

The Contribution of Factor Incomes to Income Inequality: A Process Perspective on Class (and the KOB decomposition)

2025-04-15

Bur, G.



Executive summary I (newish)

Background: use of class as an analytical lens in research on explosion of economic inequality is surprisingly rare. Yet, it seems like a good candidate: level and type of income almost certainly must be linked. By class, I mean ownership relation to means of production; others use class to just mean “large, analytically defined aggregate categories” (Weeden and Grusky 2012).

What barriers exist? Many people fit into "contradictory class locations" (Wright 1998): curse of dimensionality troubles attempts to catalogue all such locations. Many surveys ask vaguely if at all about class.

What does my paper do? I solve both problems adapt the CV decomposition (from demography) to decompose income into capital + labor components using CPS/SOI, making use of class-theoretic information without slotting people into classes.

What do I find? Capital income matters enormously if SOI data are used; less so if CPS is used despite dataframes otherwise being very similar.

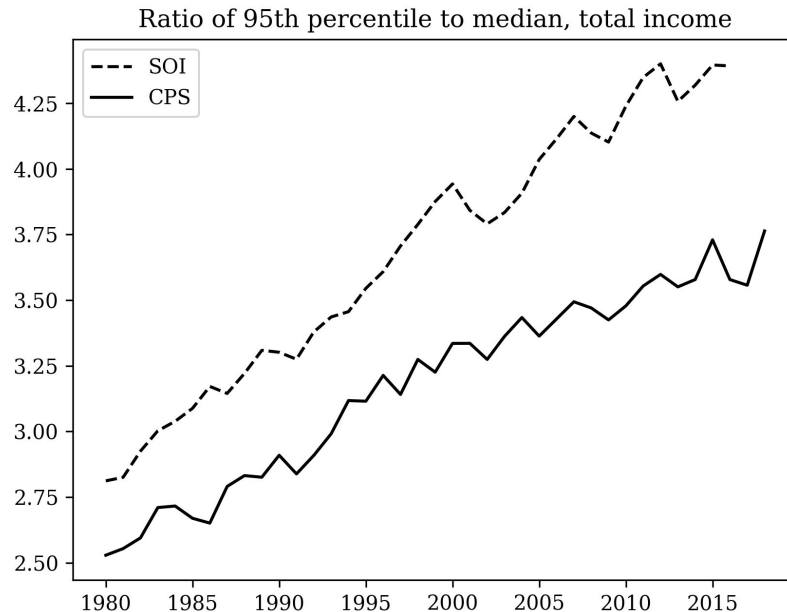
The puzzle I: explaining/characterizing increasing inequality

Economic inequality on rise, not much agreement on the causes.

Largely a "right-tail" phenom.

Proposed causes include...

occupational closure (Weeden and Grusky 2014), **globalization** (Alderson 2002), **“skill-biased technical change”** (Autor 2013), and/or **financialization** (Lin and Tomaskovic-Devey 2013).



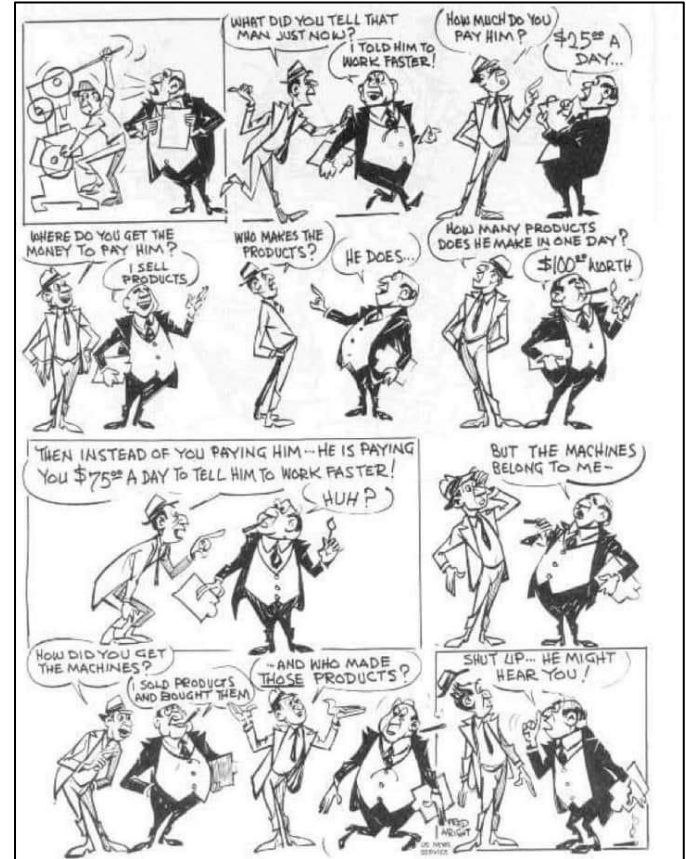
Here, I show quantile ratios, but my primary outcome measure is the coefficient of variation.

The puzzle II: where is class in this literature?

On one hand, it seems obvious that class should play some role in the ineq. literature, but it largely hasn't.

Again, class for my purposes := an "economic" category, i.e. ownership relationship of people to resources necessary for industrial production.

So, class is isomorphic to factor income:
capital income: capitalist :: labor income: worker. Should play some role in understanding inequality.



The puzzle III: one approach I'll set aside

What would it look like for class to matter?

Wodtke (2016) is a rare entry into this type of literature—where people are, in essence, carrying out a regression analysis—which uses class.

He does a fancy sort of ANOVA and finds that income variance both between and within classes increased.

Key limitation: this requires slotting people into fixed categories.

TABLE 1
DECOMPOSITION OF TRENDS IN PERSONAL INCOME INEQUALITY

PERIOD	COMPOSITIONAL EFFECT		BETWEEN-GROUP EFFECT		WITHIN-GROUP EFFECT	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Social class position:						
1980s–1990s	-.04	(-.06, -.01)	-.15	(-.24, -.08)	1.19	(1.11, 1.28)
1980s–2000s	-.11	(-.13, -.09)	.18	(.15, .20)	.94	(.91, .96)
1990s–2000s	-.10	(-.13, -.07)	.33	(.29, .37)	.77	(.73, .81)
Social class strata:						
1980s–1990s	.02	(-.01, .06)	-.19	(-.28, -.11)	1.16	(1.08, 1.25)
1980s–2000s	-.10	(-.13, -.08)	.29	(.26, .32)	.81	(.78, .84)
1990s–2000s	-.16	(-.20, -.12)	.55	(.50, .61)	.61	(.56, .65)

NOTE.—The sample includes respondents who are 18–65 years old and work full-time in the 1980–2010 GSS waves. Results are based on 10 multiple imputation data sets. Confidence intervals (CI) are based on quantiles of 500 bootstrap sample estimates.

The puzzle IV: theoretical barriers

- Often messy; unclear where to stop stratifying (Wright 1985).
- Wright's classic table to right: three forms of exploitation (worker/capitalist, organization, skill).
- Merges meanings of exploitation that are very different, sneaks a version of labor theory of value in back-door.

88

TABLE 3.3
Typology of Class Locations in Capitalist Society

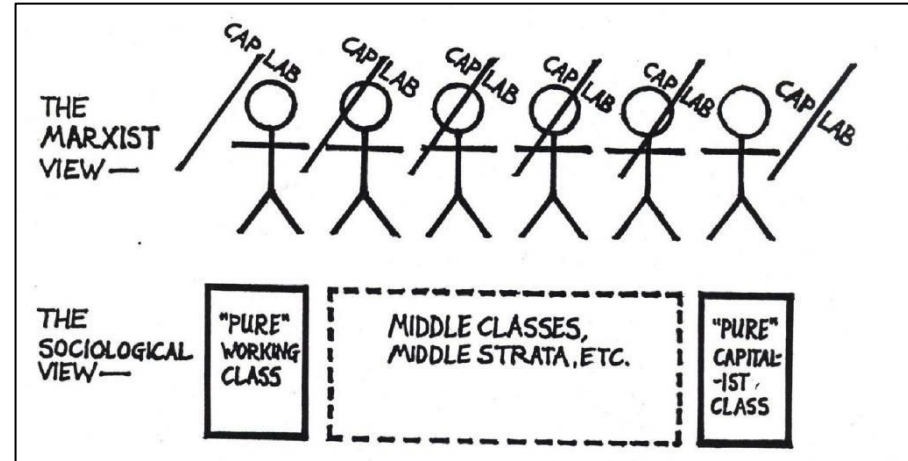
Assets in the means of production		Non-owners [wage labourers]			
Owners of means of production					
Owens sufficient capital to hire workers and not work	1 Bourgeoisie	4 Expert Managers	7 Semi Credentialed Managers	10 Uncredentialed Managers	+
Owens sufficient capital to hire workers but must work	2 Small Employers	5 Expert Supervisors	8 Semi Credentialed Supervisors	11 Uncredentialed Supervisors	>0
Owens sufficient capital to work for self but not to hire workers	3 Petty Bourgeoisie	6 Experts non-managers	9 Semi Credentialed Workers	12 Proletarians	-
		+	>0	-	
		Skill/credential assets			

Organization assets

The puzzle V: a theoretical suggestion

Marx himself provides an interesting alternative by spending ~0 time defining who sits in what class and a ton of time on process:

... class is not a group identity but the name for a process. Workers are workers *to the extent that* they perform wage-labor and, in many cases, also capitalists to the extent that they (or their pension) employ capital.



