

Inequality Statistics from the LEHD

James R. Spletzer
U.S. Census Bureau

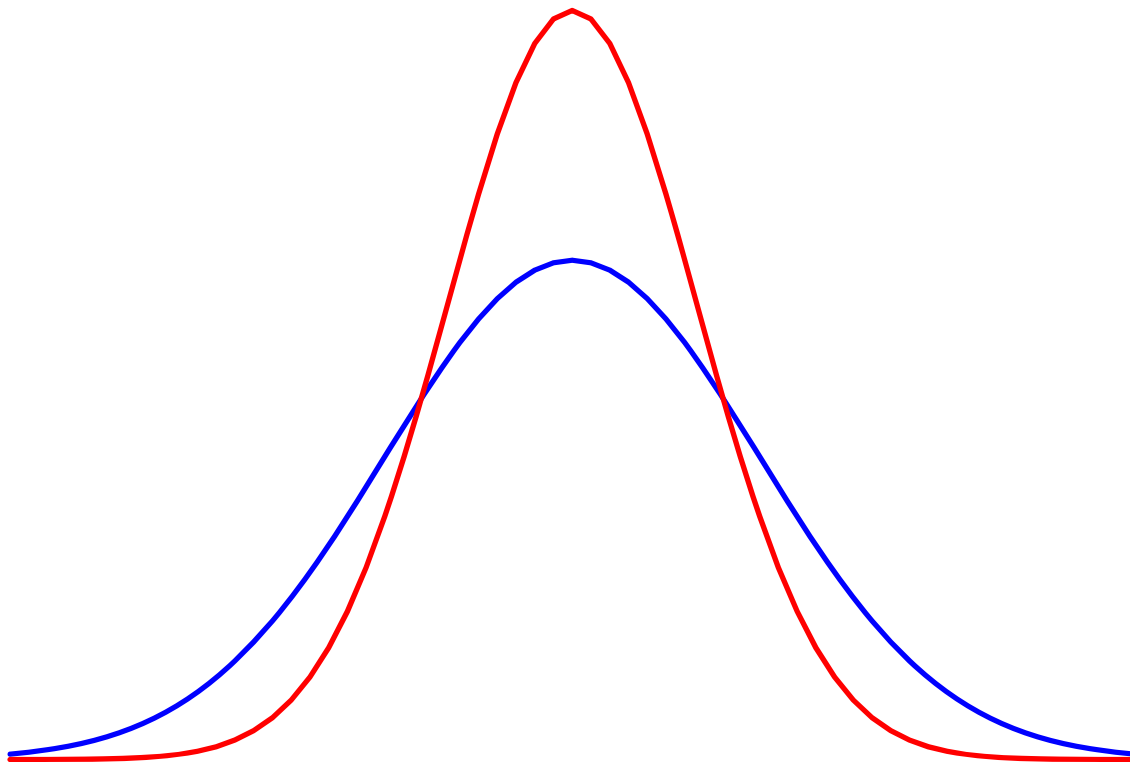
Federal Economic Statistics Advisory Committee
June 13, 2014

Disclaimer

Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau.

All results have been reviewed to ensure that no confidential information is disclosed.

Increasing Inequality

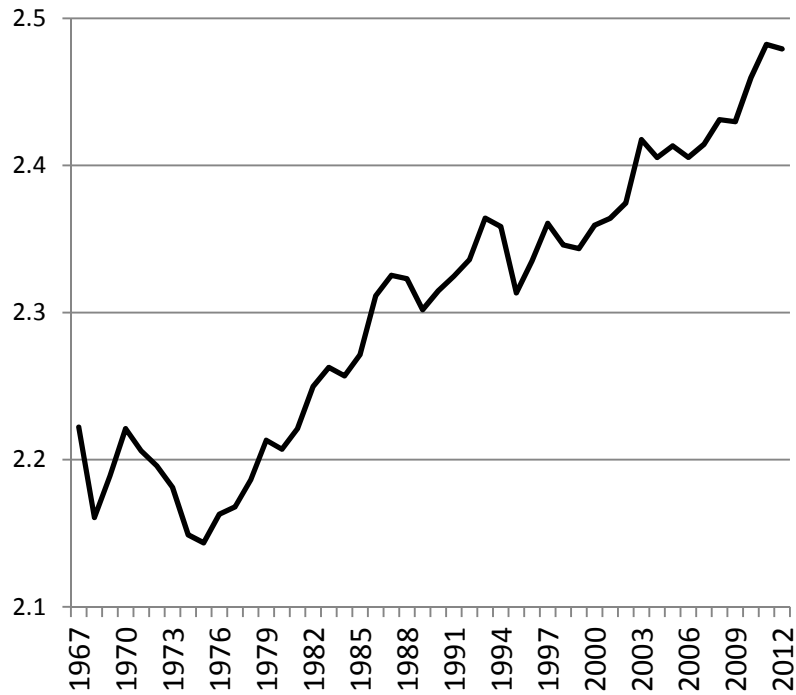


Increasing inequality refers to a widening distribution of income or earnings (from red to blue)

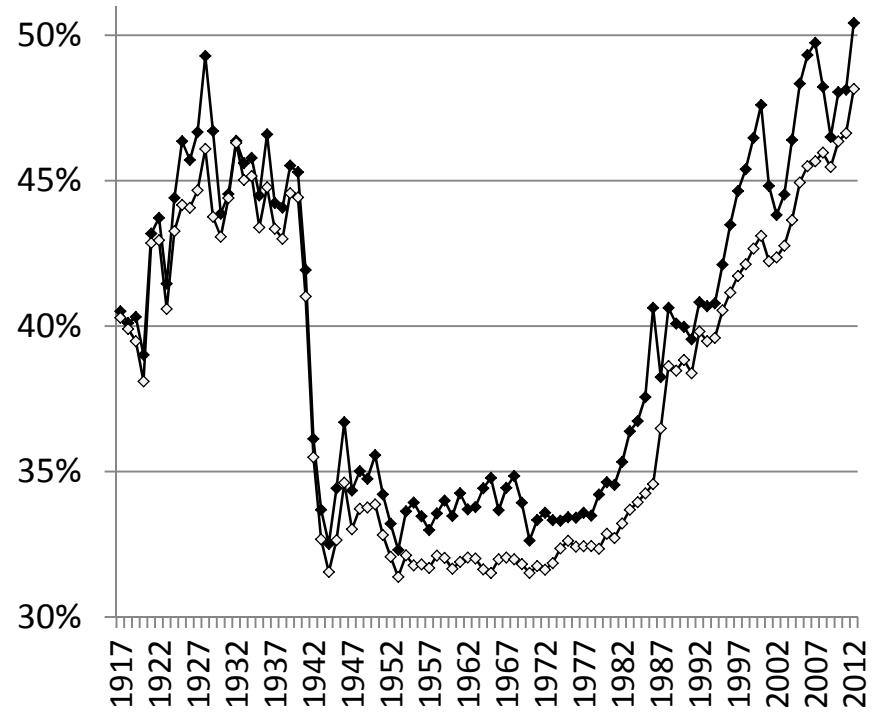
Increasing inequality is often measured using:

