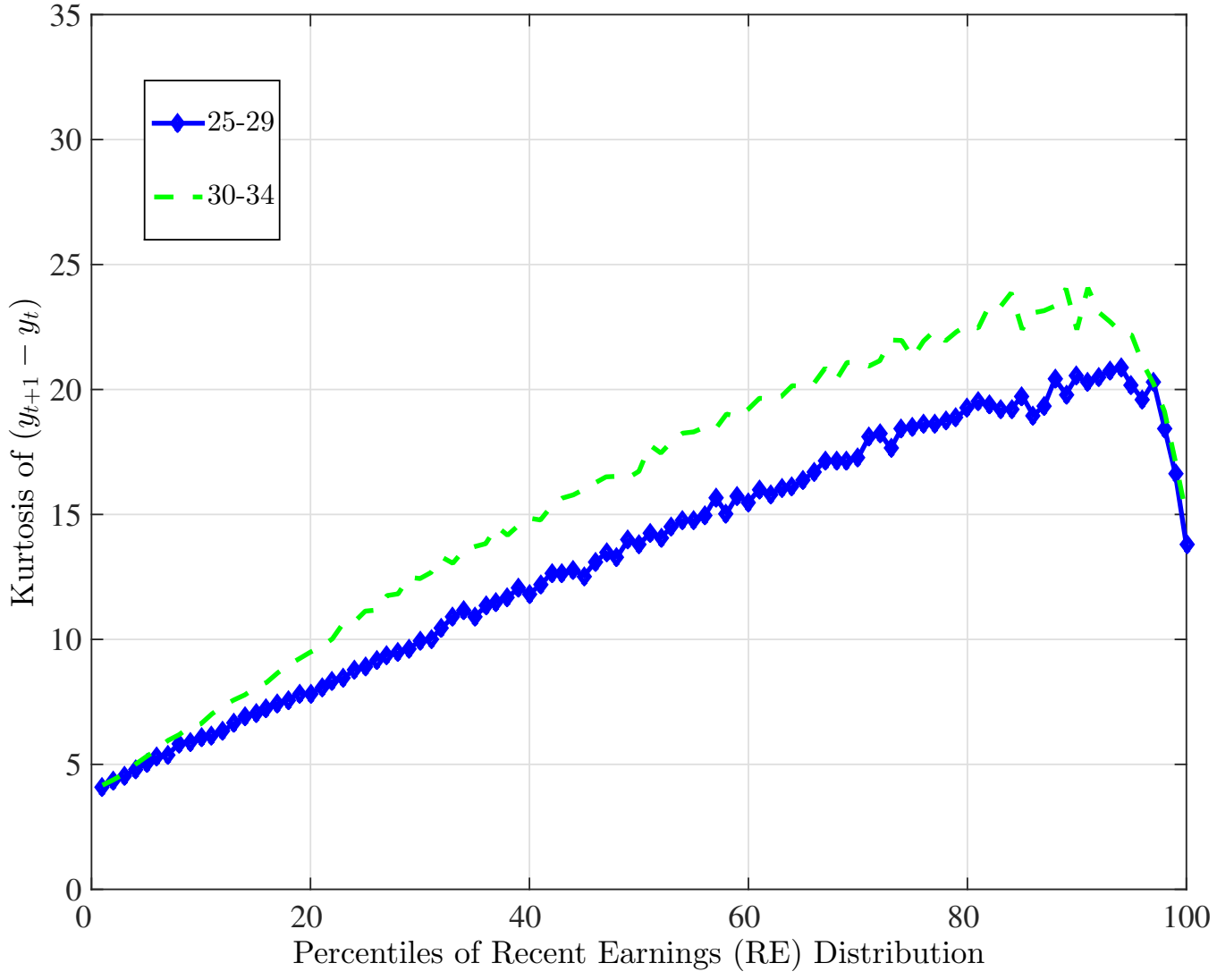
II.c Distribution of Income Changes

	Prob($\lvert y^i_{t+1} - y^i_t \rvert < x$)		
x :	Data*	$\mathcal{N}(0, 0.48^2)$	Ratio
0.05	0.35	0.08	4.38
0.10	0.54	0.16	3.38
0.20	0.71	0.32	2.23
0.50	0.86	0.70	1.22
1.00	0.94	0.96	0.98