Source: Gunn (1987).

The puzzle VI: a theoretical suggestion, 2

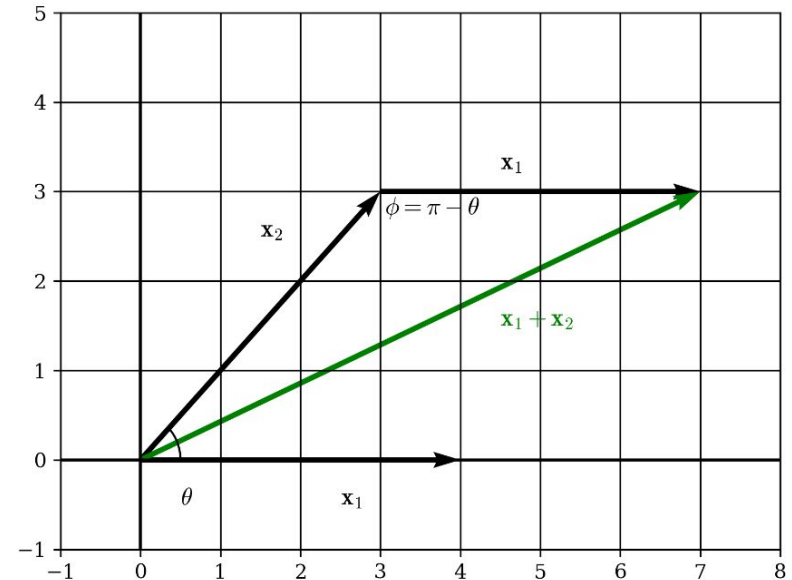
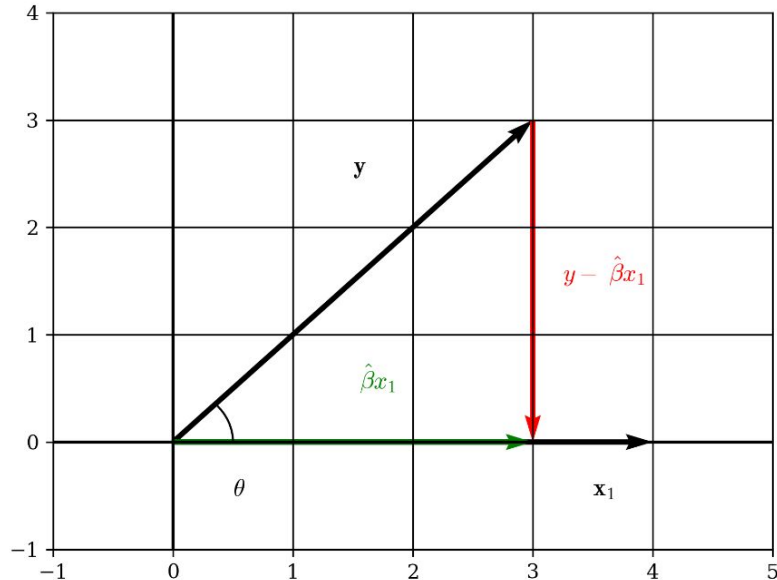
- I will **sidestep the problem** of defining exact class boundaries and consider people on a capitalist-worker scale, $\propto$ their capital- (labor-) income. We don't have to assign people to groups.
- **Many of the most robust measures of income inequality *do not require slotting people into fixed groups***, e.g. the coefficient of variation. Some do implicitly, as in quantile ratios or Gini.
- **There are essentially different ways of thinking about inequality.**
 - **Put people into groups, regress inequality on those groups** (or interactions, etc.). Class in the sociological sense: static category at some point in time t_0 could be one.
 - **Treat inequality as general dispersion or entropy.** Class here shows up as sources of income, or factor incomes.

Method 0: big idea (new)

- **Take robust measure of inequality:** normed variance of income.
- **Assume it is a composition of capital and labor income.**
 - Not a causal argument in the direct sense: no least squares problem, no error term.
- **Decompose into components** that come fundamentally from coefficients of variation of capital income and labor income individually, their shares of income, and correlation between them.
- **What does it tell us? What is counterfactual?**
 - **If inequality is rise of super-managers, all about micro-classes or professions** (broadly: Piketty, Weeden styles args. respectively) ... **there should be little contribution from capital income.**

Method I: regression vs. decomposition

In subject space, regression is the application of the law of cosines shown on the left; decomposition is the application shown on the right.



Method II: choice of outcome (new)

- **Why relvariance/CV?** Many choices of outcome, all shown in presentation: Gini (generalized shares), various quantile ratios.
- **Desirable properties of inequality measure** (Cowell 2000):
 - **Mean-independent** (CV and most measures succeed)
 - **Passes principle of population** (CV and most succeed)
 - Ineq. should not depend on N
 - **Passes principle of transfers** (CV succeeds, some fail)
 - Loosely transfer of income from rich person to poor person should reduce inequality so long as ranks aren't swapped.
 - **Decomposable** (CV succeeds, many fail)
 - If ineq. $\downarrow$ in all sub-groups, overall inequality $\downarrow$.
 - General: inequality should be a function of subgroups.
 - CV succeeds, Gini fails. To succeed, must be a member of generalized entropy family.

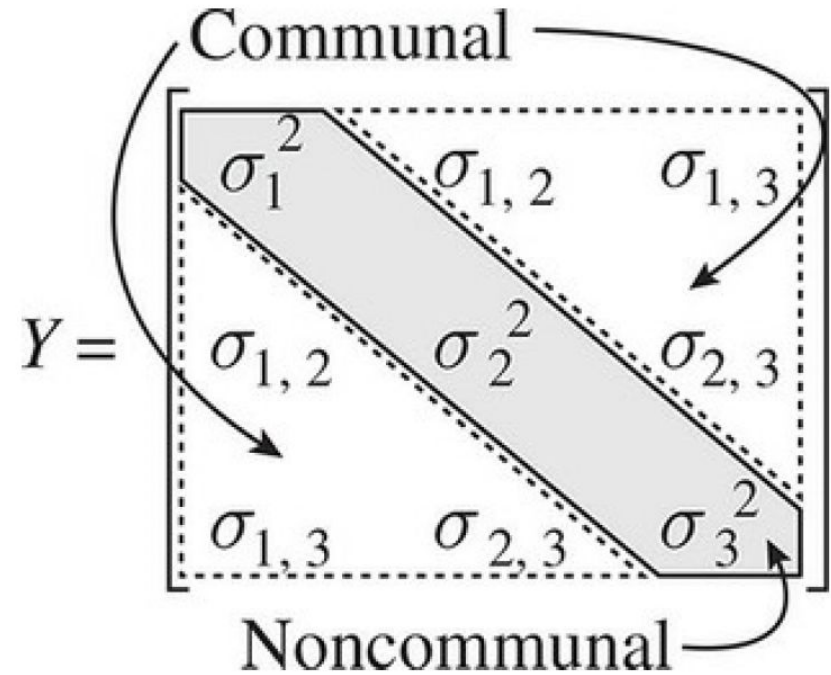
Method III: the Bienaymé decomposition

This approach draws on a fundamental identity sometimes called the Bienaymé formula.

Simply put, for any composite variable $Y = \sum Y_j$, total σ^2_Y is the sum of all the variances σ^2_j , plus twice the sum of all of the unique pairwise covariances $\sigma(X_j, X_k) : j \neq k$.

Illustrated on the right with a generic variance-covariance matrix.

Holds for square of CV (relvariance).



Source: DeVellis (1991).

Method III.5: the RV decomposition (newish)

Nice decomposition from Cancian and Reed that I proved (prob. not first to get there).

Write RV equation: Bienaymé decomp. divided by constant.

To get ind. RVs for each var, divide by its mean (or square), then mult. by it. Write correlation as cov. by SDs.

Result: get expr. where RV is a function of corr, CV or RV of each var, share of each var (or square).

$$\begin{aligned}\frac{\sigma_Y^2}{\mu_Y^2} &= \sum_{k=1}^p \frac{\sigma_k^2}{\mu_Y^2} + 2 \sum_{j \neq k} \sum_k \frac{\sigma_{jk}}{\mu_Y^2} \\ &= \sum_{k=1}^p \frac{\sigma_k^2}{\mu_k^2} \frac{\mu_k^2}{\mu_Y^2} + 2 \sum_{j \neq k} \sum_k \rho_{jk} \frac{\sigma_k}{\mu_k} \frac{\sigma_j}{\mu_j} \frac{\mu_j}{\mu_Y} \frac{\mu_k}{\mu_Y} \\ &= \sum_{k=1}^p RV_k s_k^2 + 2 \sum_{j \neq k} \sum_k \rho_{jk} RV_j RV_k s_j s_k\end{aligned}$$

Then, the contribution to the total relvariance is as follows.

$$C_f = RV_f s_f^2 + 2 \sum_{f \neq k} \rho_{fk} RV_f RV_k s_f s_k$$

Method IV: meaning (new)

The total relvariance is a function of the following:

- Each individual variable's RV times its squared share of income.
- The product of ...
 - each pairwise correlation involving that variable
 - its CV and that of every other variable
 - its share and every other variable

Here, it's simple: the components are 1) RV/CV of capital and labor income; 2) their correlation; 3) their individual shares and the squares of those shares.