- the 90/10 ratio
- the amount of mass in the tails

Increasing Inequality



—ln(90)-ln(10) Published CPS ASEC



◆ Top 10% Income Share (including capital gains)
◇ Top 10% Income Share (excluding capital gains)

Source: Saez (2013)

Increasing Inequality

- ❖ The growth of inequality is a frequent topic in current policy discussions
 - Inequality and upward mobility has been referred to as “the defining challenge of our time” (President Obama, 12/4/13)
- ❖ Claim: almost everything we know empirically about inequality in the U.S. comes from CPS and IRS data
 - Published time series of 90-10, 90-50, and 50-10 from CPS
 - Published top percentile shares from IRS

Goals of this Presentation

- ❖ Two goals of this presentation:
 - A] Review publicly available statistics from BLS, Census, and the IRS that inform us about increasing inequality
 - B] Introduce another data source (LEHD) with time series information about increasing inequality

- ❖ What is the value added of LEHD statistics?
 - Comparison and confirmation
 - Utilize the large sample of the LEHD to provide inequality statistics by detailed demographic and job characteristics

Outline of Today's Presentation

- 1) Publicly available 90-10 statistics from CPS
 - a) Outgoing Rotation Group (ORG)
 - b) Annual Social and Economic Supplement (ASEC)

- 2) Publicly available top percentile shares from IRS, SSA, and CPS ASEC

- 3) New LEHD inequality statistics
 - a) 90-10, 90-50, and 50-10 statistics
 - b) Top percentile shares
 - c) New statistics by demographic & job characteristics

Published {10,50,90} Percentiles

CPS ORG

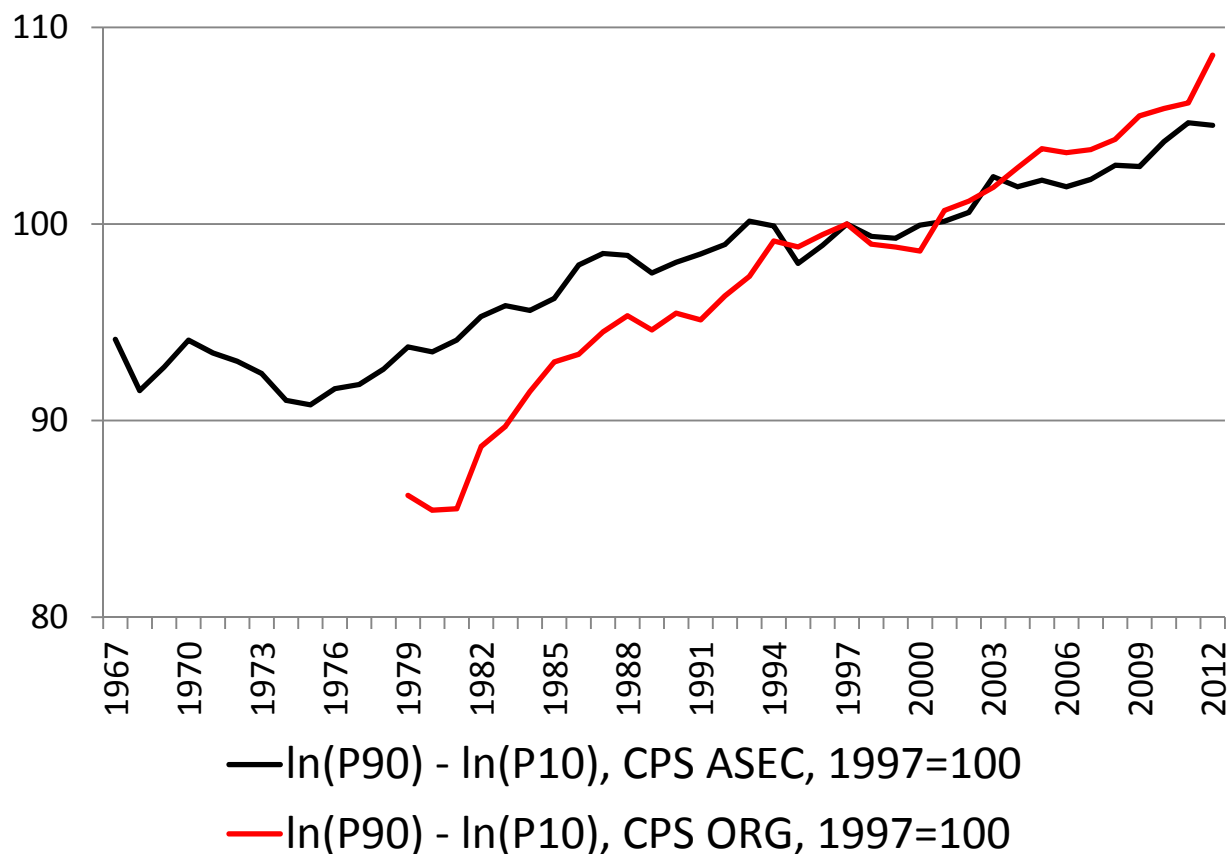
- Annual 1979 – current
Quarterly 1994 - current
- <http://www.bls.gov/webapps/legacy/cpswktab5.htm>
- Usual weekly earnings of full time wage & salary workers currently employed at their main job

CPS ASEC

- Annual 1967 – current
- <http://www.census.gov/prod/2013pubs/p60-245.pdf>
- Previous year income from all jobs for all persons currently residing in the household