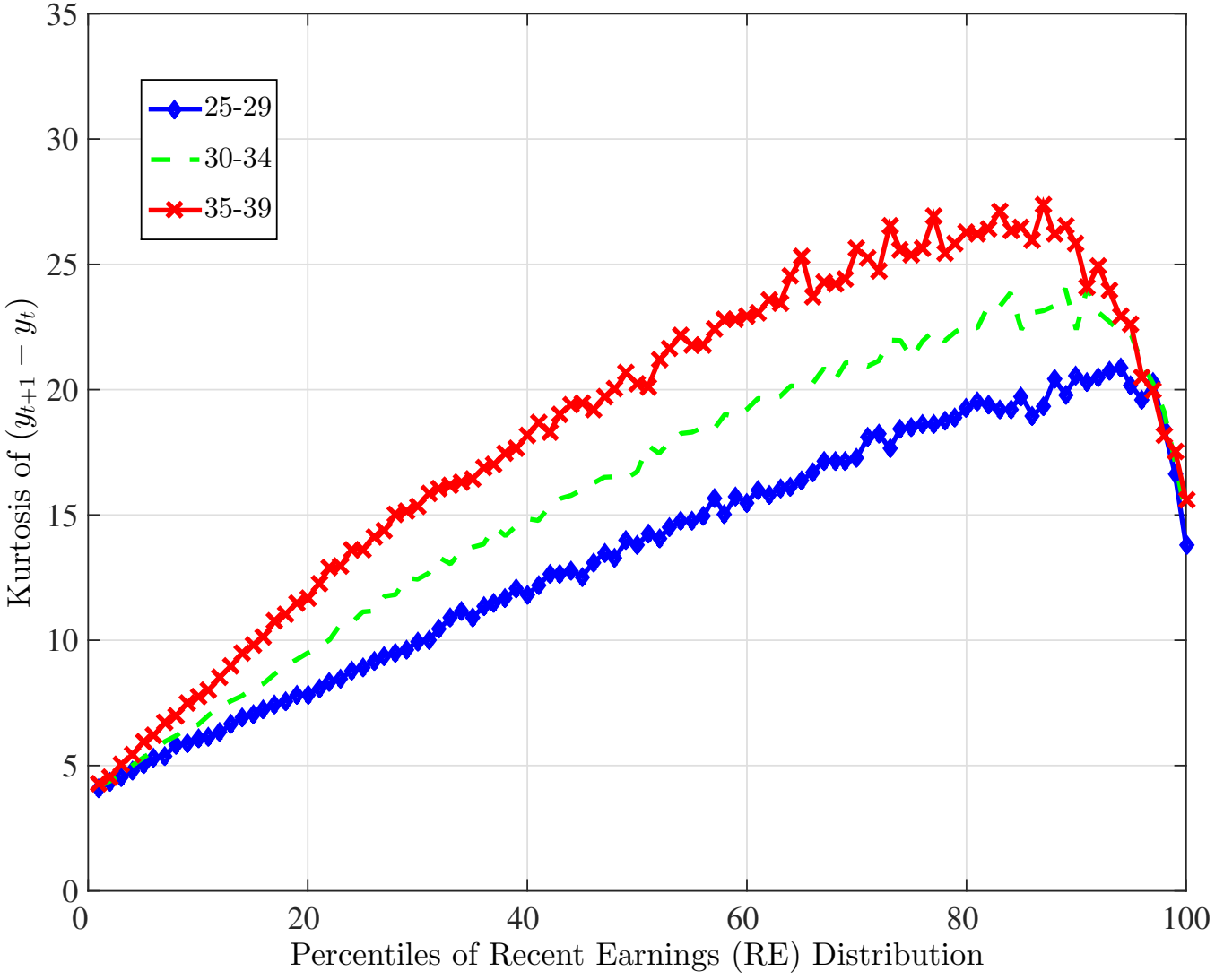
II.c Kurtosis of $y_{t+1} - y_t$



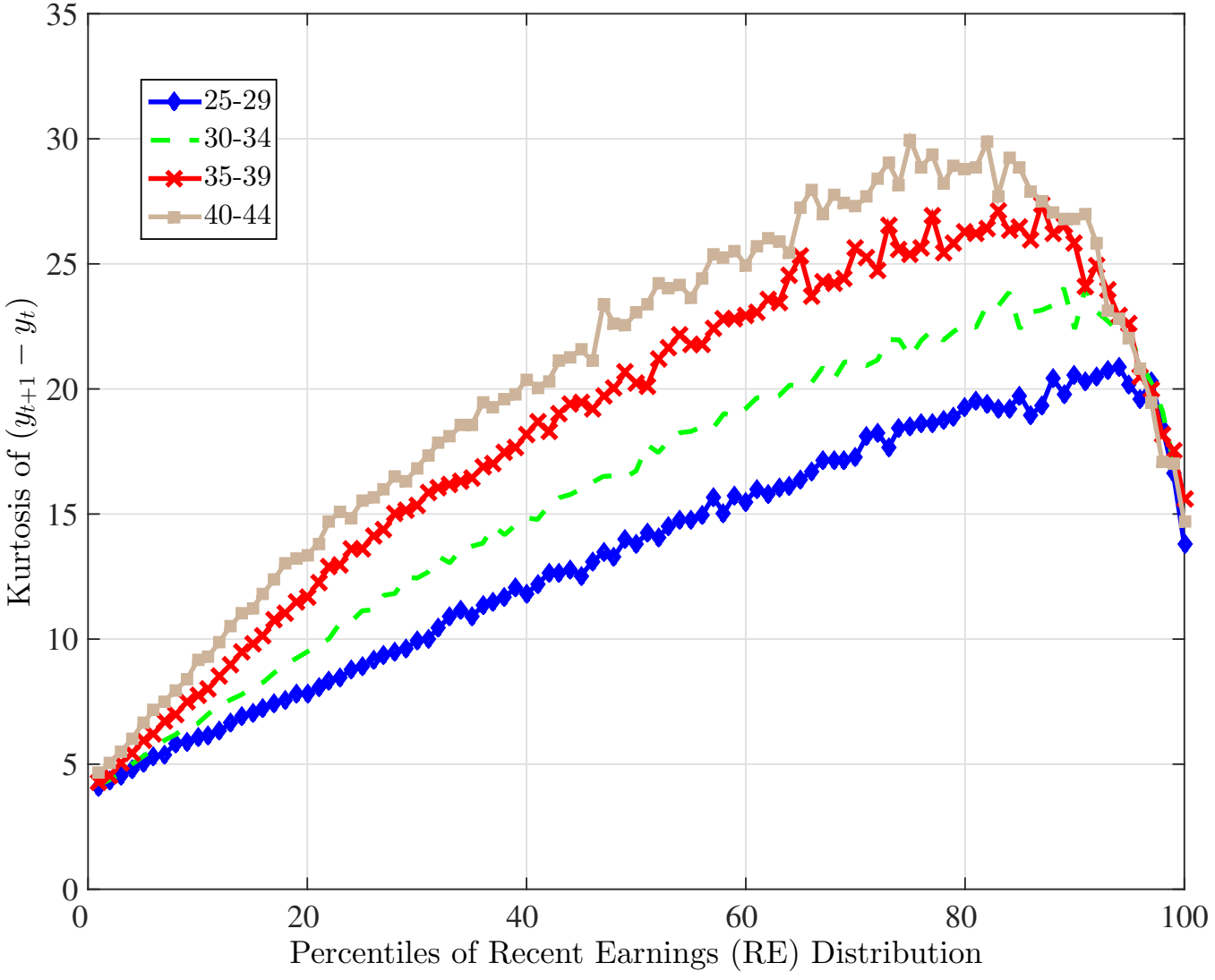
II.c Kurtosis of $y_{t+1} - y_t$



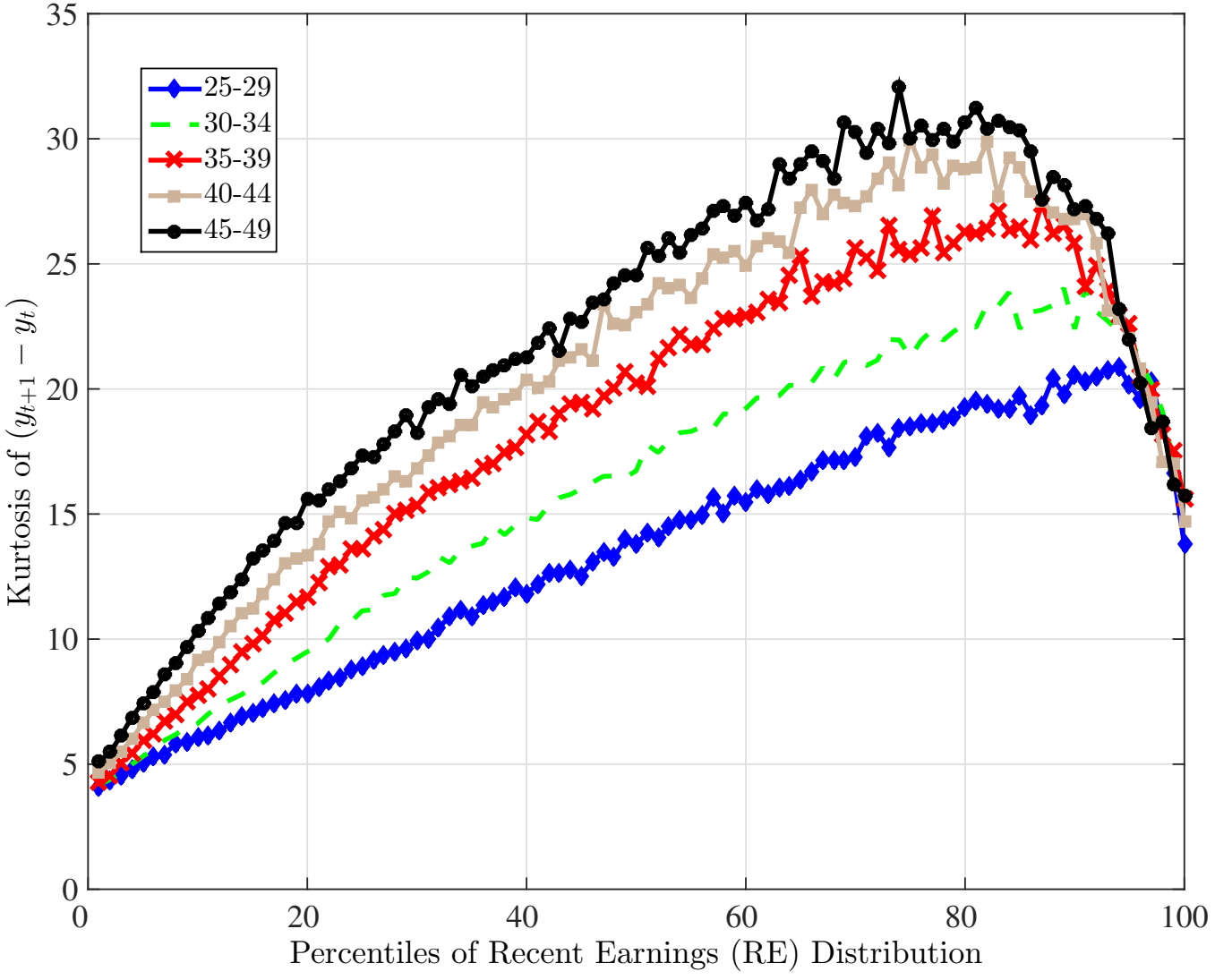
II.c Kurtosis of $y_{t+1} - y_t$



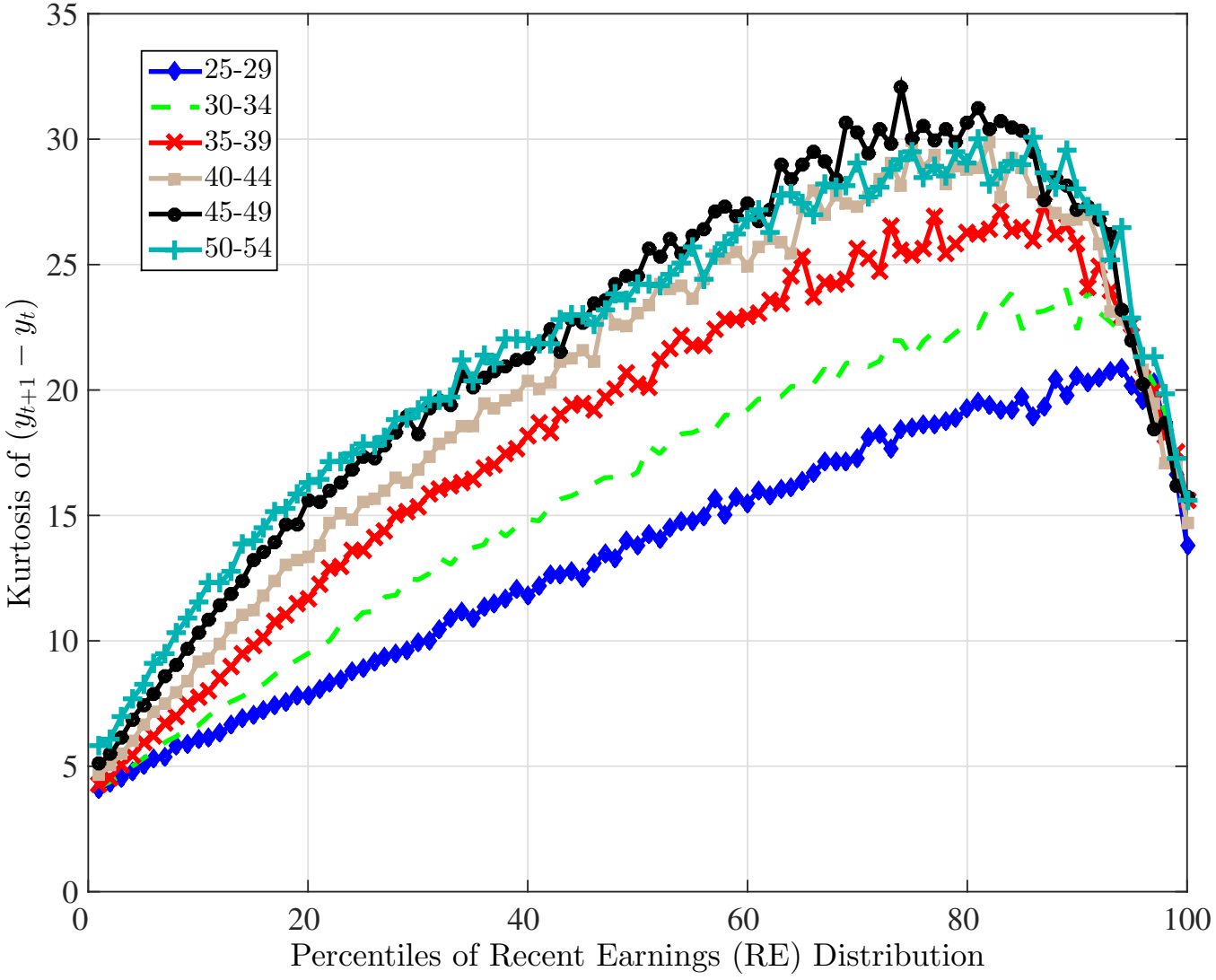
II.C Kurtosis of $y_{t+1} - y_t$



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



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



Risk Premium: Skewness and Kurtosis

Let $\tilde{\delta}$ be a static gamble. And π is the risk premium to avoid it:

$$U(c \times (1 - \pi)) = \mathbb{E}\left[U(c \times (1 + \tilde{\delta}))\right].$$

Risk Premium: Skewness and Kurtosis

Let $\tilde{\delta}$ be a static gamble. And π is the risk premium to avoid it:

$$U(c \times (1 - \pi)) = \mathbb{E}\left[U(c \times (1 + \tilde{\delta}))\right].$$

	Risk Premium (π)	
Gamble:	$\tilde{\delta}^A$	$\tilde{\delta}^B$
Mean	0.0	
Standard Deviation	0.10	
Skewness	0.0	
Excess Kurtosis	0.0	
Premium	.	

Risk Premium: Skewness and Kurtosis

Let $\tilde{\delta}$ be a static gamble. And π is the risk premium to avoid it:

$$U(c \times (1 - \pi)) = \mathbb{E}\left[U(c \times (1 + \tilde{\delta}))\right].$$