Method IV.5: kind of argument (new)

- **Want to emphasize: not a causal argument.** Definitionally, all variance in income comes from factor income.
- Instead, two things:
 - **Capital income matters**; it's not just the rise of super-managers, ultra-high-labor incomes (which people have always suggested is misreported capital income).
 - **Relatedly, lots of inequality has to do with old distinction between labor income and capital income.** It's not "class causes inequality", but "inequality data do not falsify Marx's analysis".
- Counterfactual: if it's all just inequality within occupations/professions, **we should expect most inequality to be driven by labor income.**

Method V: data (new)

Previous versions of the paper used PSID data. Messy, complicated dependence between waves. **This paper uses...**

- **Current Population Survey (CPS)**
- **What I call the SOI (Statistics of Income) data**, which are based on IRS records (happy Tax Week!)
 - Produced by Saez and Zucman (2018). Unfun fact, IRS no longer even gives out the data they based their method on, called Public Use Files (I called them and asked!)
 - I describe SOI data on the next page in more detail.

Method VI: SOI data (new)

- Saez and Zucman pick relatively constant variables from IRS forms.
- They stratify records by itemization/marriage status/age dummy.
 - Within strata, highest adjusted gross income (AGI) was used as a benchmark, and other records in the stratum were ranked by their squared distance on a set of 15 income variables. The five closest records were taken out; process was repeated until all records had been grouped.
 - Then, **actual observations were constructed by taking an average of the groups.** Frequency weights used (many joint-filers).
 - **"Blurring" the data.**

Method VII: SOI data continued (new)

- **Authors identify extreme records in the data. I chose to take these out.** This generally means extremely high or low income.
- No age, but **I drop persons who were 1 on 65+ dummy.**
- To try and capture active labor force population, **I drop people with below ½ full-time minimum wage in 2010 constant dollars** (converted to this because CPS is in 2010 dollars).
- **Income: restricted to capital + labor income.** Capital income defined to be the sum of dividend income, interest income (taxable and not), rental income, royalties income, estate income, schedule C (sole proprietorship), net income, partnership income, S corporation net income, and capital gains income. I remove self-employment income, subcomponent of some of above, and pension income.

Method VII: CPS data (new)

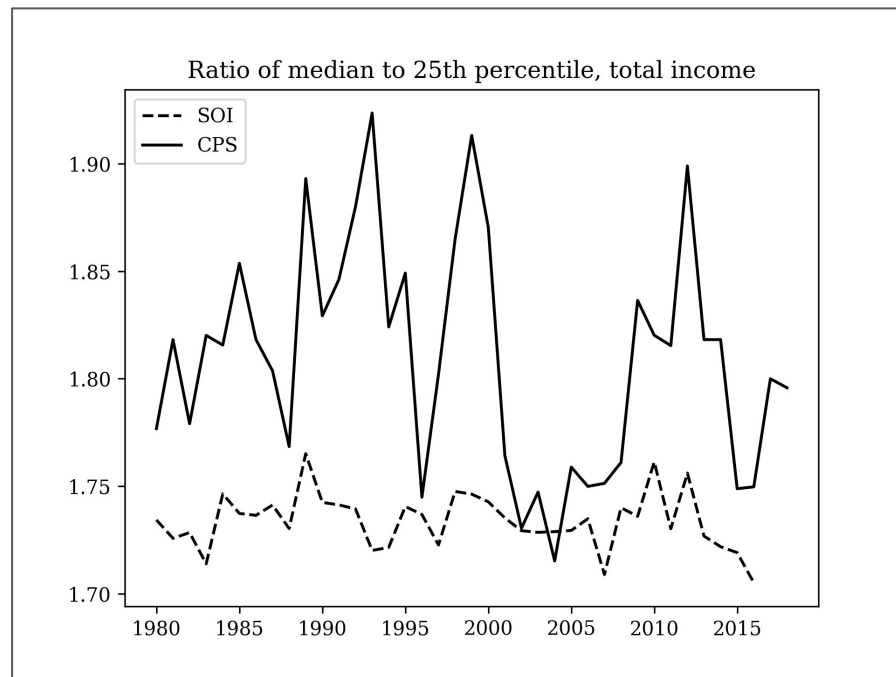
- Much better known. Used ASEC March extracts.
 - Nearly equal probability sample; weights don't change results much.
 - **Some amount of topcoding and imputation** (technically no MVs in CPS), which probably explains divergent results.
 - **Drop NILF, people below ½ FTM.**
 - **Capital income:** interest, dividends, royalties, trusts. Omit pensions.
- Used CEPR extract; they have bad data labeling that makes it appear that some variables double-count, but I figured out what was going on by tracking down original CPS codebooks.
- **To minimize dependence among individuals, only kept heads of households.** CPS has light panel structure (4-8-4); dropped one obs. for small subset of people who happened to be in two Marches.

Findings 0 (new): overall inequality

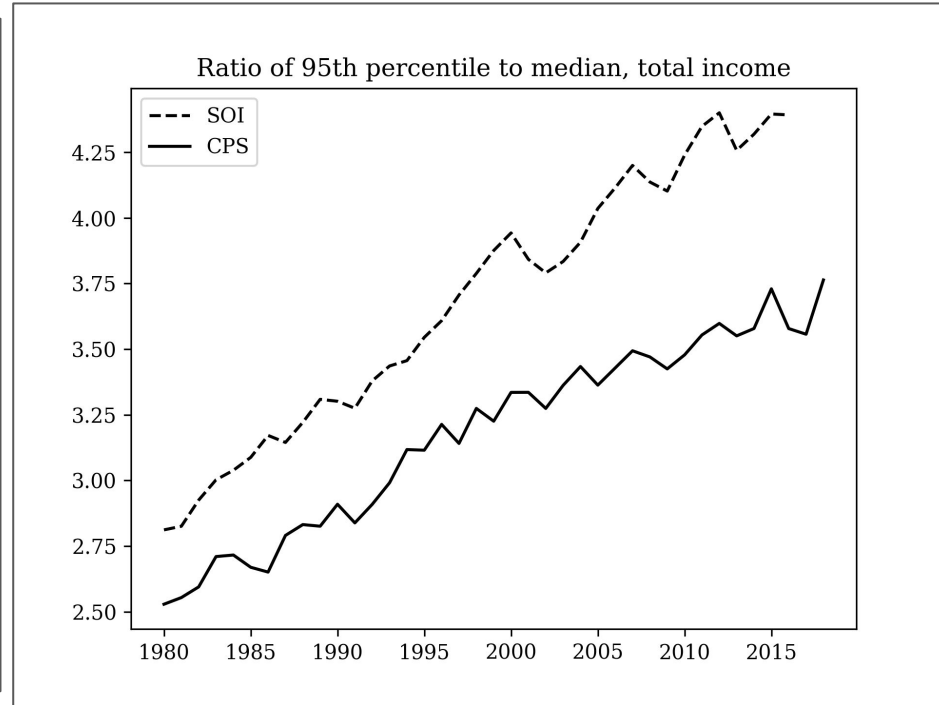
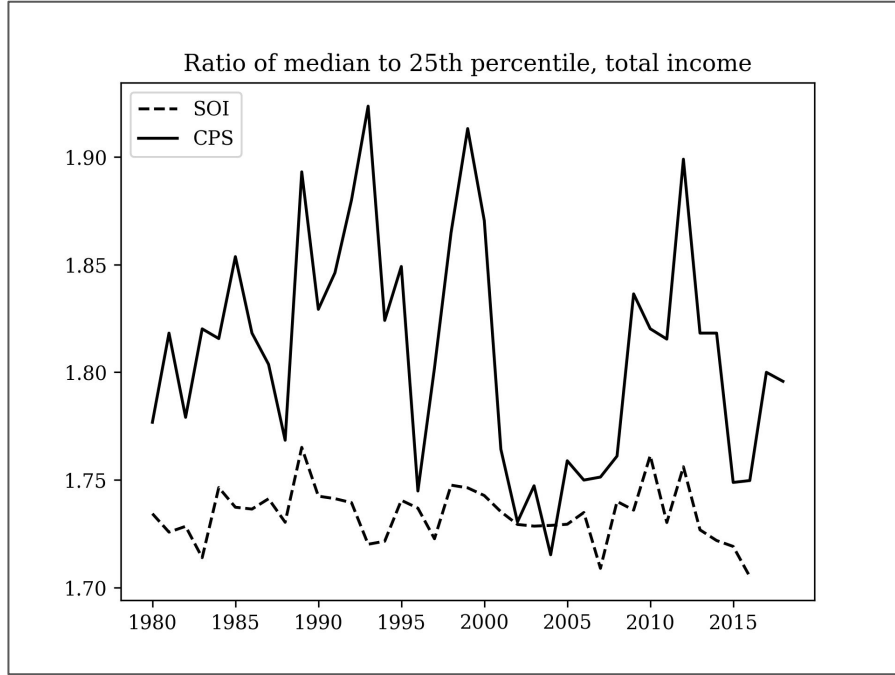
Some measures of inequality are very similar across dataframes, including the Gini coefficient and some quantile ratios.

