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

¹The 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?
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.
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.
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.
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

![Graph showing the age profile of labor income with a 127% rise from age 25 to 30.](image)

- **Log Average Income**: The graph indicates a log average income increase of 127% from age 25 to 30.
I. Income Growth Over Life Cycle

Income Growth from Pooled Regression
I. Income Growth Over Life Cycle

Income Growth from Pooled Regression

Top 1\%: 15-fold increase!

38\% increase (median worker)
I. Income Growth Over Life Cycle

- Top 1%: 15-fold increase!
- Random Walk Model
- Income Growth from Pooled Regression
- 38% increase (median worker)
I. Income Growth Over Life Cycle

Top 1%: 15-fold increase!

Income Growth from Pooled Regression
Random Walk Model

HIP (Guvenen (2009))
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 $y_{t+1} - y_t$

[Graph showing standard deviation of $y_{t+1} - y_t$ across percentiles of past 5-year average income distribution for prime age males.]
II.a Standard Deviation of $y_{t+1} - y_t$
II.a Standard Deviation of $y_{t+1} - y_t$
II.b Skewness of $y_{t+1} - y_t$

Age=25-34

Percentiles of Past 5-Year Average Income Distribution
II.b Skewness of $y_{t+1} - y_t$

Percentiles of Past 5-Year Average Income Distribution
II. B Skewness of $y_{t+1} - y_t$

Skewness of $(y_{t+1} - y_t)$

- Age=25-34
- Age=35-44
- Age=45-49

Percentiles of Past 5-Year Average Income Distribution
II. B Skewness of $y_{t+1} - y_t$
Kurtosis
II.c Histogram of $y_{t+1} - y_t$

Density

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

Density

US Data
Normal (0, 0.48)

Std. Dev. = 0.48
Skewness = -1.35
Kurtosis = 17.80
## II.c Distribution of Income Changes

\[
\text{Prob}(|y_{t+1} - y_t| < x)
\]

<table>
<thead>
<tr>
<th>$x$</th>
<th>Data</th>
<th>$N(0, 0.43^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.42</td>
<td>0.10</td>
</tr>
<tr>
<td>0.10</td>
<td>0.63</td>
<td>0.20</td>
</tr>
<tr>
<td>0.20</td>
<td>0.79</td>
<td>0.39</td>
</tr>
<tr>
<td>0.50</td>
<td>0.90</td>
<td>0.80</td>
</tr>
<tr>
<td>1.00</td>
<td>0.96</td>
<td>0.99</td>
</tr>
</tbody>
</table>
### II.c Distribution of Income Changes

<table>
<thead>
<tr>
<th>$x$</th>
<th>Data</th>
<th>$N(0, 0.43^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.42</td>
<td>0.10</td>
</tr>
<tr>
<td>0.10</td>
<td>0.63</td>
<td>0.20</td>
</tr>
<tr>
<td>0.20</td>
<td>0.79</td>
<td>0.39</td>
</tr>
<tr>
<td>0.50</td>
<td>0.90</td>
<td>0.80</td>
</tr>
<tr>
<td>1.00</td>
<td>0.96</td>
<td>0.99</td>
</tr>
</tbody>
</table>
II.c Kurtosis of $y_{t+1} - y_t$

Percentiles of Past 5-Year Average Income Distribution

Kurtosis of $(y_{t+1} - y_t)$

Ages 25-29
II.c Kurtosis of $y_{t+1} - y_t$

Percentiles of Past 5-Year Average Income Distribution

Kurtosis of $(y_{t+1} - y_t)$

- Ages 25-29
- Ages 30-34
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

Prime Age Workers with Median Average Past Income

\[ y_{t+k} - y_t \]

Permanent

Transitory
**Impulse Response Functions**

Prime Age Workers with Median Average Past Income

\[
y_{t+k} - y_t = y_t - y_{t-1}
\]

- Transitory
- Permanent

\(k=1\)
Prime Age Workers with Median Average Past Income

Impulse Response Functions

\[ y_t + k - y_t \]

- Transitory
- Permanent

-1.5 -1 -0.5 0 0.5 1 1.5

-1.5 -1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1

-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1
Prime Age Workers with Median Average Past Income