	Risk Premium (π)	
Gamble:	$\tilde{\delta}^A$	$\tilde{\delta}^B$
Mean	0.0	
Standard Deviation	0.10	
Skewness	0.0	
Excess Kurtosis	0.0	
Premium	**4.88%**	

Risk Premium: Skewness and Kurtosis

Let $\tilde{\delta}$ be a static gamble. And π is the risk premium to avoid it:

$$U(c \times (1 - \pi)) = \mathbb{E}\left[U(c \times (1 + \tilde{\delta}))\right].$$

	Risk Premium (π)	
Gamble:	$\tilde{\delta}^A$	$\tilde{\delta}^B$
Mean	0.0	0.0
Standard Deviation	0.10	0.10
Skewness	0.0	.
Excess Kurtosis	0.0	.
Premium	**4.88%**	.

Risk Premium: Skewness and Kurtosis

Let $\tilde{\delta}$ be a static gamble. And π is the risk premium to avoid it:

$$U(c \times (1 - \pi)) = \mathbb{E}\left[U(c \times (1 + \tilde{\delta}))\right].$$

	Risk Premium (π)	
Gamble:	$\tilde{\delta}^A$	$\tilde{\delta}^B$
Mean	0.0	0.0
Standard Deviation	0.10	0.10
Skewness	0.0	–2.0
Excess Kurtosis	0.0	27.0
Premium	4.88%	.

Risk Premium: Skewness and Kurtosis

Let $\tilde{\delta}$ be a static gamble. And π is the risk premium to avoid it:

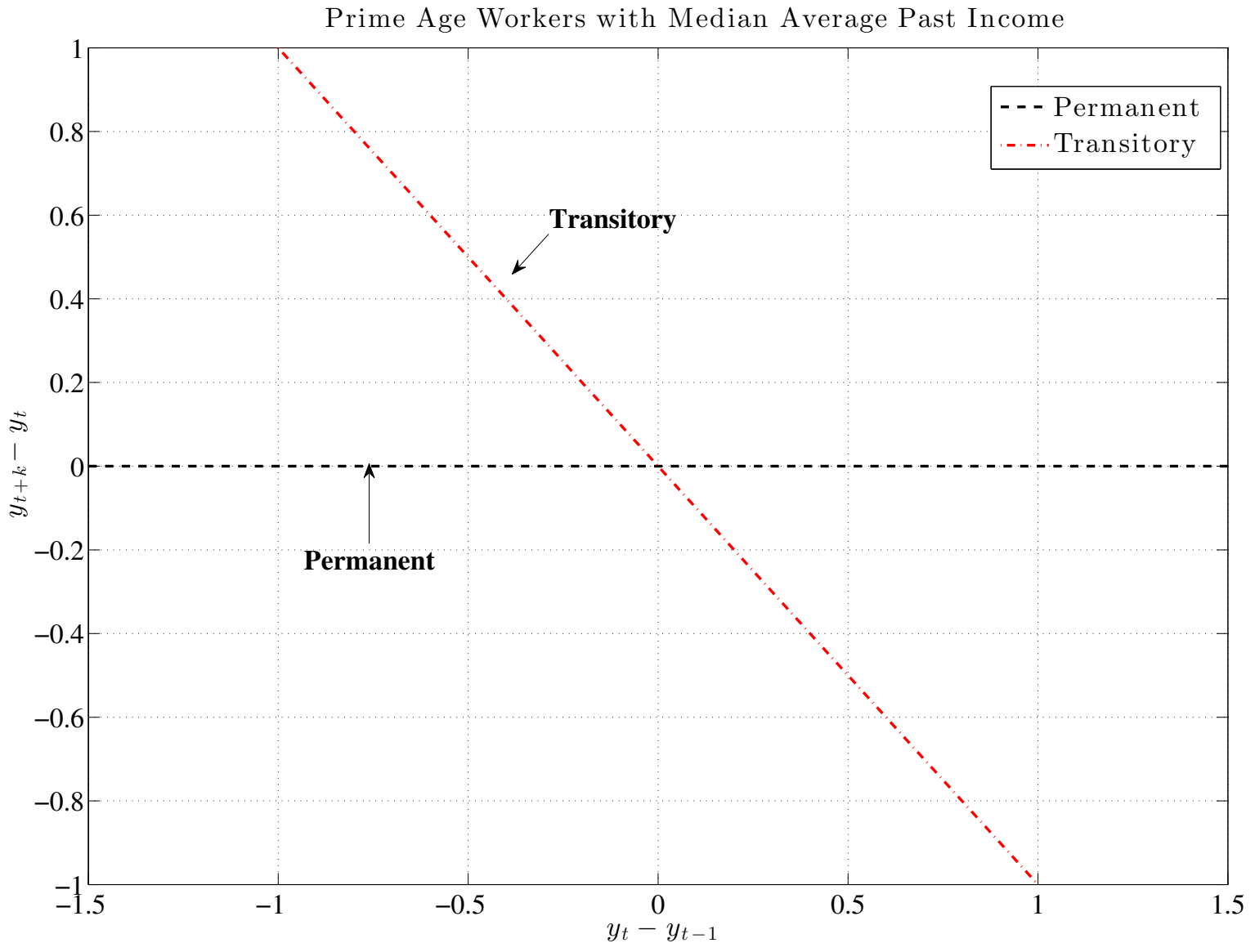
$$U(c \times (1 - \pi)) = \mathbb{E}\left[U(c \times (1 + \tilde{\delta}))\right].$$