These aren't bootstrapped; that is on my to-do list, although these measures have funny asymptotic properties and there are tens of thousands of people each year in each set anyways.

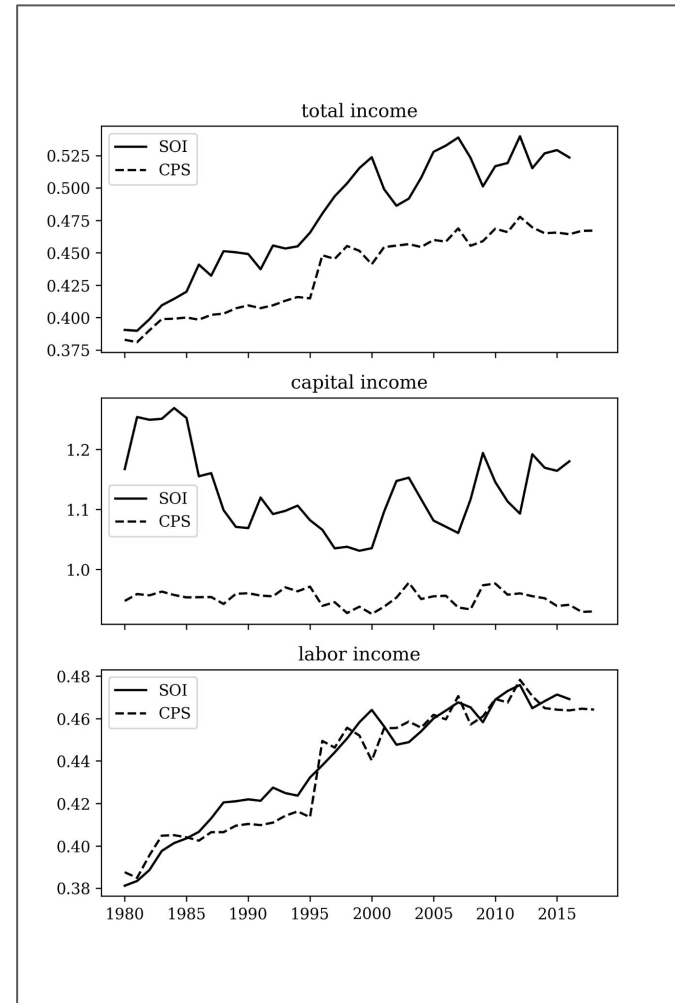
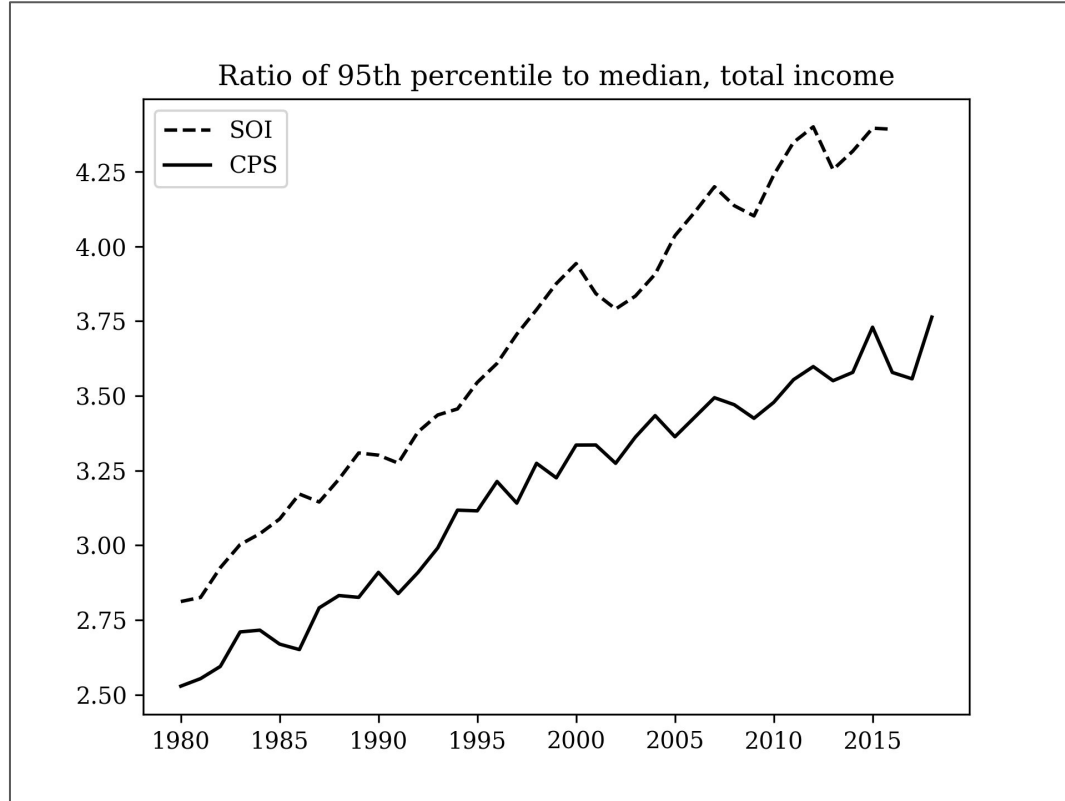
Inequality largely an upper-tail growth (note scale here).



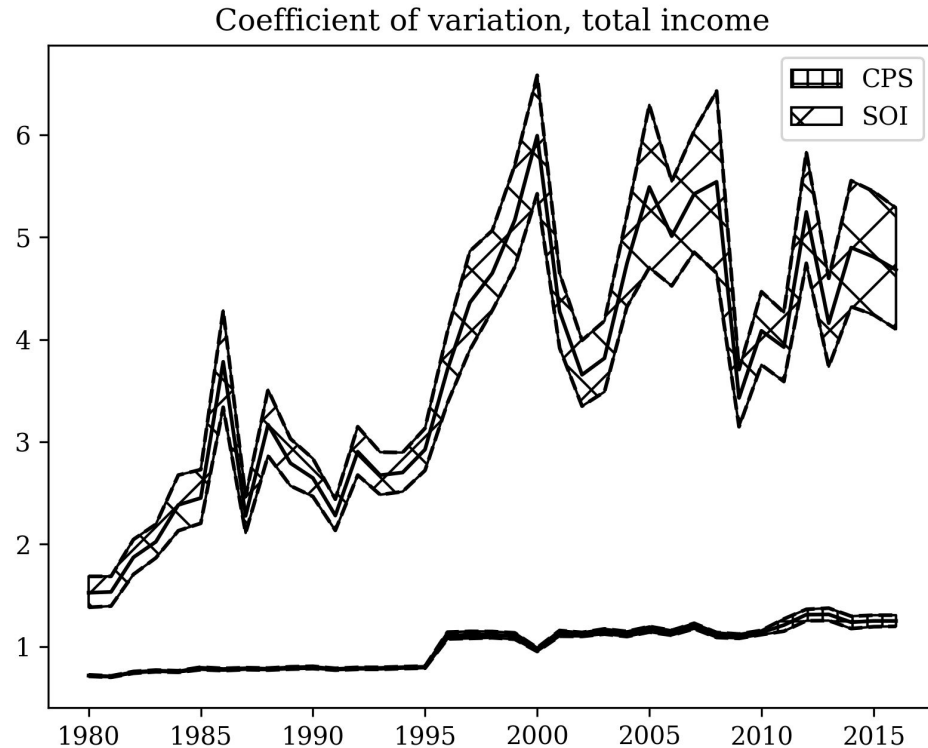
Findings I (new): right-tail story



Findings I.5 (new): Gini coefficient, percentile ratios



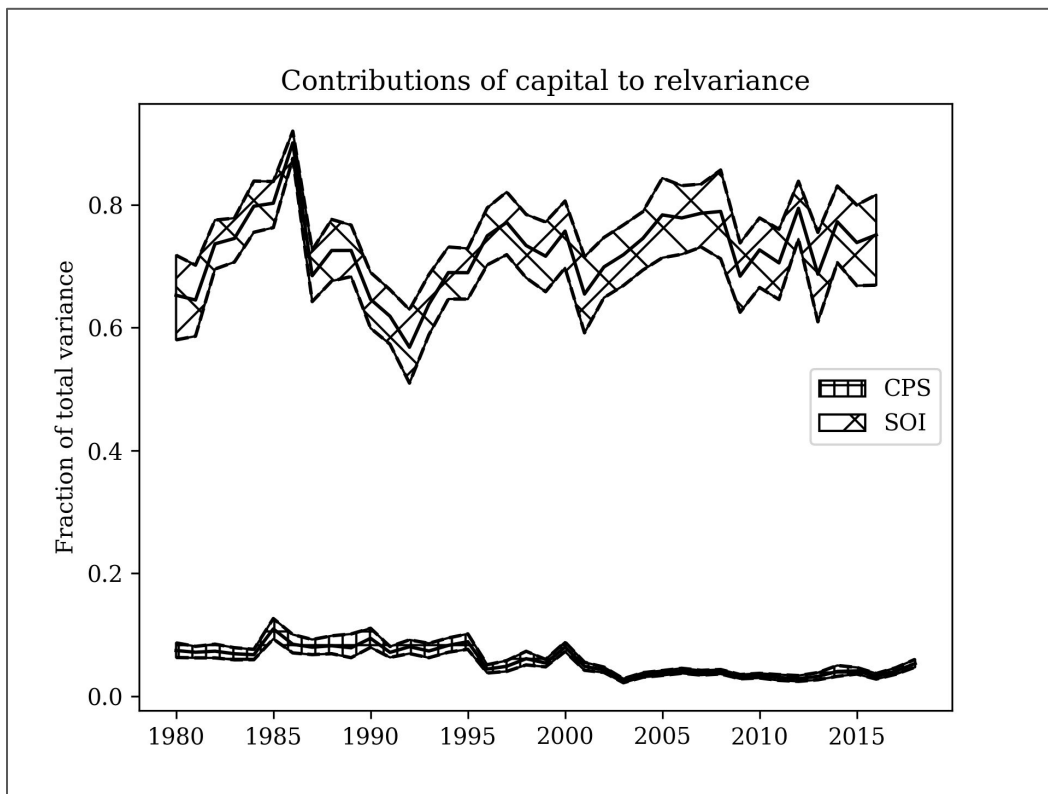
Findings I.75 (new): CV in the two data-sets