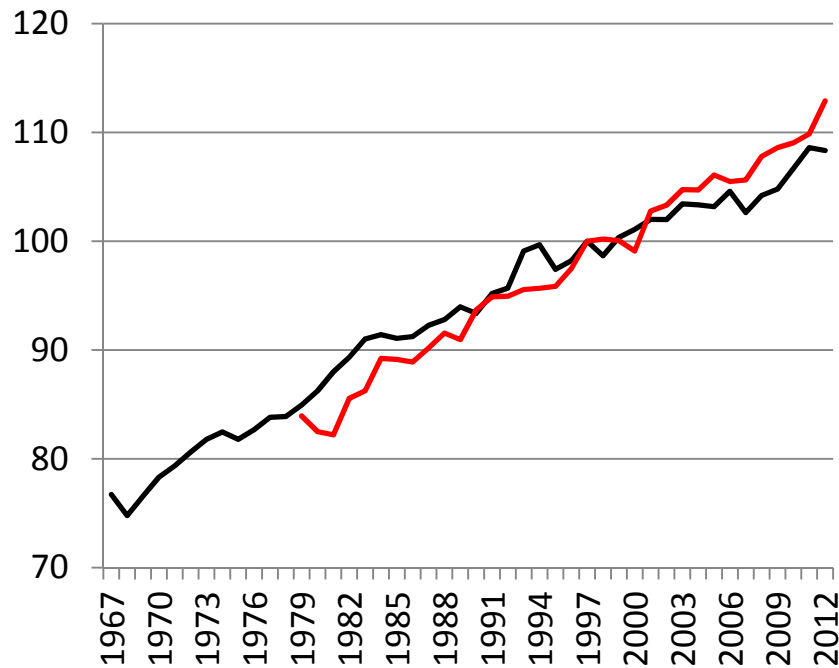
Increasing Inequality

CPS ORG and CPS ASEC



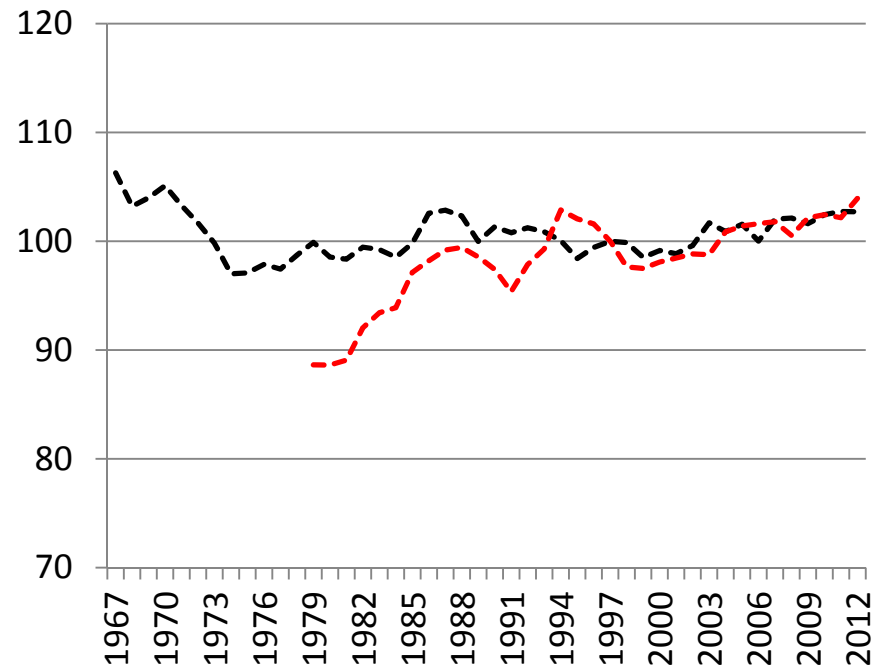
Increasing Inequality

CPS ORG and CPS ASEC



— ln(P90) - ln(P50), CPS ASEC, 1997=100

— ln(P90) - ln(P50), CPS ORG, 1997=100



- - - ln(P50) - ln(P10), CPS ASEC, 1997=100

- - - ln(P50) - ln(P10), CPS ORG, 1997=100

Published Top Percentile Shares

IRS (Saez, Table A2)

- Annual 1917 – current
- <http://elsa.berkeley.edu/~saez/TabFig2012prel.xls>
- (1) Annual income of tax units;
(2) Annual salaries and wages of tax units

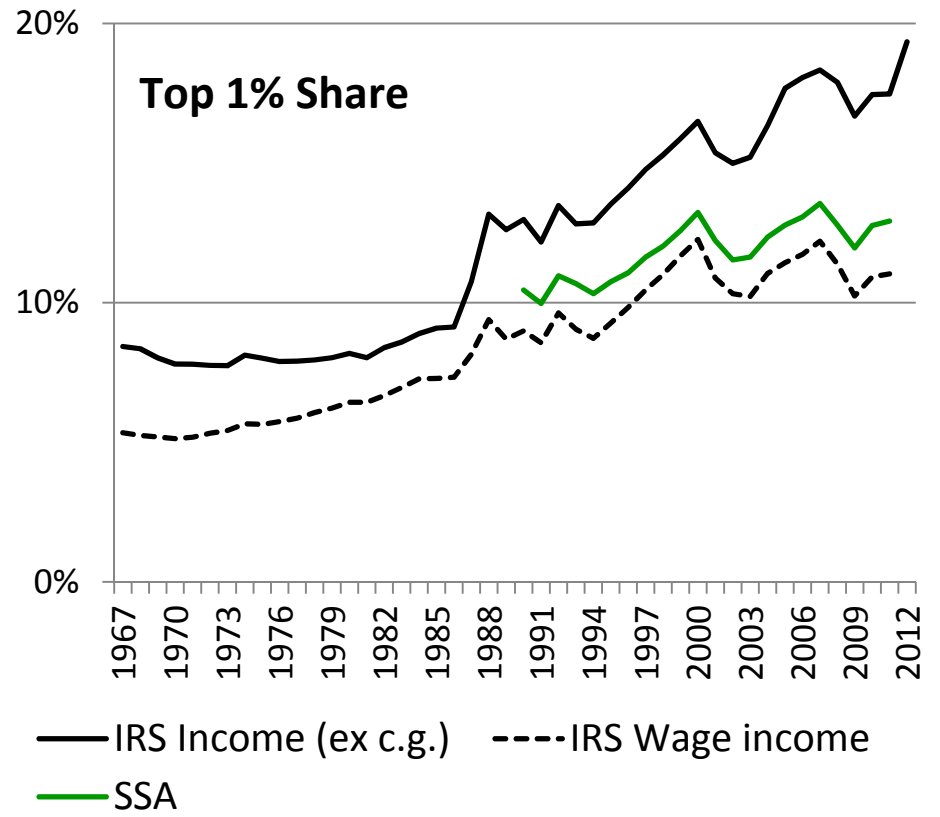
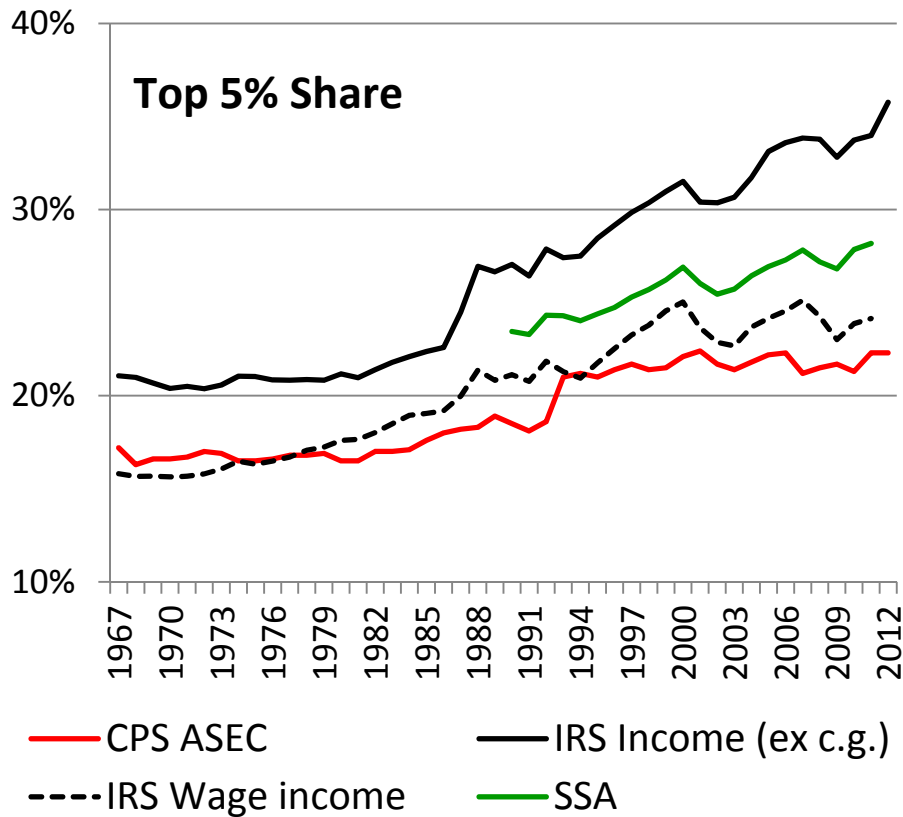
SSA (Saez, Table B5)