	Risk Premium (π)	
Gamble:	$\tilde{\delta}^A$	$\tilde{\delta}^B$
Mean	0.0	0.0
Standard Deviation	0.10	0.10
Skewness	0.0	–2.0
Excess Kurtosis	0.0	27.0
Premium	4.88%	**22.15%**

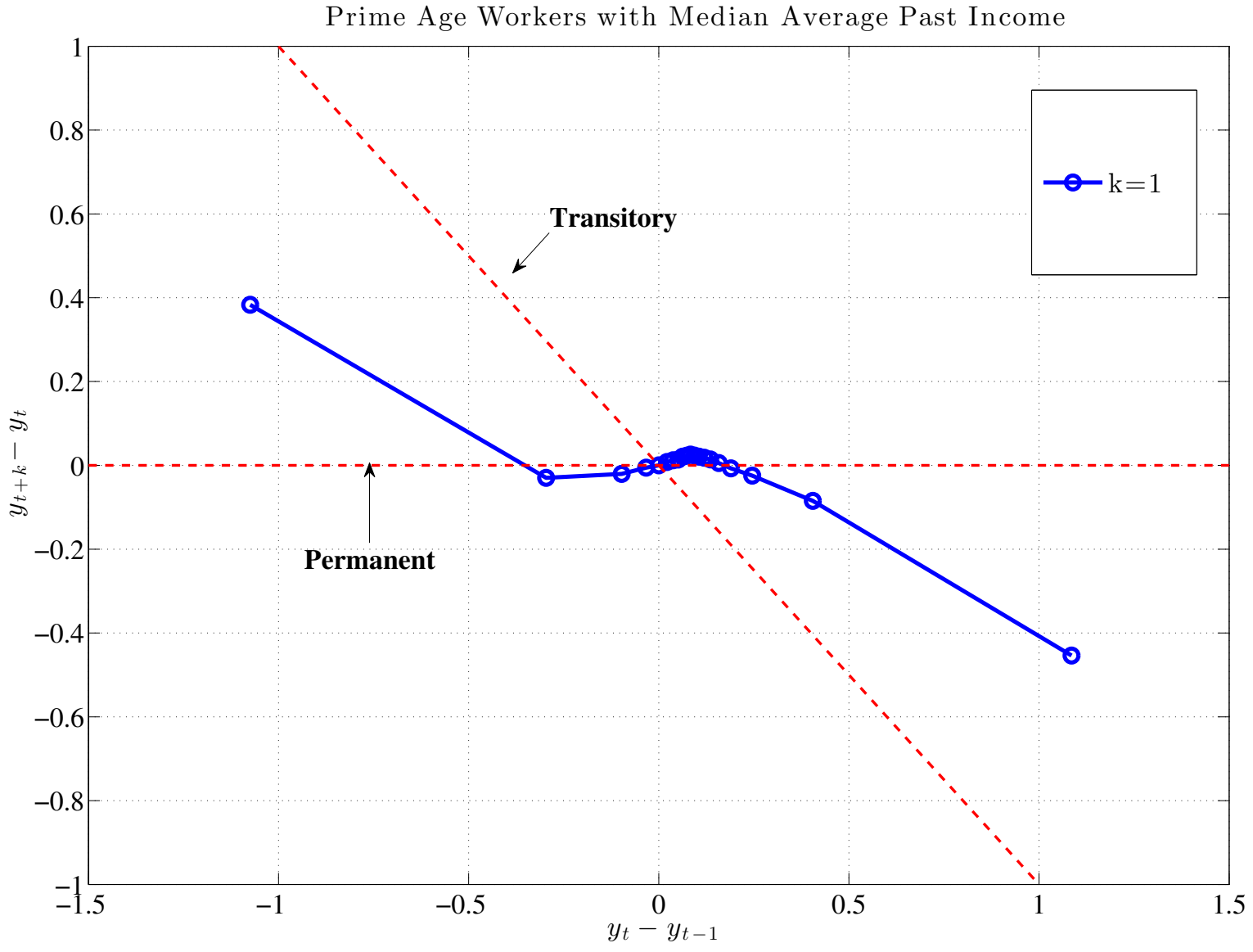
Four Sets of Empirical Facts

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.