Findings II (new): *contribution from capital vastly higher in SOI*

This is a pretty shocking finding (to me). I sanity checked the data in every way I can think of (obviously made sure every round of bootstrap was algebraically correct). **I'll show more sanity checks soon.**

My guess: **less censoring of top incomes, especially capital income obviously.**

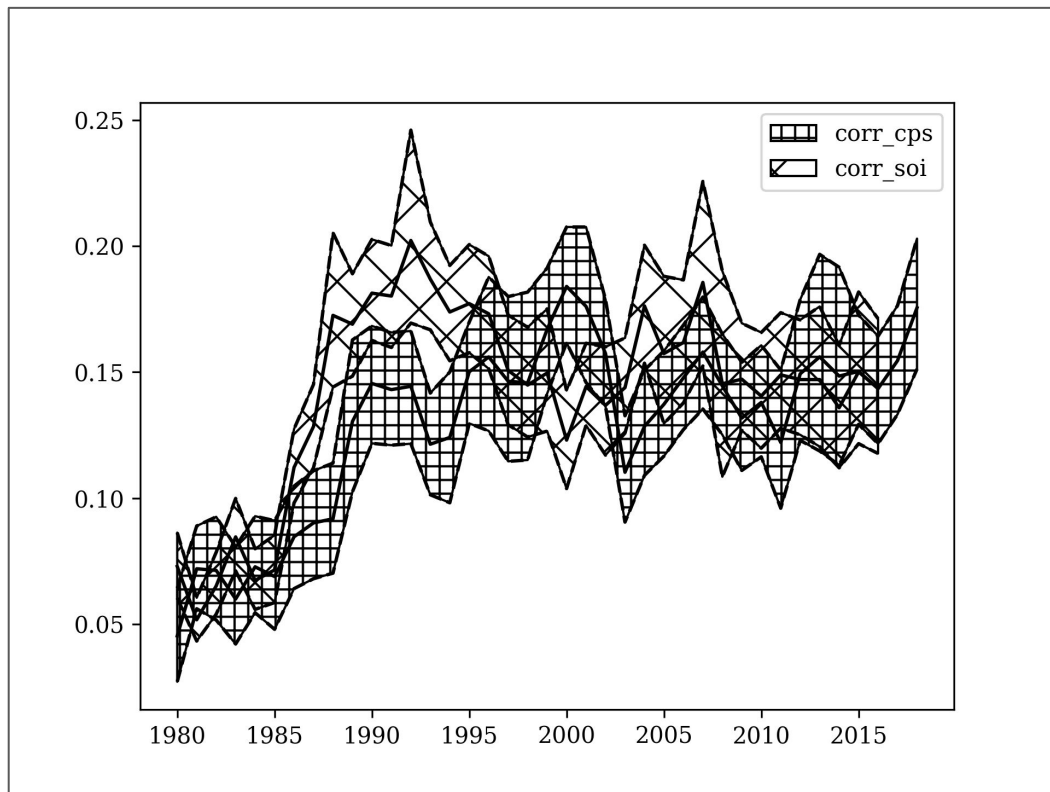


Findings III (new): correlation between types of income ~ =

Sanity check/decomposition analysis: correlation between capital and labor income.

Higher for the SOI data, but not by a lot, both trend in a very similar way, interestingly.

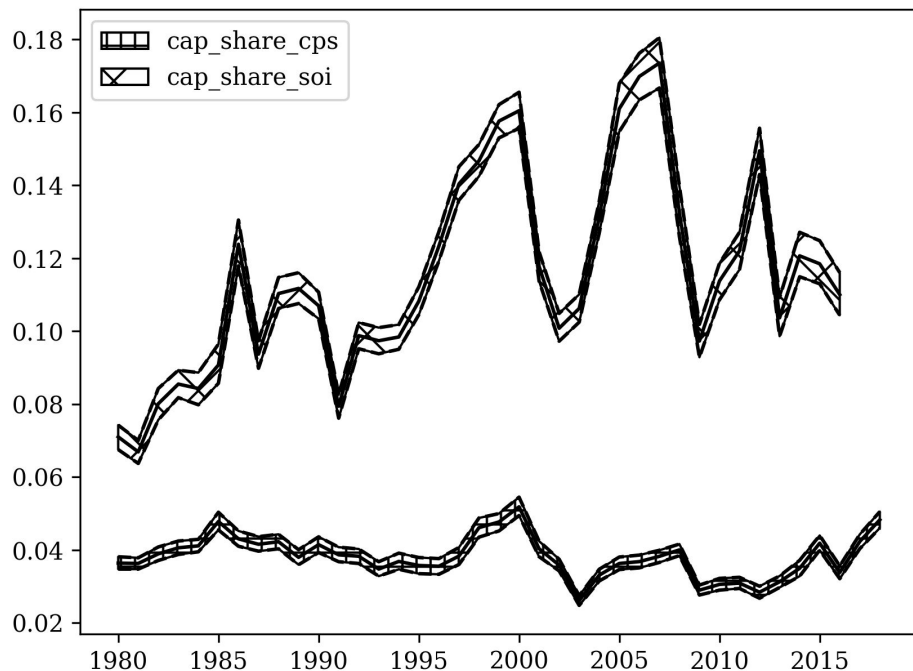
By the way, both series are backward looking.



Findings IV (new): share of capital income varies, not by tons

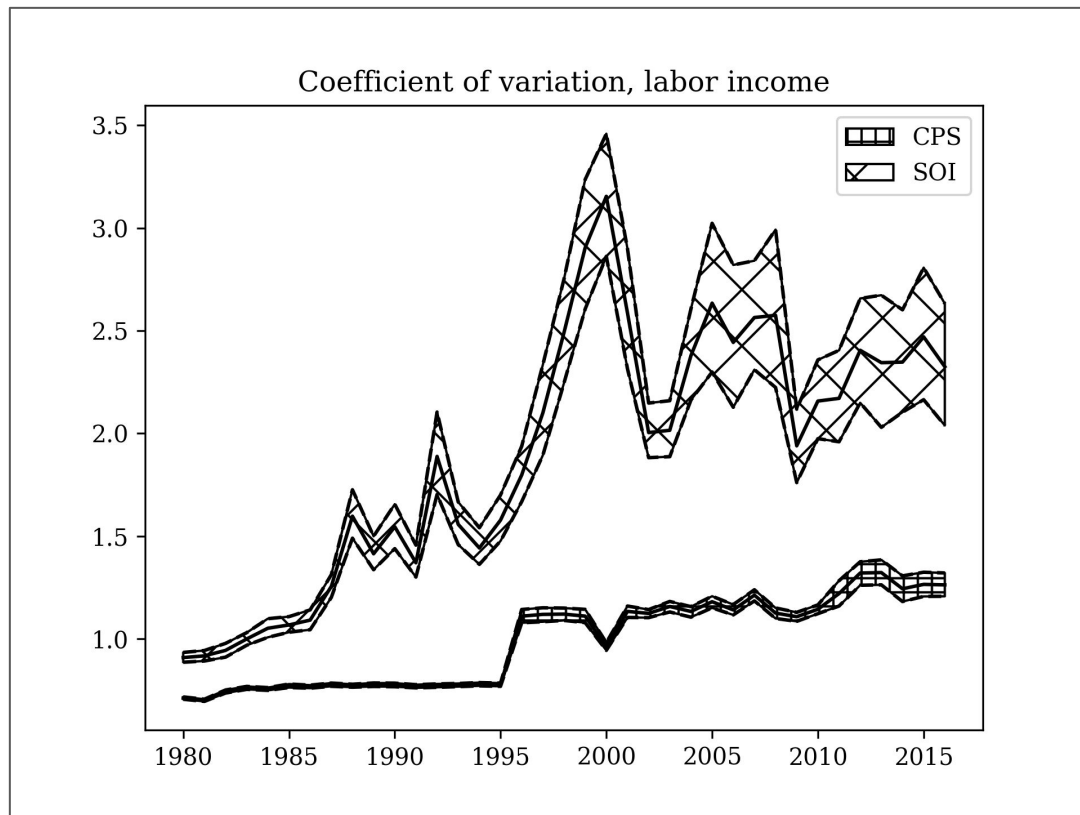
Sanity check/decomposition analysis: capital income share higher in SOI data, but not outlandish.