- Annual 1990 – current
- <http://elsa.berkeley.edu/~saez/TabFig2012prel.xls>
- Annual earnings of individuals

CPS ASEC

- Annual 1967 – current
- <http://www.census.gov/prod/2013pubs/p60-245.pdf>
- Annual income of households

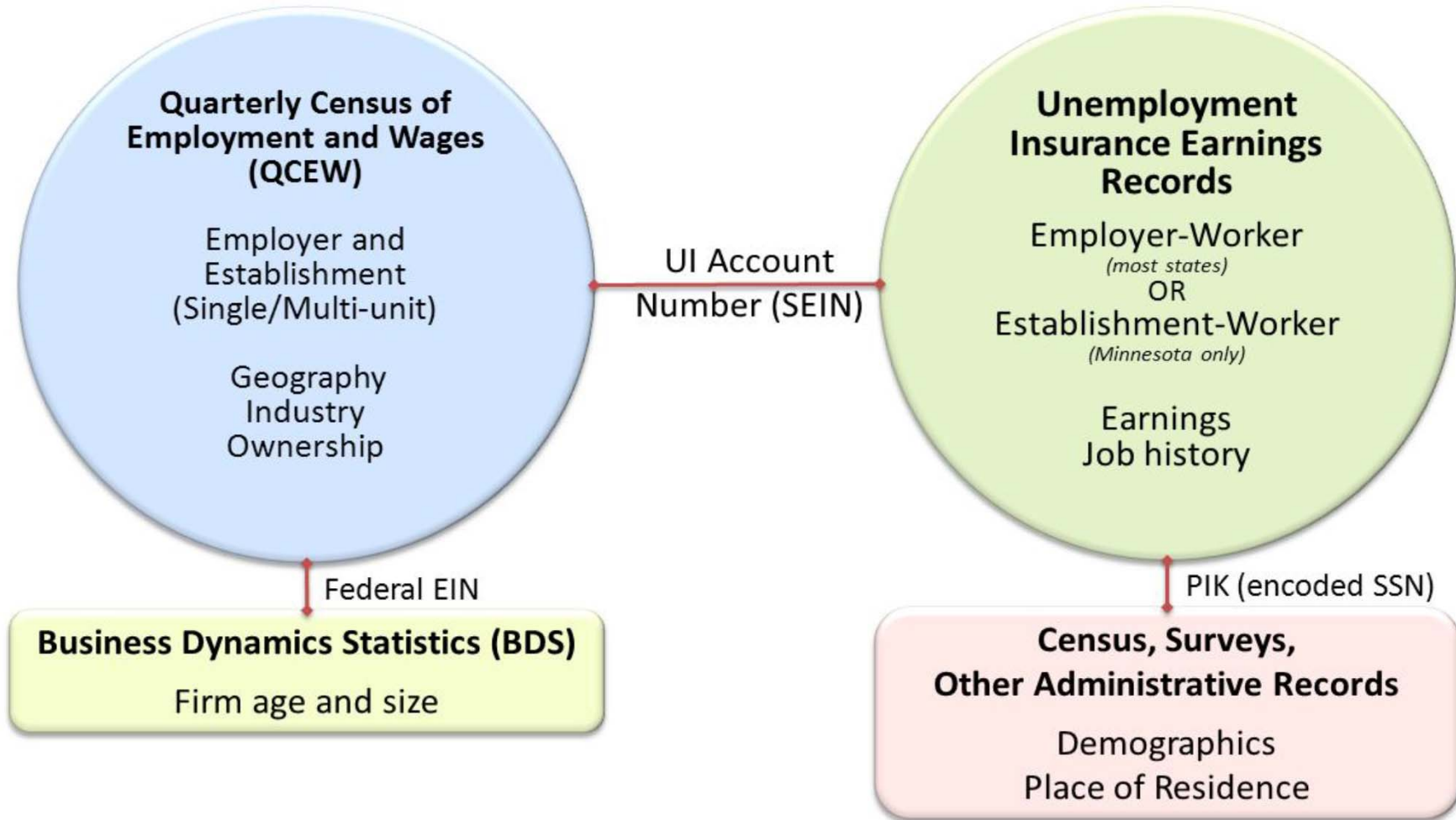
Increasing Inequality IRS, SSA, & CPS ASEC



LEHD Data

- ❖ Longitudinal Employer-Household Dynamics
 - Longitudinally linked employer-employee microdata
 - Created at the U.S. Census Bureau
 - Microdata from the State UI administrative systems wage records and QCEW establishment data
 - Enhanced with demographics (age, gender, ...)
 - Enhanced with firm information (age, size)
- ❖ Different states have joined the LEHD at different times, and have provided different amounts of historical data
 - This presentation: 20 states with data from 1996:Q2 to 2012:Q2
 - These 20 states account for 48% of national employment