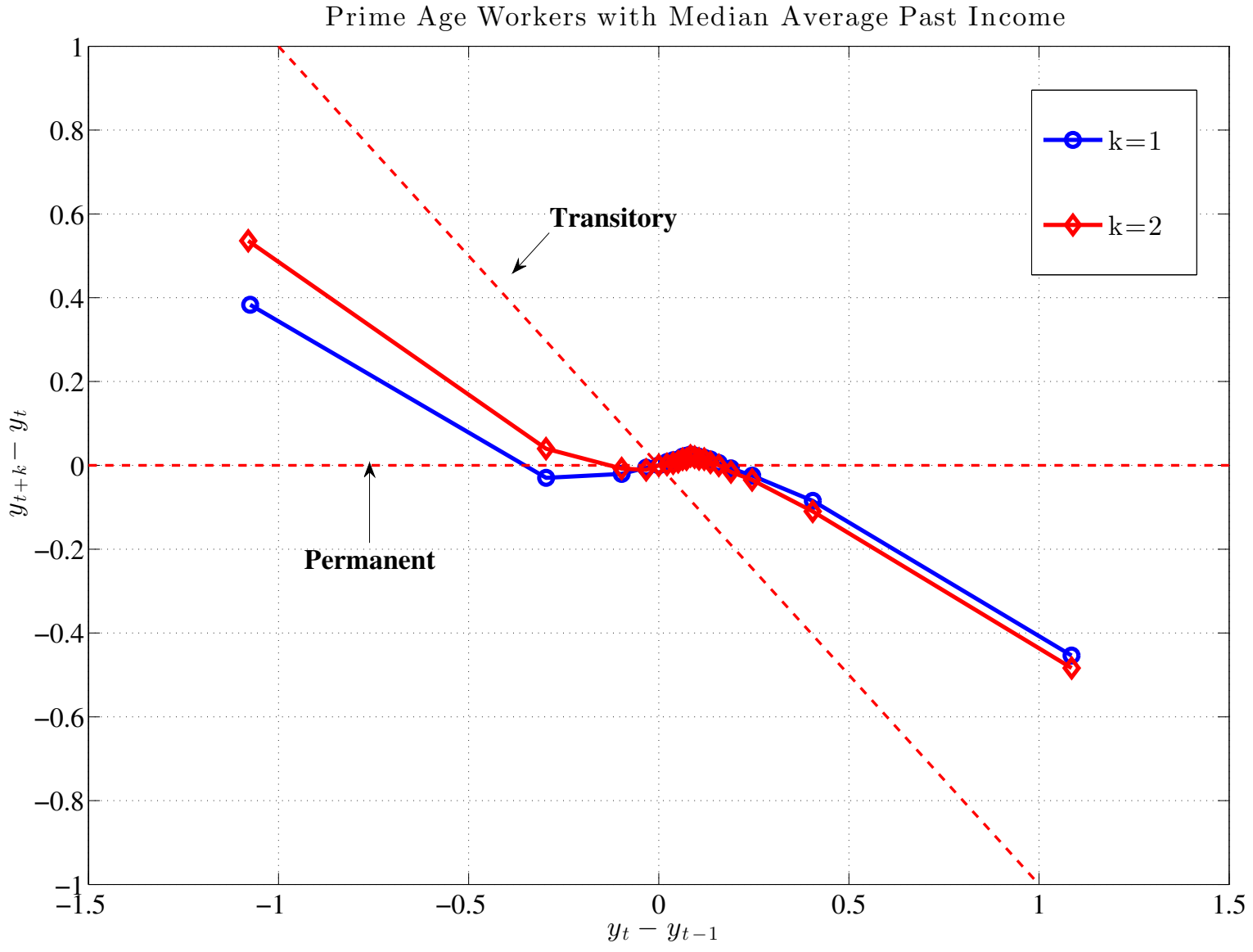
Impulse Response Functions



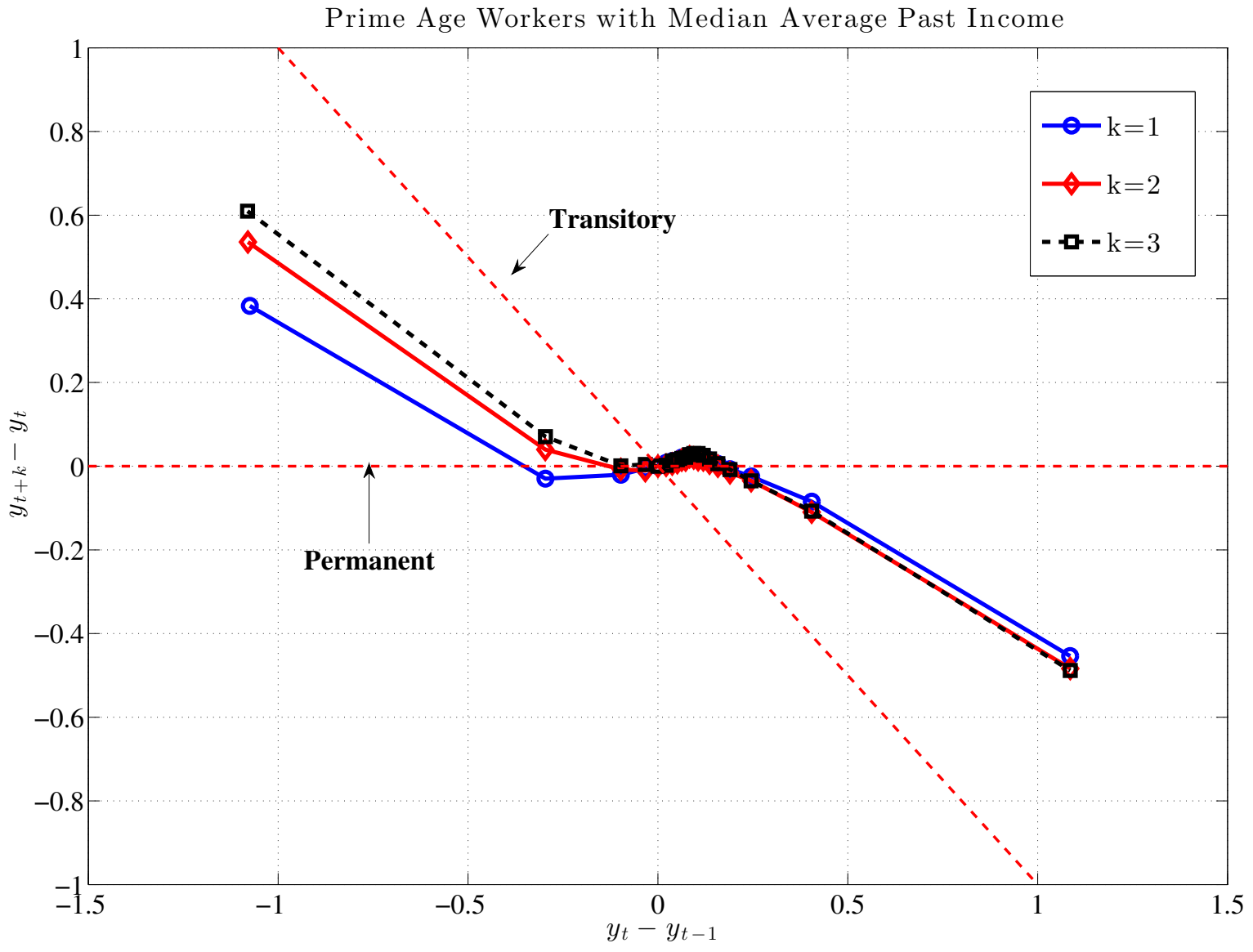
Impulse Response Functions



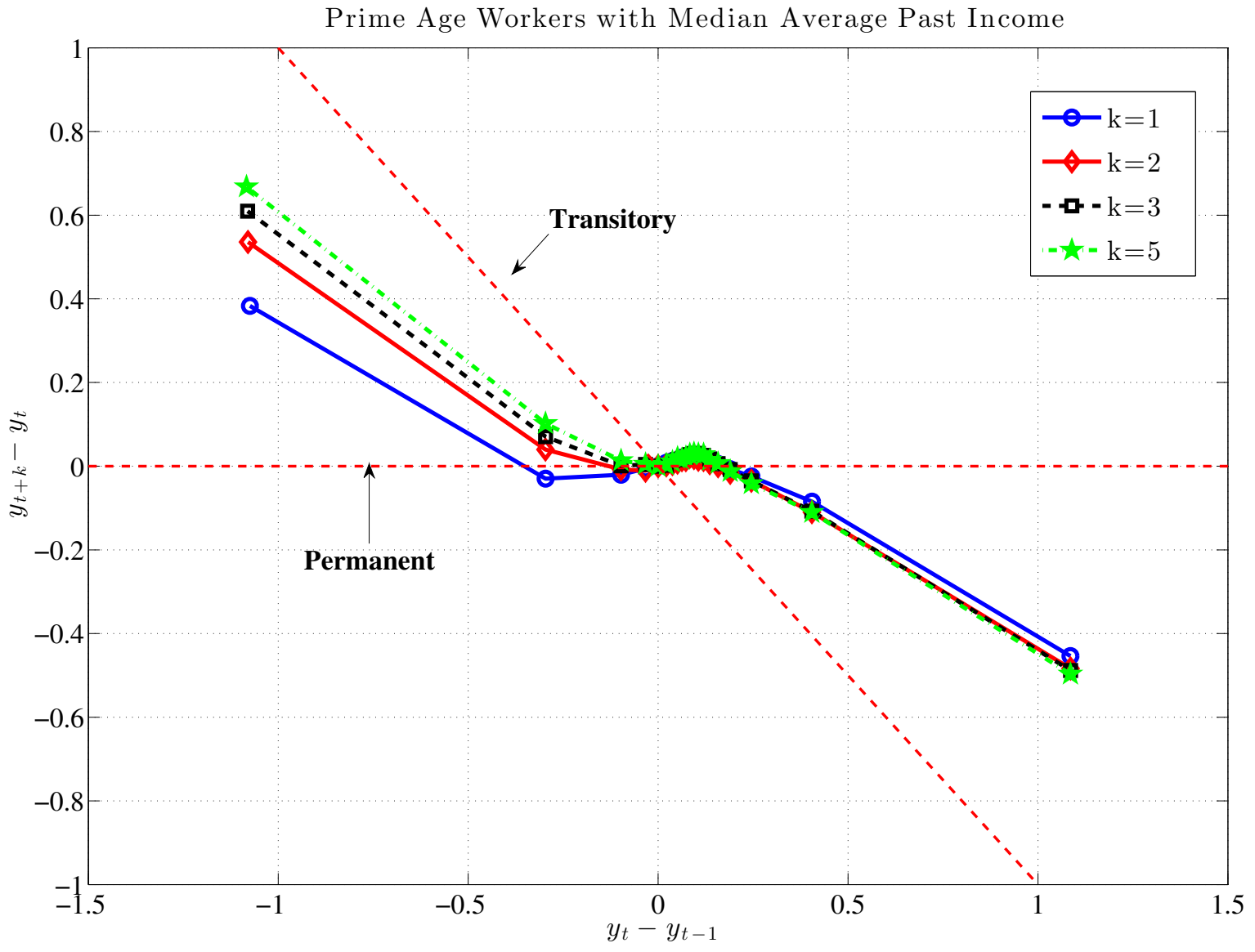
Impulse Response Functions



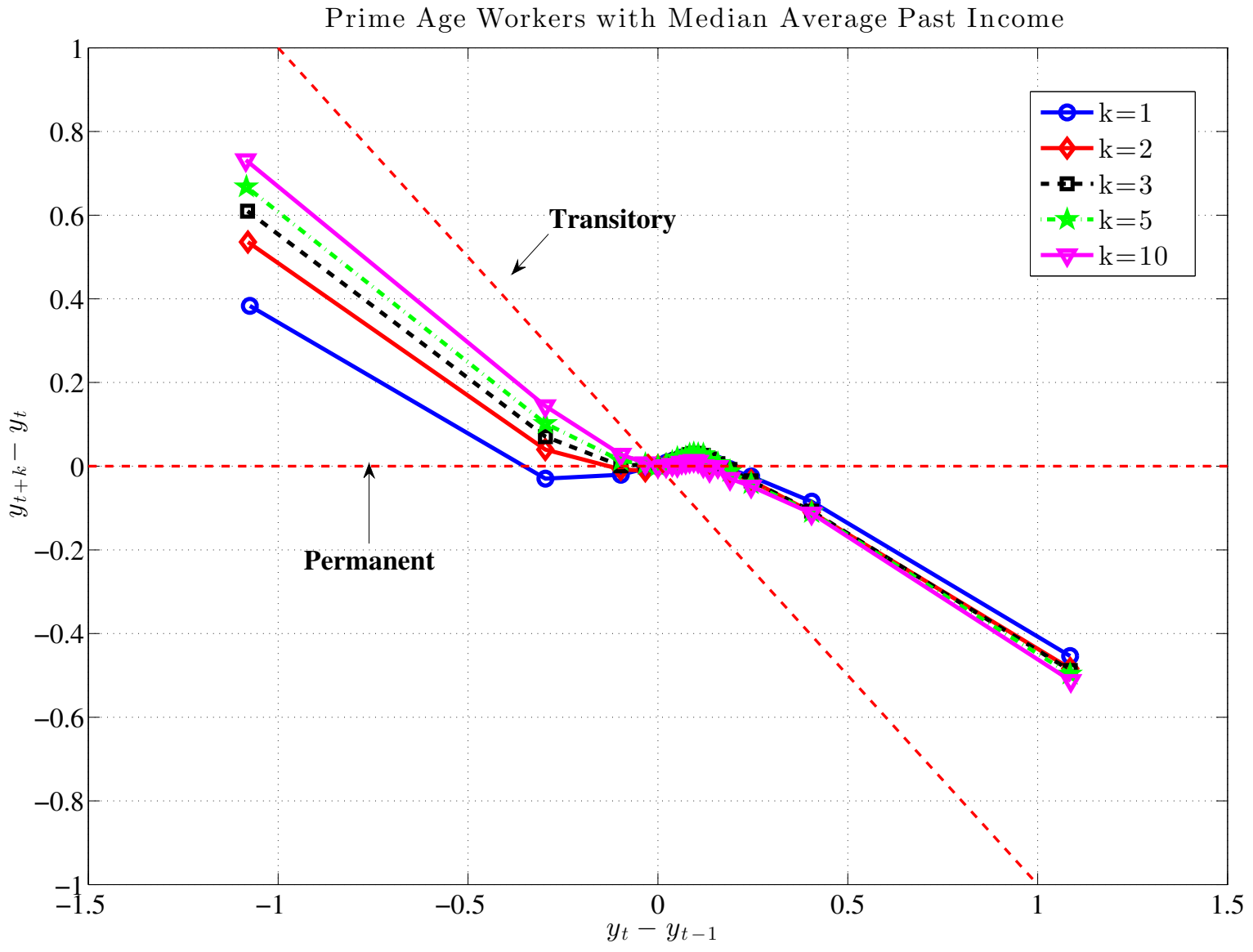
Impulse Response Functions



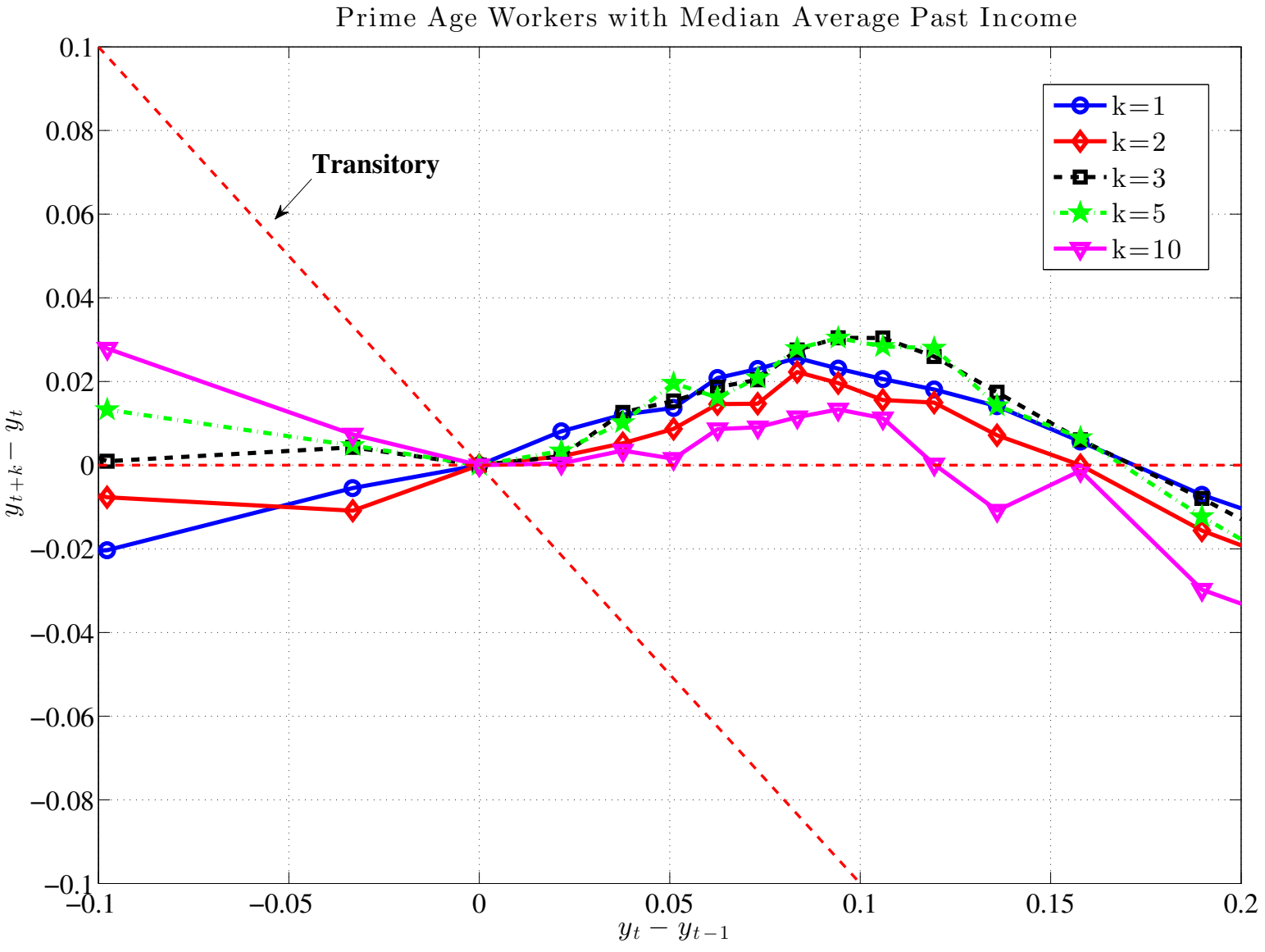
Impulse Response Functions



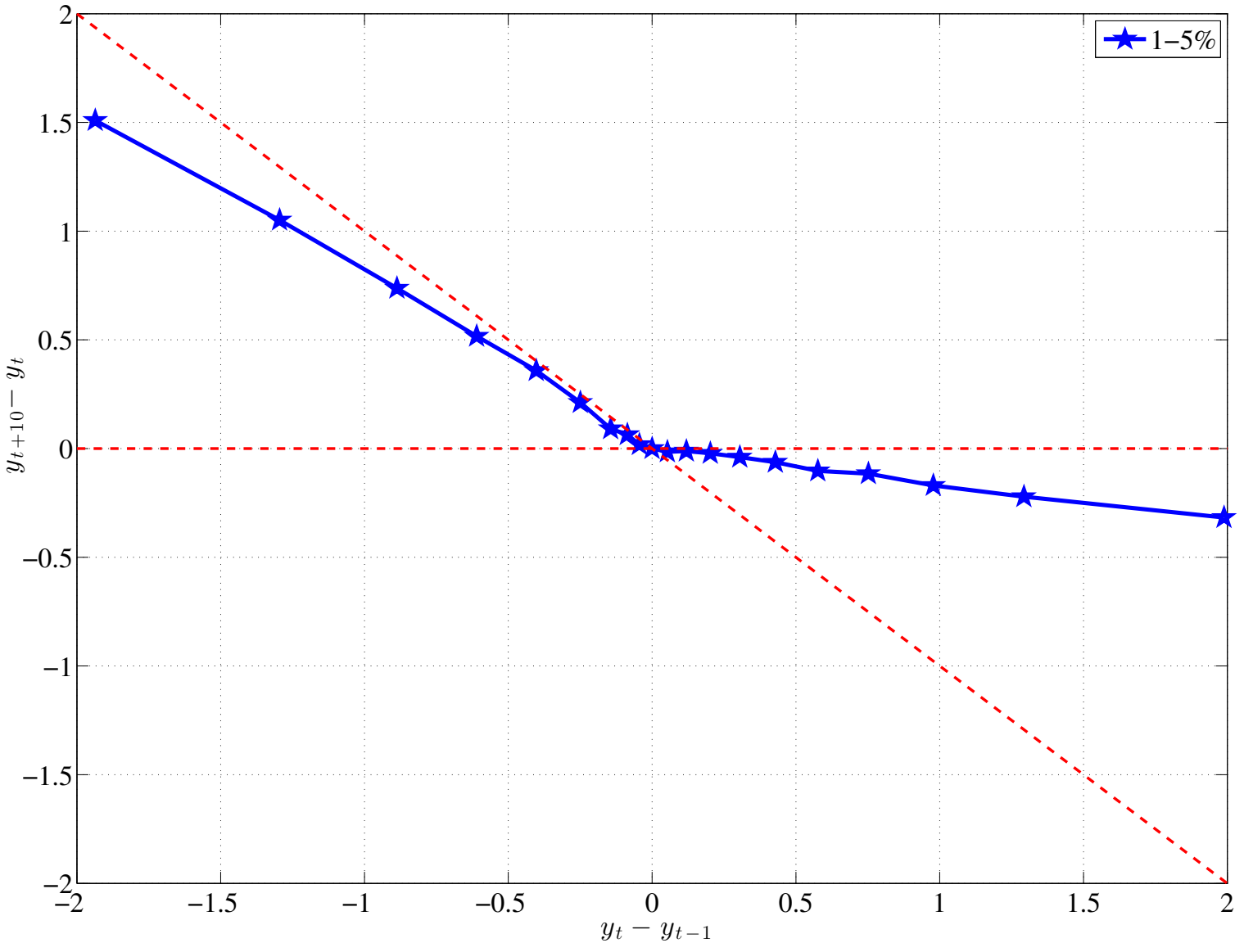
Impulse Response Functions



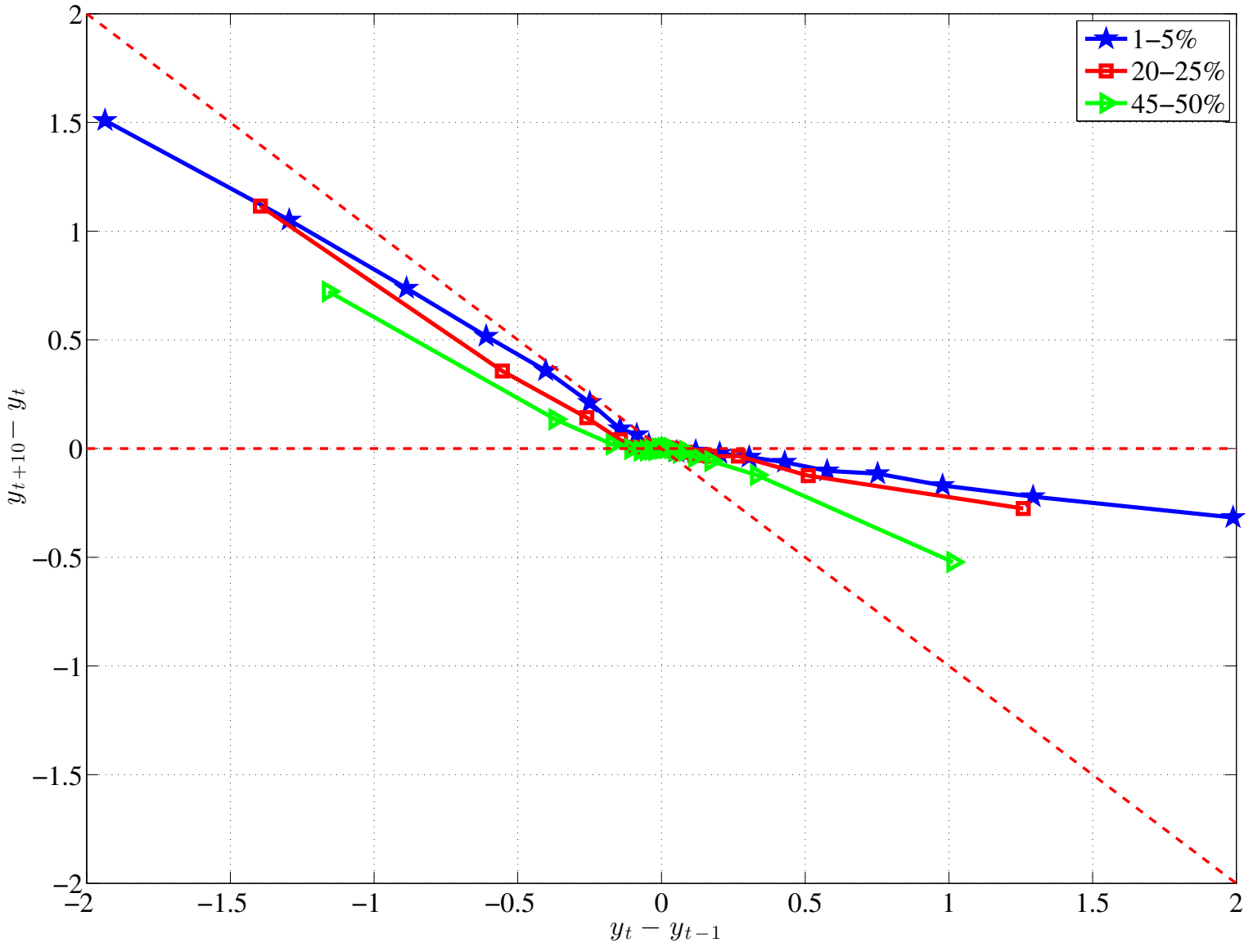
Impulse Response Functions



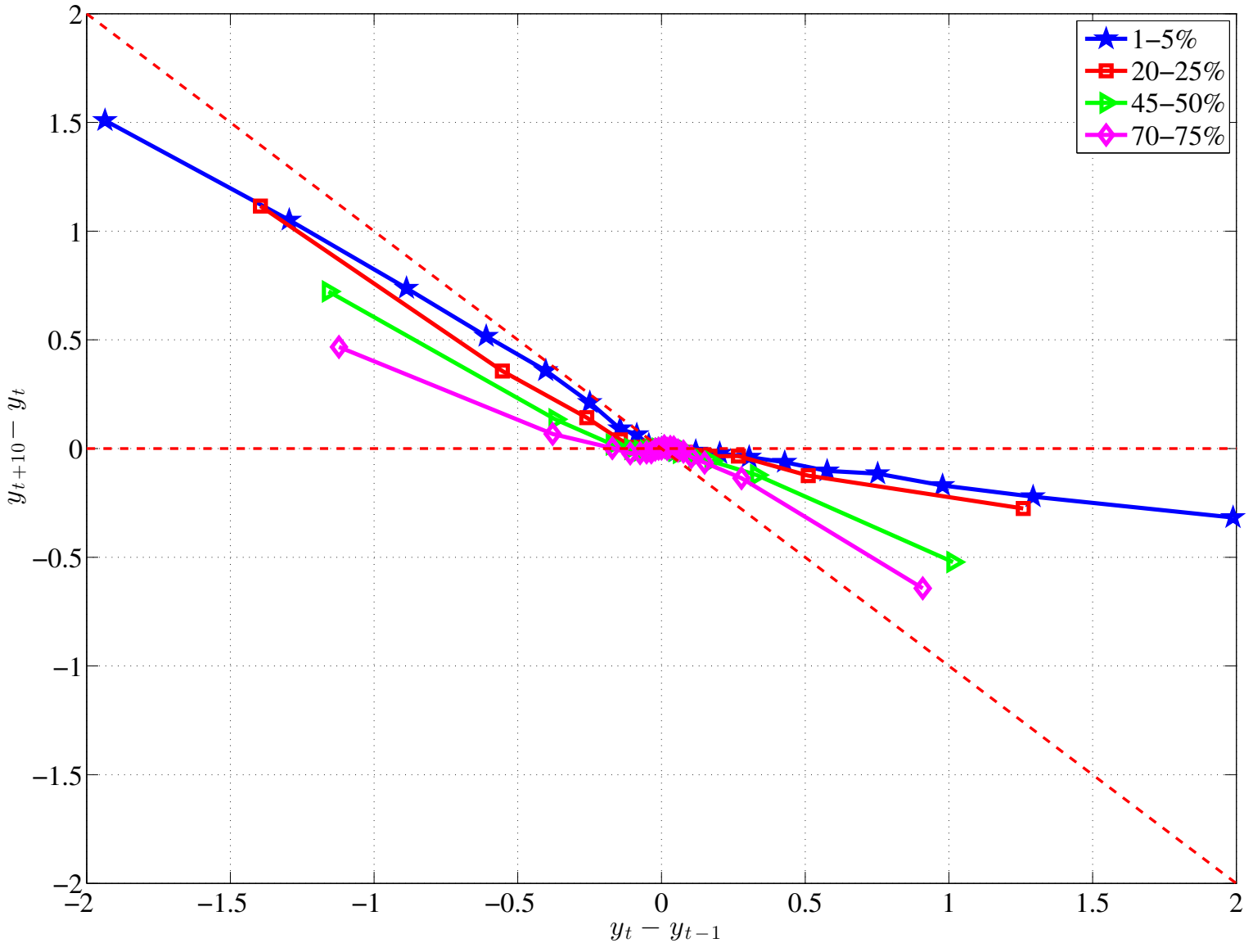
Asymmetric Mean Reversion



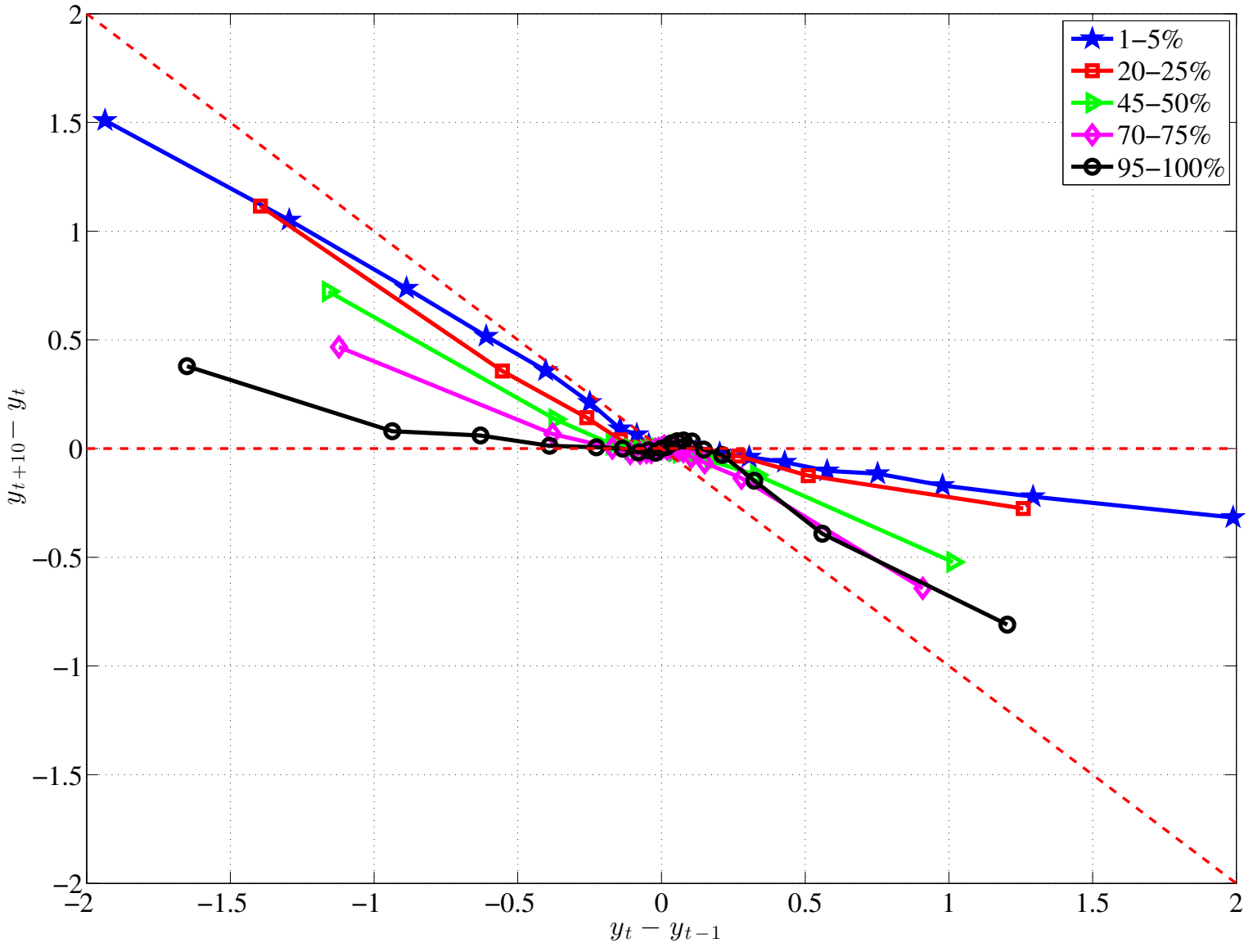
Asymmetric Mean Reversion



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

ESTIMATION

Econometric Specification

$$y_t^i = \underbrace{\left[\alpha^i + \beta^i t + \gamma^i t^2\right]}_{\text{HIP}} + \underbrace{z_{1,t}^i + z_{2,t}^i + z_{3,t}^i}_{\text{mixture of AR(1)s}} + \underbrace{\varepsilon_t^i}_{\text{i.i.d.}}$$

For $j = 1, 2, 3$:

$$z_{j,t}^i = \rho_j z_{j,t-1}^i + \eta_{j,t}^i$$

Econometric Specification