Impulse Response Functions

- $y_{t+k} - y_t$
- $y_t - y_{t-1}$

Permanent

Transitory

$k=1$

$k=2$

$k=3$
Prime Age Workers with Median Average Past Income

\[ y_{t+k} - y_t \]

Impulse Response Functions

Permanent

Transitory
**IMPULSE RESPONSE FUNCTIONS**

Prime Age Workers with Median Average Past Income

\[ y_{t+k} - y_t \]

- Permanent
- Transitory

\[ y_t - y_{t-1} \]
Prime Age Workers with Median Average Past Income

\[ y_{t+k} - y_t \]

Transitory
Asymmetric Mean Reversion
Asymmetric Mean Reversion

\[ y_{t+10} - y_t = 0.5(1 - 5\%)(y_t - y_{t-1}) + 0.5(20 - 25\%)(y_t - y_{t-1}) + 0.5(45 - 50\%)(y_t - y_{t-1}) \]
Asymmetric Mean Reversion

\[ y_{t+10} - y_t \]

\[ y_t - y_{t-1} \]
Asymmetric Mean Reversion

\[ y_{t+10} - y_t \]

\[ y_t - y_{t-1} \]

Legend:
- 1–5%
- 20–25%
- 45–50%
- 70–75%
- 95–100%
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

\[ y_{e,t+k} - \bar{y}_e \]

Average past income percentile

- \( k = 1 \)
- \( k = 2 \)
- \( k = 3 \)
- \( k = 5 \)
- \( k = 10 \)
Introduction

New Empirical Facts

Estimation

Conclusions

PRIME-AGE WORKERS: UNEMPLOYED

Average past income percentile

$y_{u,t+k} - \bar{y}_{u}$

$k=1$

$k=2$

$k=3$

$k=5$

$k=10$
**Prime-Age Workers: Diff. in Diff.**

\[
(y_{u,t+k} - \bar{y}_u) - (y_{e,t+k} - \bar{y}_e)
\]

- for \(k=1\)
- for \(k=2\)
- for \(k=3\)
- for \(k=5\)
- for \(k=10\)
**YOUNG WORKERS: DIFF. IN DIFF.**

- **Graph Description:**
  - X-axis: Average past income percentile
  - Y-axis: \((y_{u,t+k} - \bar{y}_u) - (y_{e,t+k} - \bar{y}_e)\)
  - Legend:
    - Blue diamond: \(k=1\)
    - Red dash line: \(k=2\)
    - Green circle: \(k=3\)
    - Black square: \(k=5\)
    - Cyan circle: \(k=10\)

- **Equation:**
  \[
  (y_{u,t+k} - \bar{y}_u) - (y_{e,t+k} - \bar{y}_e)
  \]
ESTIMATION
Econometric Specification

\[
y_t^i = \left[ \alpha^i + \beta^i t + \gamma^i t^2 \right] + \underbrace{z_{1,t}^i + z_{2,t}^i}_{\text{HIP}} + \varepsilon_t^i
\]

where for \( j = 1, 2 \):

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

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

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractions</td>
<td>0.10</td>
<td>0.80</td>
<td>0.10</td>
</tr>
<tr>
<td>mean(α)</td>
<td>2.21</td>
<td>2.95</td>
<td>3.57</td>
</tr>
<tr>
<td>mean(β) × 100</td>
<td>4.31</td>
<td>9.44</td>
<td>12.27</td>
</tr>
<tr>
<td>quadratic</td>
<td>-0.25</td>
<td>-0.25</td>
<td>-0.25</td>
</tr>
<tr>
<td>σ_α</td>
<td>-0.74</td>
<td>0.00</td>
<td>0.63</td>
</tr>
<tr>
<td>σ_β × 100</td>
<td>1.02</td>
<td>1.35</td>
<td>0.68</td>
</tr>
<tr>
<td>σ_α_β</td>
<td>-0.02</td>
<td>-0.41</td>
<td>0.21</td>
</tr>
<tr>
<td>ρ_1</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ_2</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ_1</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ_2</td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ_1</td>
<td>1.07 + 0.65zt_{-1} + 0.32t + 0.148tz_{t-1}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ_2</td>
<td>0.07 − 0.15zt_{-1} − 0.15t − 0.21tz_{t-1}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ_ε</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.