LEHD Data

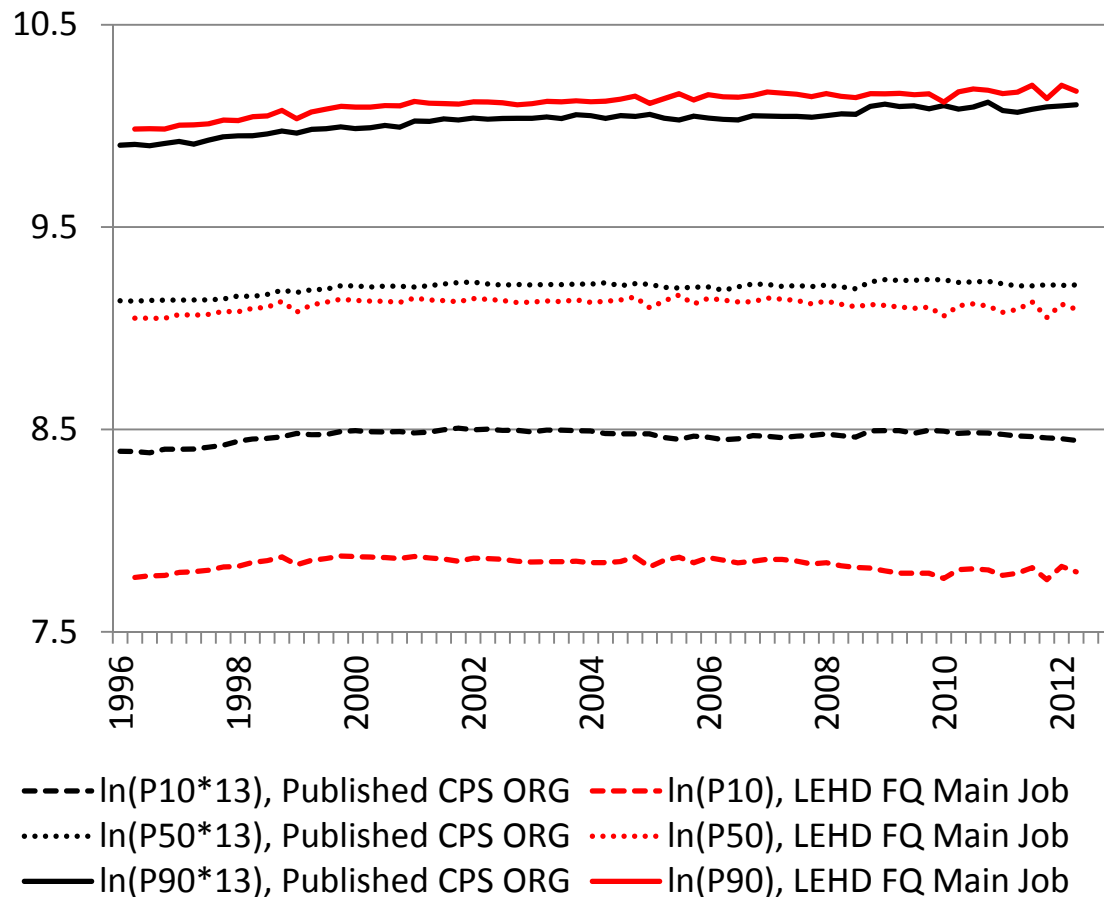


LEHD {10,50,90} Earnings Percentiles

- ❖ Attempting to mimic the CPS-ORG earnings data
 - Quarterly earnings of individuals in their main full quarter job
 - “Full quarter job” is defined as the middle quarter of 3 consecutive quarters at the same employer
 - 2.6 billion full-quarter jobs 1996:Q2 – 2012:Q2
 - approximately 39½ million FQ jobs each quarter (65 quarters)
- ❖ Compare to the CPS-ORG quarterly data
 - All statistics seasonally adjusted
 - All statistics converted to real (2012 CPI-U-RS) natural-logarithms

CPS-ORG (*13) and LEHD

{10,50,90} earnings percentiles, quarterly SA, real 2012 \$



Very similar 50th & 90th percentiles

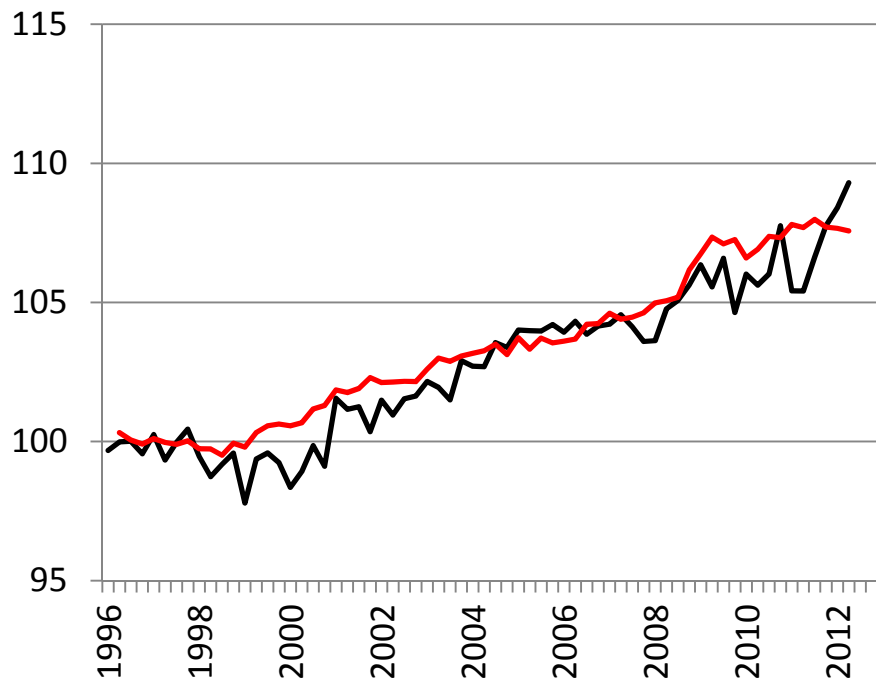
But different 10th percentiles:

-- perhaps “*13” is a poor method of transforming CPS weekly earnings to quarterly earnings