$$y_t^i = \underbrace{\left[\alpha^i + \beta^i t + \gamma^i t^2\right]}_{\text{HIP}} + \underbrace{z_{1,t}^i + z_{2,t}^i + z_{3,t}^i}_{\text{mixture of AR(1)s}} + \underbrace{\varepsilon_t^i}_{\text{i.i.d.}}$$

For $j = 1, 2, 3$:

$$z_{j,t}^i = \rho_j z_{j,t-1}^i + \eta_{j,t}^i$$

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

Econometric Specification

$$y_t^i = \underbrace{\left[\alpha^i + \beta^i t + \gamma^i t^2\right]}_{\text{HIP}} + \underbrace{z_{1,t}^i + z_{2,t}^i + z_{3,t}^i}_{\text{mixture of AR(1)s}} + \underbrace{\varepsilon_t^i}_{\text{i.i.d.}}$$

For $j = 1, 2, 3$:

$$z_{j,t}^i = \rho_j z_{j,t-1}^i + \eta_{j,t}^i$$

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

and

$$p_j(t, z_{t-1}) = a_j + b_j \times z_{t-1} + c_j \times t + d_j \times z_{t-1} \times t$$
$$1 = p_1 + p_2 + p_3$$

Estimation Results

Specification:	(2)	(4)	(8)
	Benchmark model	**Best fit**	**Standard model**
HIP order	1	1	0
AR(1)	2	3	1
RW	1	0	0
Heterog. variances?	yes	yes	no
Parameter			
σ_α	0.552	0.475	0.673
$\sigma_\beta \times 10$	0.130	0.137	—
$\text{corr}_{\alpha\beta}$	–0.49	–0.04	—
Objective value	**17.17**	**15.29**	**38.75**

Note: $^*\rho_\nu = 1.0$ *is imposed.*

Estimation Results