Unsurprising: topcoding in CPS, social desirability response bias, many reasons in total survey error framework, lots of prior research suggesting this should be true.



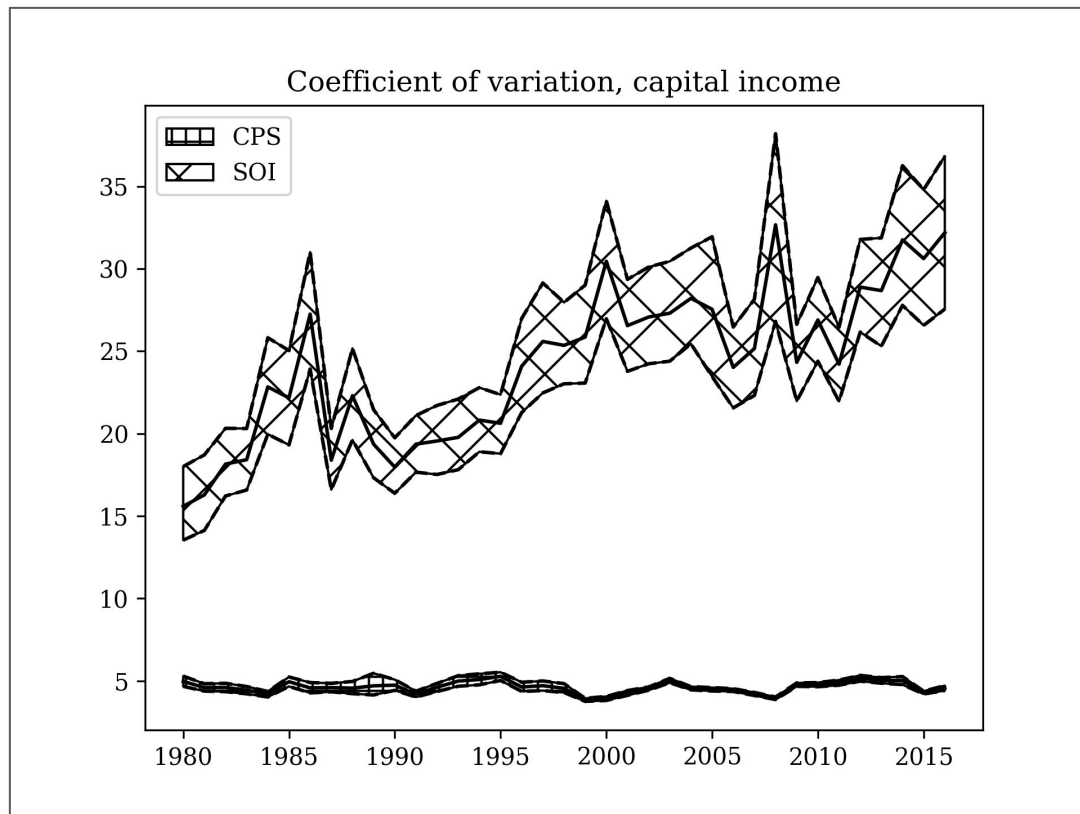
Findings V (new): CV of labor generally higher in SOI

- I'm a little suspicious of this, fairly noisy series, but not too wild.
- Not as much of a way to sanity-check this directly beyond algebraic checks that contributions sum to total in every bootstrap (verified); everything I could think to try was pretty much what you'd expect.



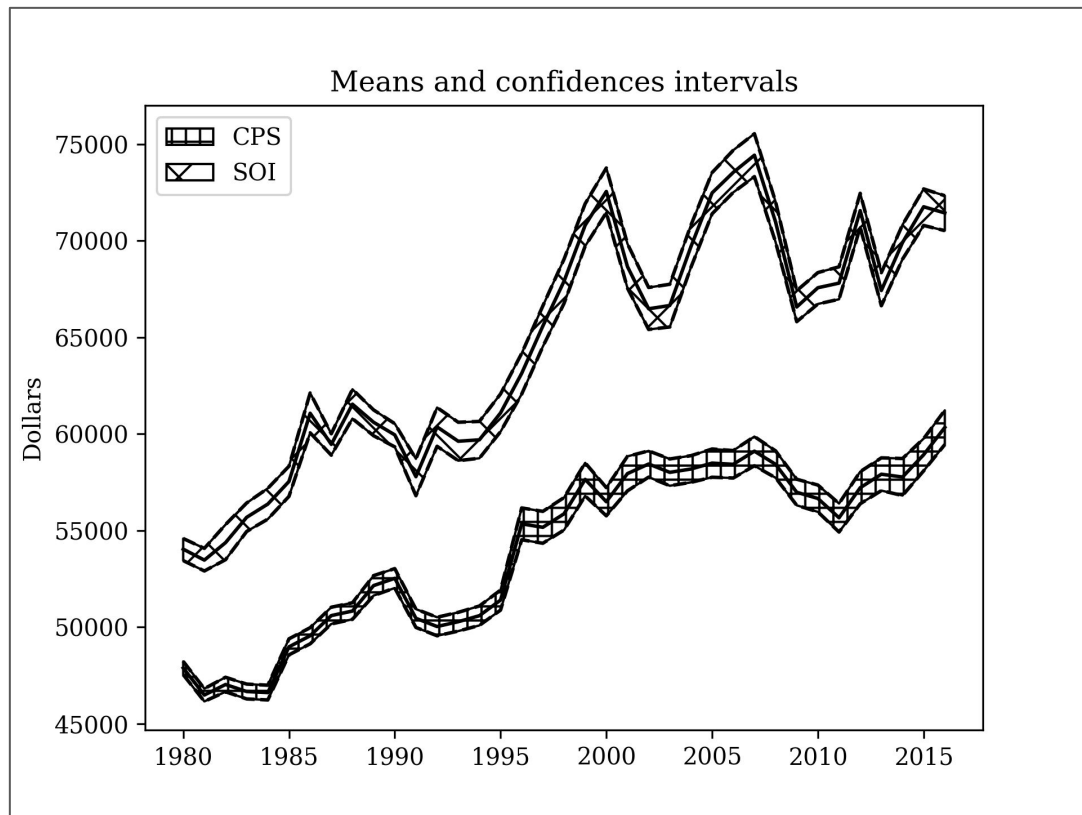
Findings VI (new): CV of capital income much higher in SOI

- Reasonably well-behaved series. *This is obviously the main source of the divergence.*
- Sanity-check: both should be much higher than respective labor income.



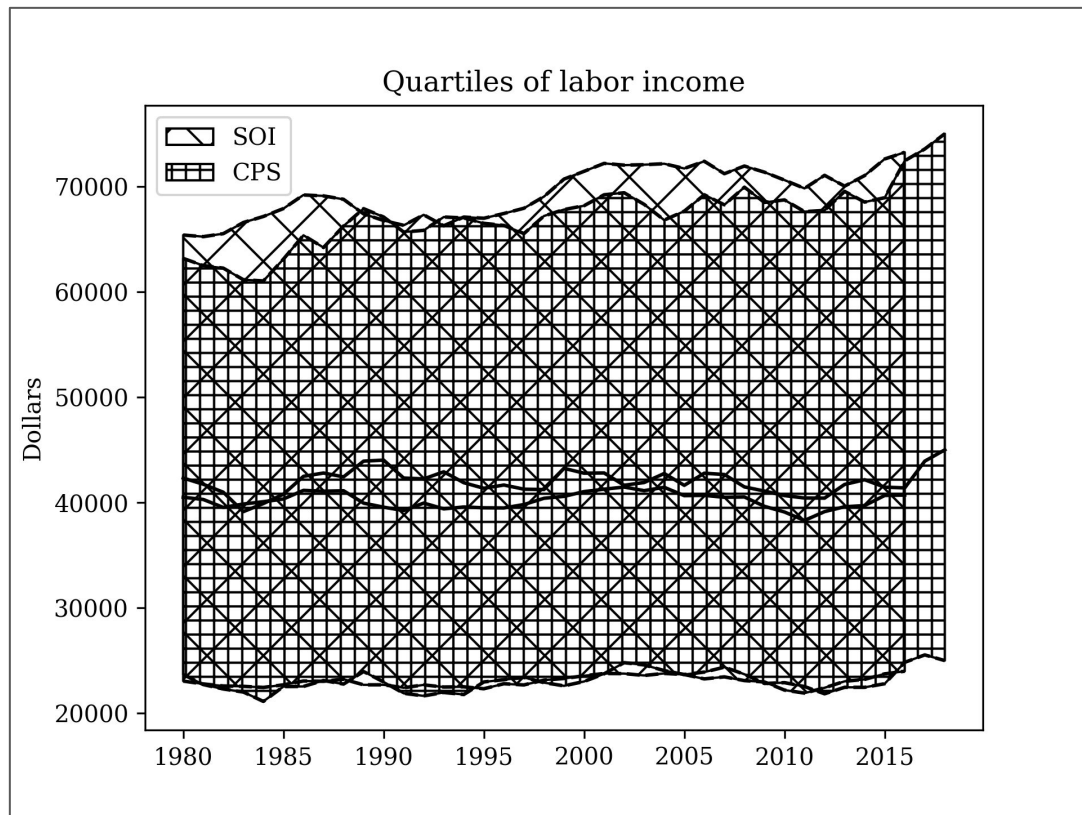
Findings VII (new): some more sanity-checking—means

- **Mean income higher in the SOI.**
- **These are both in the right ballpark.**
- Unsurprising that SOI is higher. No SDRB with IRS; lie to a pollster, don't lie to the tax cops.
- Cocktail napkin adjustment to income using CPI (doesn't matter for CV).