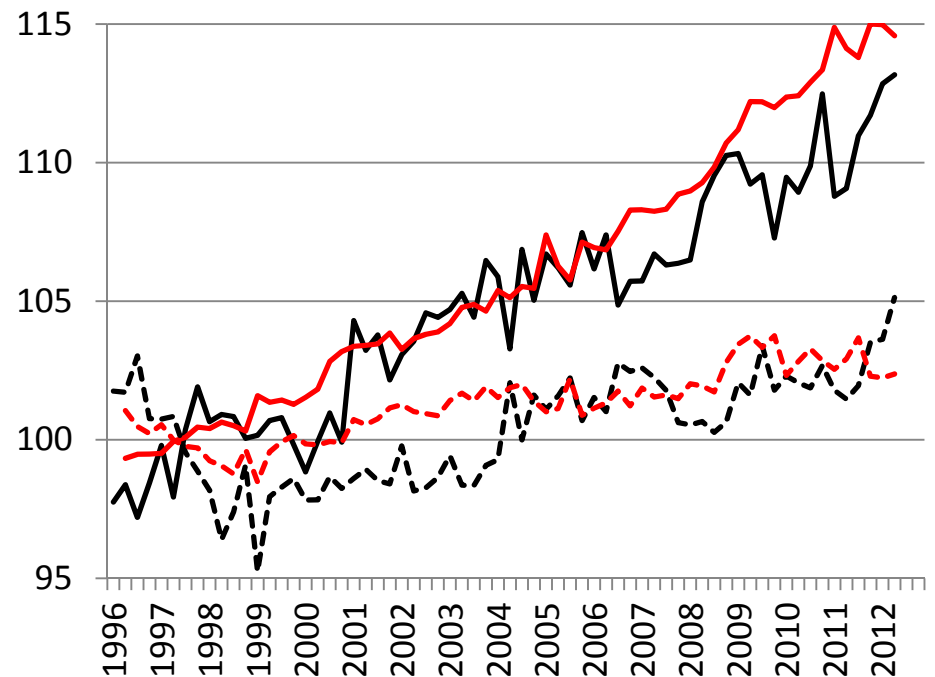
-- perhaps there are part-time (<35 hours) workers in the LEHD

Increasing Inequality, 1996 - 2012

CPS-ORG (*13) & LEHD, quarterly SA, real 2012 \$



— $\ln(P90^*13) - \ln(P10^*13)$, Published CPS ORG, 1997=100
 — $\ln(P90) - \ln(P10)$, LEHD FQ Main Job, 1997=100

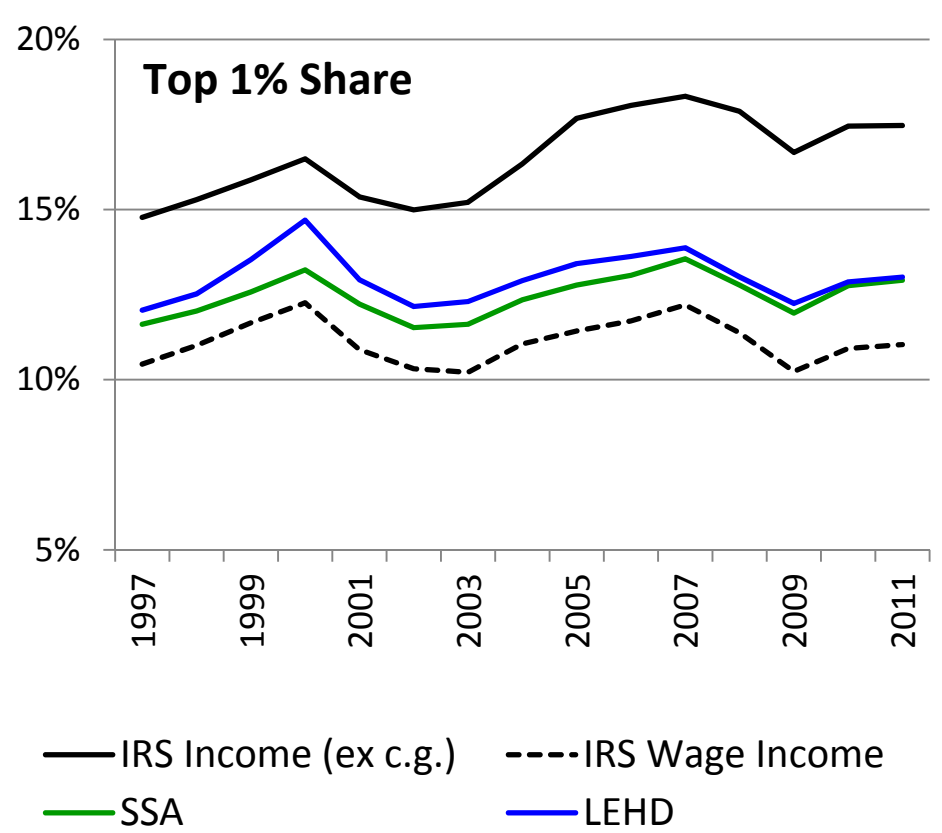
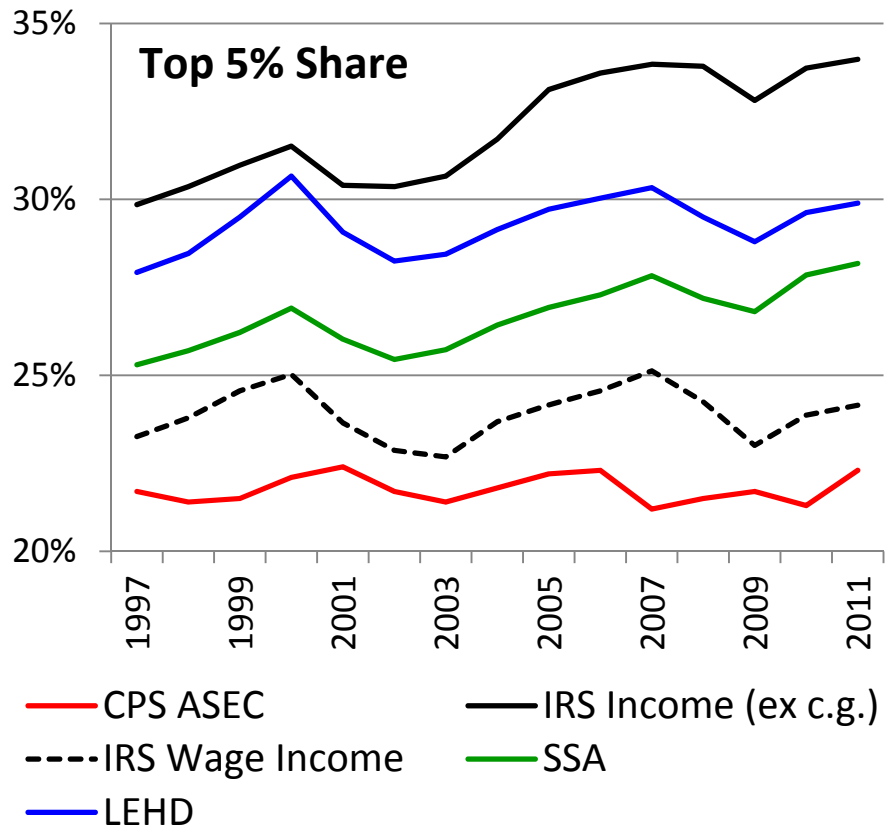