Specification:	(2)	(4)	(8)
	Benchmark model	**Best fit**	**Standard model**
ρ_z	0.259	0.083	1.00
ρ_x	0.425	0.512	—
ρ_ν	1.0*	0.796	—
$\bar{\sigma}_z$	0.847	0.656	0.172
$\bar{\sigma}_x$	0.361	0.337	—
$\bar{\sigma}_\nu$	0.087	0.082	—
σ_ε	0.029	0.019	0.040
$\sigma_{\nu 0}$	—	0.110	—
σ_{zz}	0.107	0.167	—
σ_{xx}	0.194	0.294	—

Selected Estimated Parameters

		z_1	z_2	z_2
p_j	Probability	0.11	0.23	0.68
ρ	Persistence	0.08	0.51	0.80
σ	Innov. std. dev	0.66	0.34	0.08
σ_α		0.48		
$\sigma_\beta \times 10$		0.14		

What to Use in Calibration?

- These estimated processes are complex and richly parameterized.
    - How to use them for calibration?

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.

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.
- Catalan-Civale-Fazilet (2015) 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.

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

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.

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

- New empirical facts regarding individual earnings.

Conclusions

- New empirical facts regarding individual earnings.
- Existing specifications do not capture these salient features of the data.

Conclusions

- New empirical facts regarding individual earnings.
- Existing specifications do not capture these salient features of the data.
- We propose a richer specification that captures many of these patterns.

Conclusions

- New empirical facts regarding 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 a well-known puzzle in the CME literature.
    - Ongoing research (Guvenen & Tonetti (2014)) shows this more rigorously.

Conclusions

- New empirical facts regarding 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 a well-known puzzle in the CME literature.
    - Ongoing research (Guvenen & Tonetti (2014)) shows this more rigorously.
- New benchmarks and targets for calibration.

Initial Conditions

$$\eta_{jt}^{*i} \sim \mathcal{N}(\mu_j, \sigma_j^i) \quad \text{and} \quad \eta_{jt}^i = \eta_{jt}^{*i} \times \mathbb{I}\{s_{i,t} \in I_{p_j}\} \tag{1}$$

$$\log\sigma_j^i \sim \mathcal{N}(\overline{\sigma}_j - \frac{\sigma_{jj}^2}{2}, \sigma_{jj}^2), j = z, x, \qquad \sigma_v^i \equiv \sigma_v \tag{2}$$