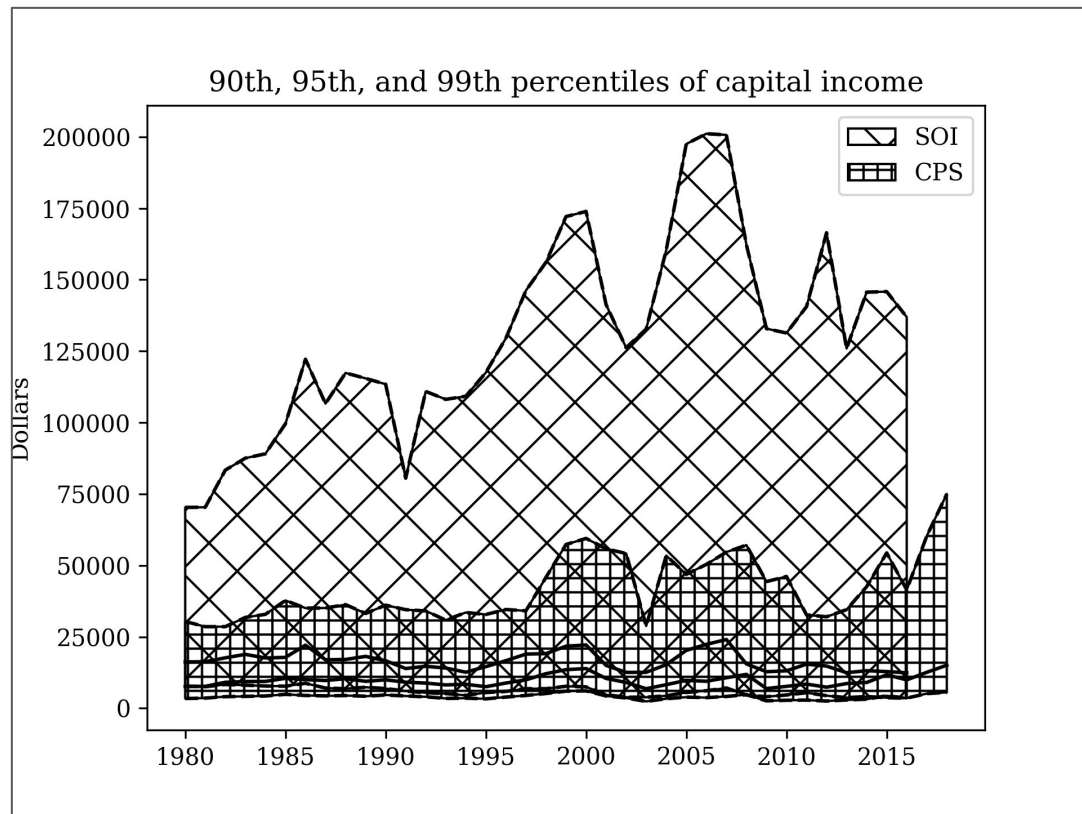
Findings VII (new): some more sanity-checking: quantiles

- **The SOI and CPS are identical, essentially, for quantiles of total income** (makes sense).
- Both series very skewed distributions and show very little growth in median or even 75th percentile income.



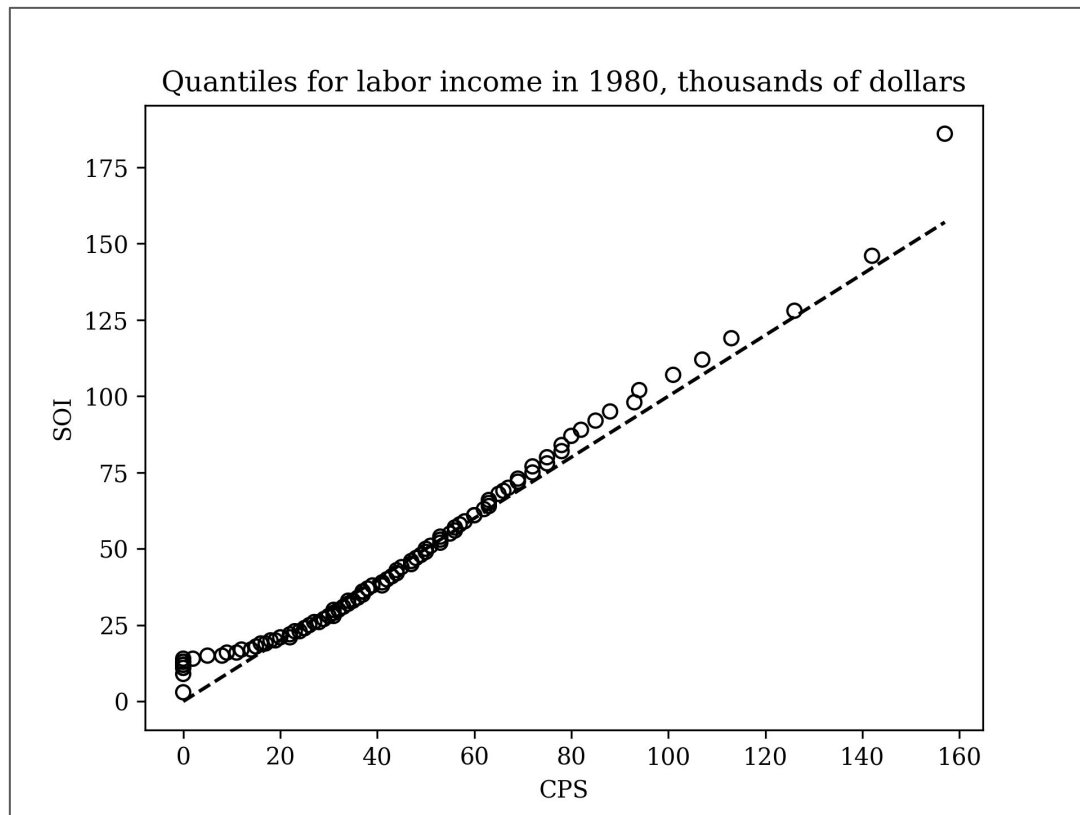
Findings VIII (new): some more sanity-checking: quantiles

- **For top-tier capital incomes, they're quite divergent, which is not super surprising.**
- It's really only at the very top, though, which would explain the large diverge in variances.

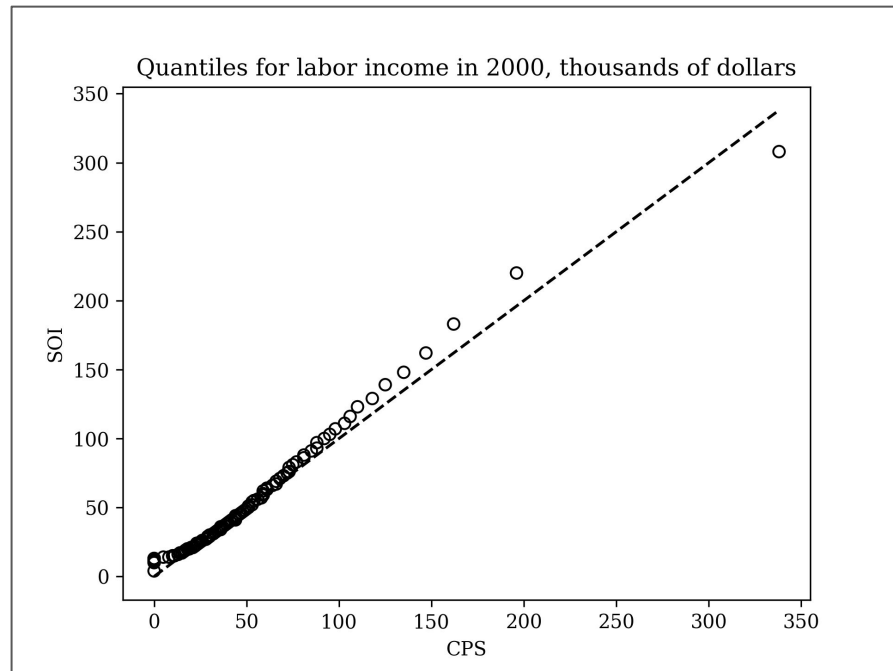
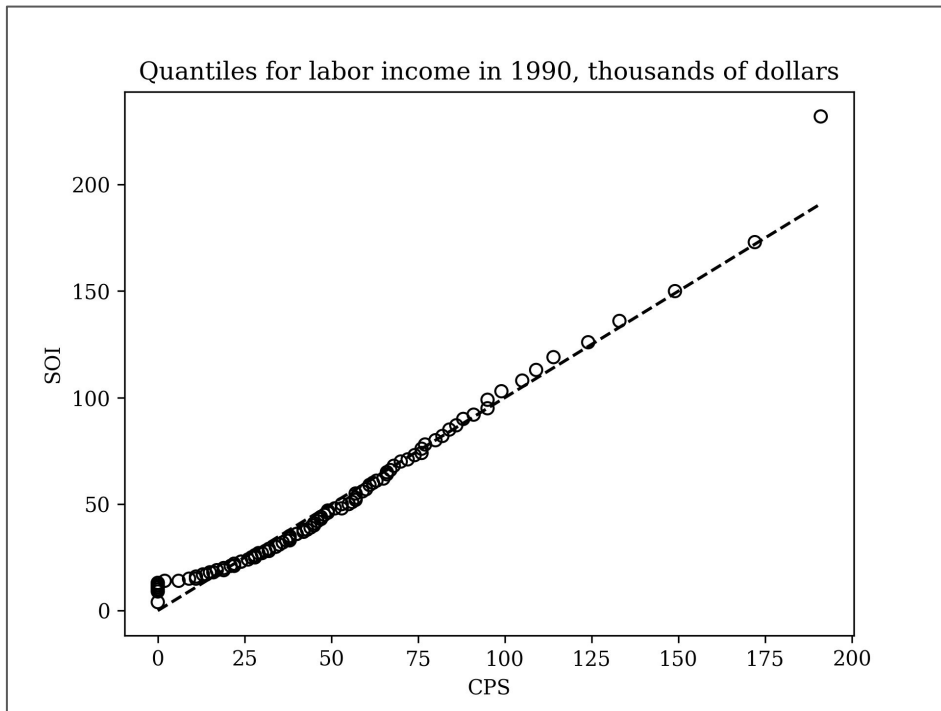