— $\ln(P90^*13) - \ln(P50^*13)$, Published CPS ORG, 1997=100
 — $\ln(P90) - \ln(P50)$, LEHD FQ Main Job, 1997=100
 - - - $\ln(P50^*13) - \ln(P10^*13)$, Published CPS ORG, 1997=100
 - - - $\ln(P50) - \ln(P10)$, LEHD FQ Main Job, 1997=100

LEHD Top Percentile Shares

- ❖ Attempting to mimic the CPS-ASEC and IRS data
 - Annual earnings of individuals from all jobs during the year
 - 943 million individual-year observations, 1997 – 2011
 - approximately 63 million persons each year (15 years)

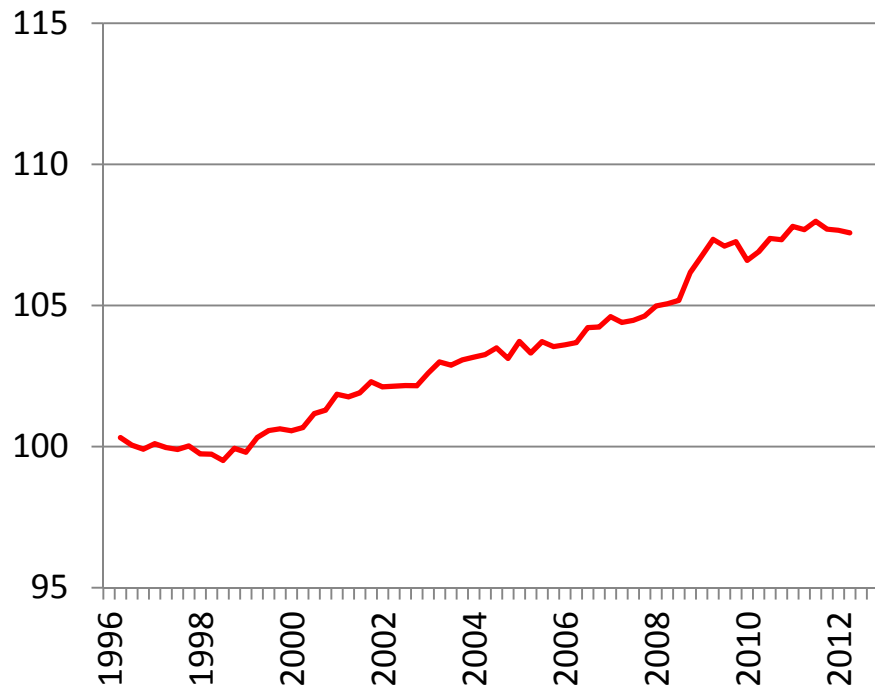
IRS, SSA, CPS-ASEC, and LEHD



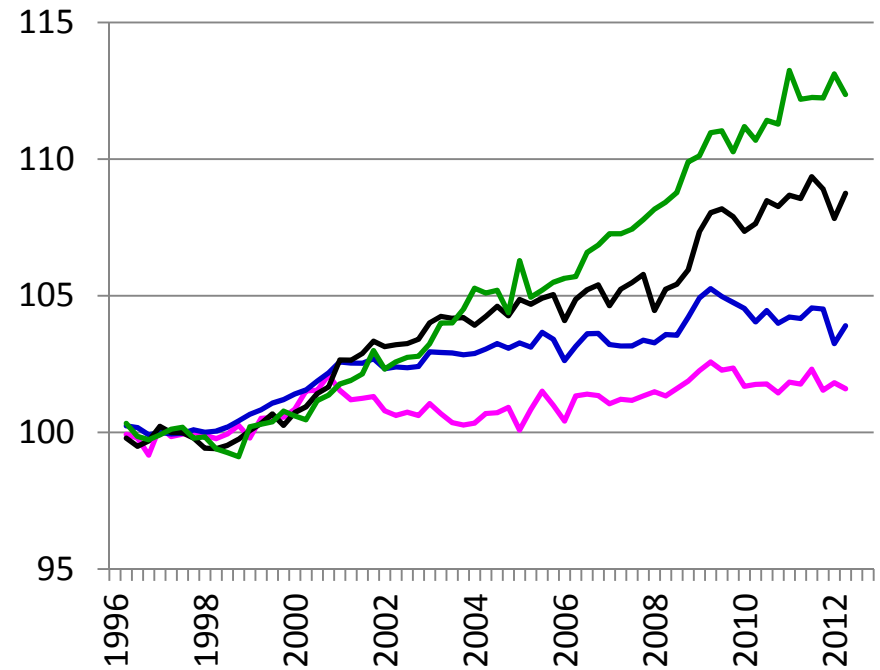
Summary of Comparison

- ❖ Acknowledge differences in scope and definitions
- ❖ Comparing LEHD {10, 50, 90} with CPS ORG:
 - 50th & 90th percentiles almost identical, 10th different
 - 90/10, 90/50 and 50/10 trends very similar
- ❖ Comparing LEHD top % shares with IRS wages & SSA:
 - Levels similar, all show growth during 1997-2011
 - Time series correlations are $>.8$
- ❖ Now turn to the value added of LEHD statistics

90/10 Ratios by Firm Size, LEHD



— $\ln(P90) - \ln(P10)$, LEHD FQ Main Job, 1997=100



- $\ln(P90) - \ln(P10)$ Firm Size <20, 1997=100
- $\ln(P90) - \ln(P10)$ Firm Size 20-99, 1997=100
- $\ln(P90) - \ln(P10)$ Firm Size 100-999, 1997=100
- $\ln(P90) - \ln(P10)$ Firm Size 1000+, 1997=100