Findings IX (new): some more sanity-checking: quantiles

- Finally, for good measure, some **Q-Q plots to visualize these differences**.
- These really only make sense yearwise, so I picked a few typical years. First, labor income on several pages, then capital.

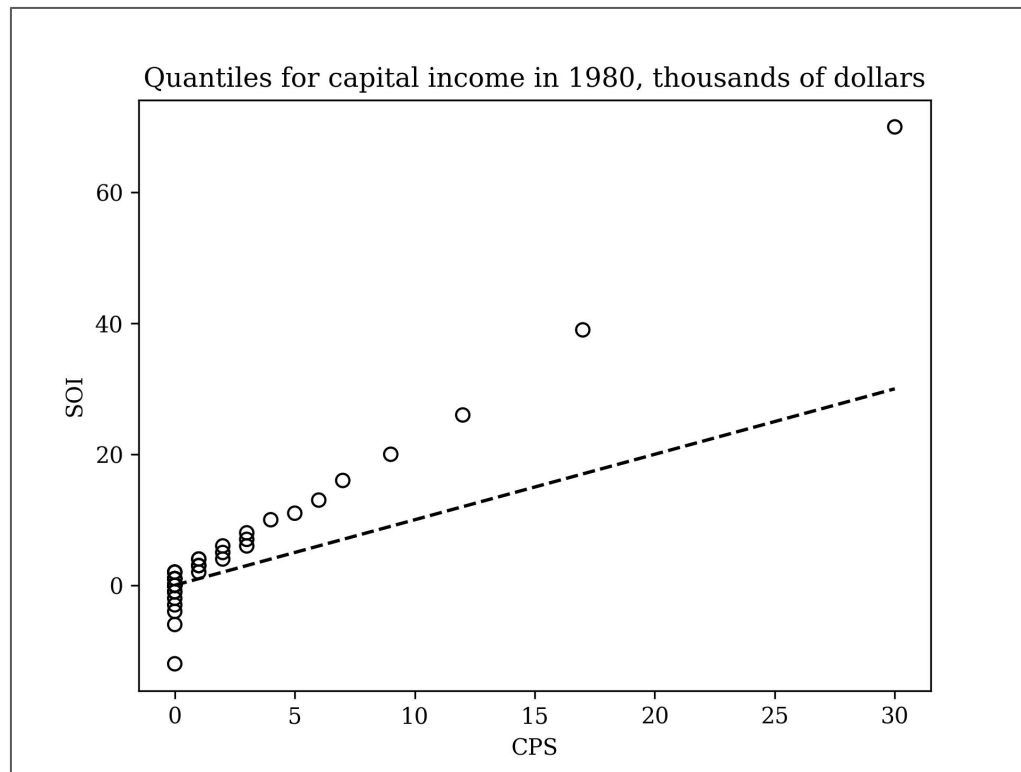


Findings X (new): some more sanity-checking: quantiles

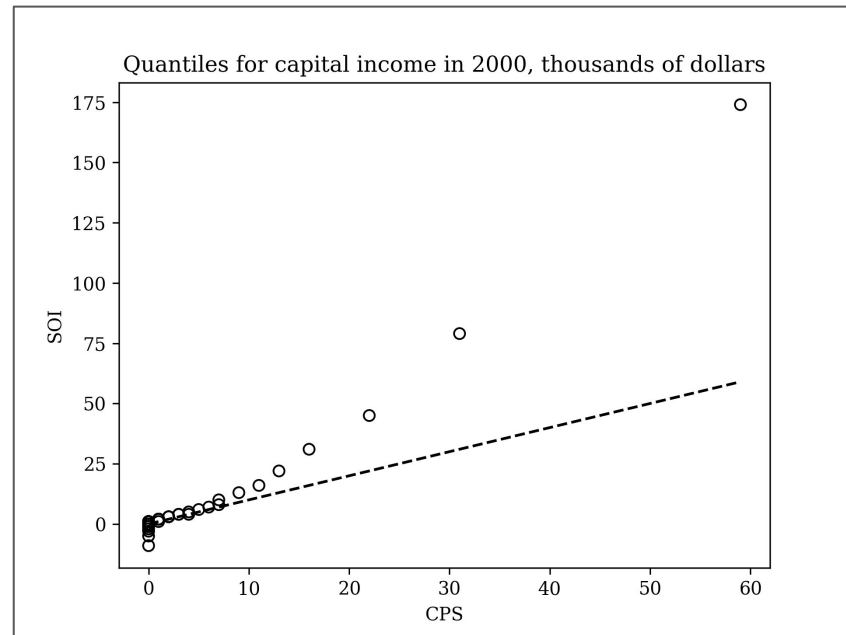
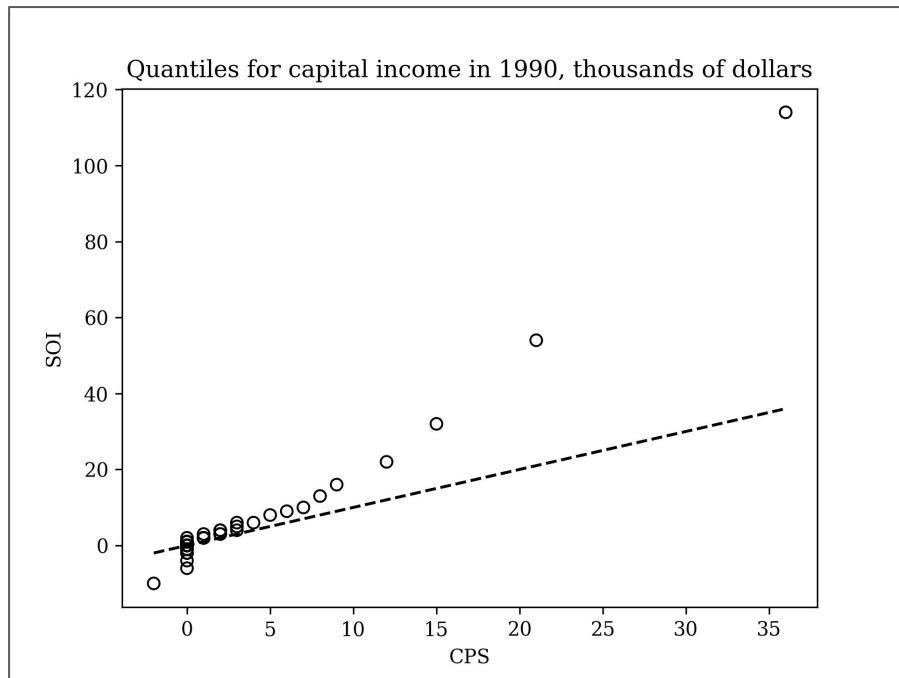


Findings XI (new): some more sanity-checking: quantiles

- **Here's capital income.**
It's obviously much more divergent (and does include some negative income).
- Incidentally, there are people with massively negative capital incomes I dropped.



Findings XII (new): some more sanity-checking: quantiles



Findings V: towards a greater emphasis on capital income

- If inequality is all professions, labor income – which includes labor income of very specialized workers; IRS forces you to split, e.g. partner income into capital/labor components – should play majority role.
- Not a typical causal argument: by definition, only factor incomes can contribute to variance in outcome. Not ruling out other ways of looking at it.
- **More indirect causality: the emphasis on professions and top labor incomes would mean that these results should not exist.** They do, so it's evidence that we should continue taking Marx seriously.