The Role of the Firm

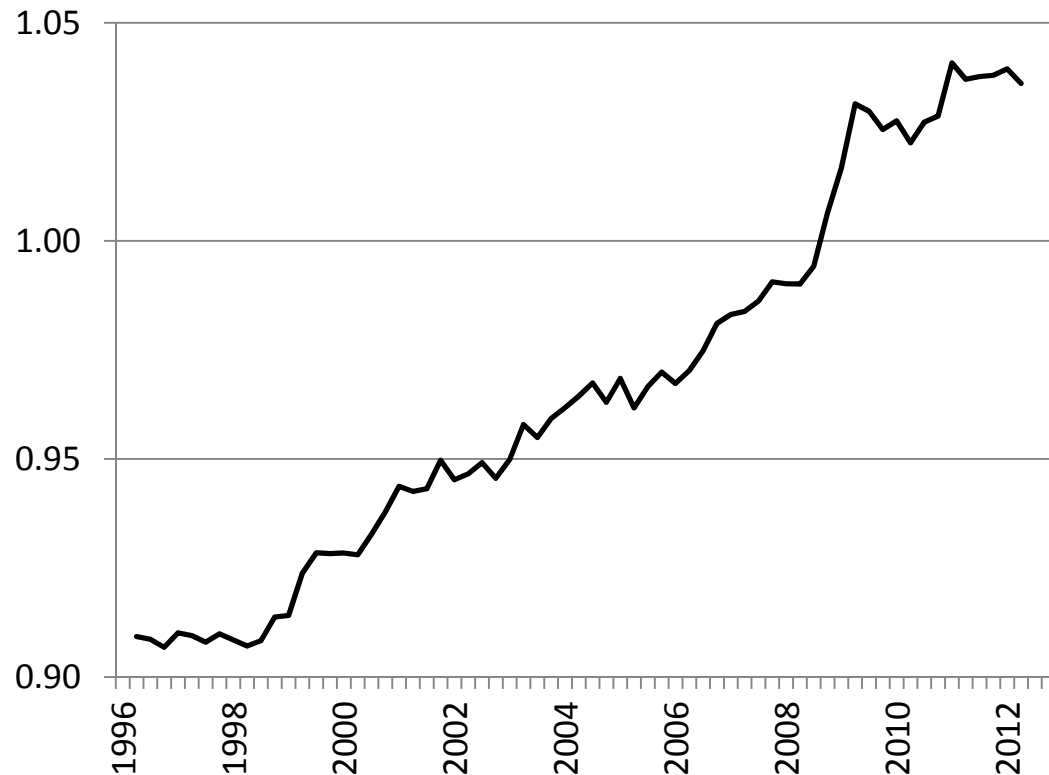
We know that quite a bit of cross-sectional earnings variance is across establishments:

- a) Groshen (1991): 17-85% [6 manufacturing industries]
- b) Davis & Haltiwanger (1991): 51-58% [Manufacturing]
- c) Dunne, Foster, Haltiwanger, & Troske (2004): 53-69% [Manufacturing]
- d) Lane, Salmon, & Spletzer (2007): 50% [All industries]
- e) Handwerker & Spletzer (2014): 55% [All industries]

We also know that a sizeable amount of the *growth* in earnings variance is across establishments:

- a) Davis & Haltiwanger (1991): 48% [Manufacturing]
- b) Dunne, Foster, Haltiwanger, & Troske (2004): 90% [Manufacturing]
- c) Handwerker & Spletzer (2014): 73% [All industries]

The Role of the Firm (continued)



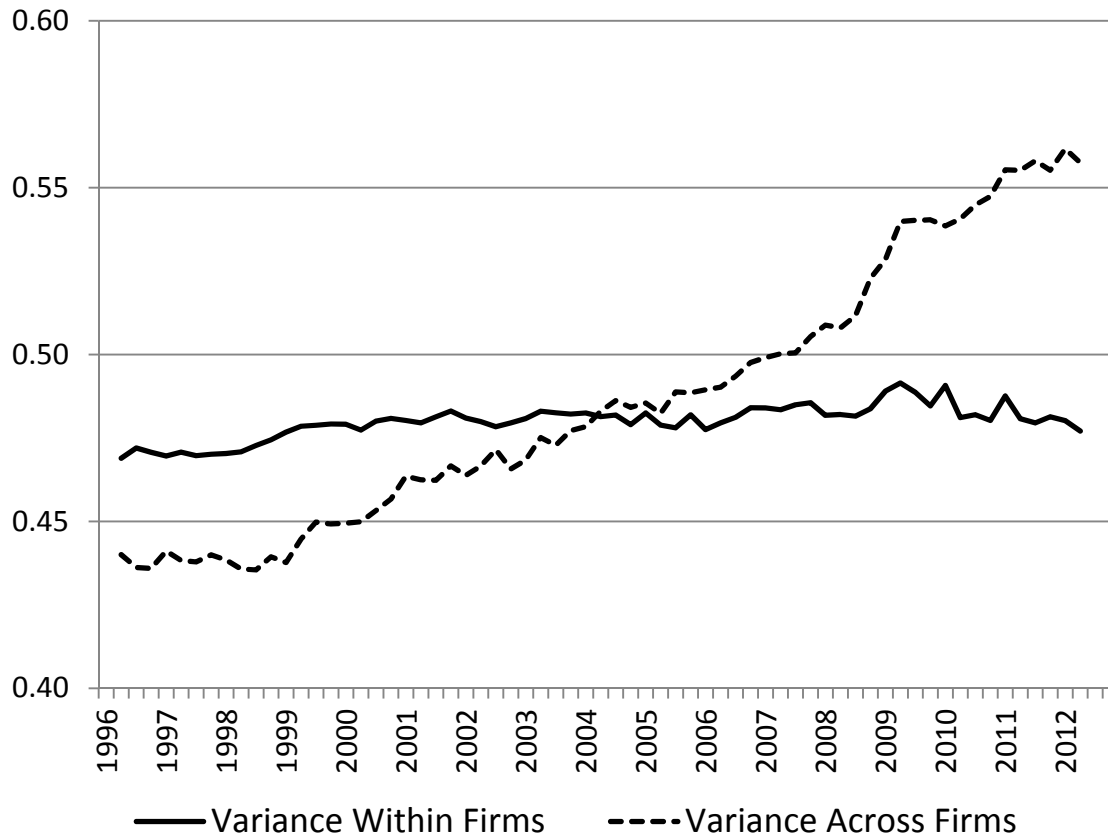
— Variance LN(Winsorized FQ Real Earnings)

We switch from 90-10 ratios to variances

Why? To use the simple decomposition

Total Variance =
Variance Across Firms +
Variance Within Firms

Variance Decomposition: LEHD



On average, 50.3% of cross-sectional earnings variance is across firms

93.5% of the growth in earnings variance is across firms

LEHD Top Percentile Shares

What are the age and gender distributions of workers in the top 5%?

Age	All Workers	Workers in the top 5%
<30	31.8%	2.7%
30-34	11.2%	8.8%
35-39	11.4%	15.1%
40-44	11.4%	18.5%
45-49	10.6%	18.7%
50-54	8.9%	16.2%
55-59	6.6%	11.3%
60-64	4.1%	5.8%
>65	4.1%	2.9%

Gender	All Workers	Workers in the top 5%
Male	53.1%	79.0%
Female	46.9%	21.0%

Questions for FESAC

- 1) Are the inequality statistics currently in the public domain sufficient for analysts and policymakers, or should the statistical agencies publish similar statistics from other data sources?
- 2) Do you have any thoughts about the similarities and differences in inequality statistics from different data sources? Differences undoubtedly reflect different concepts and definitions across data sources, but differences might also signal issues that the statistical agencies should research and understand.
- 3) One goal of publishing inequality statistics is to help understand why inequality is increasing. As such, are there any specific statistics from the LEHD that you would find useful?