Measuring Trends in Trimmed Annual Earnings

David Pattison

Social Security Administration

FCSM, March 2018

Trimmed annual earnings data

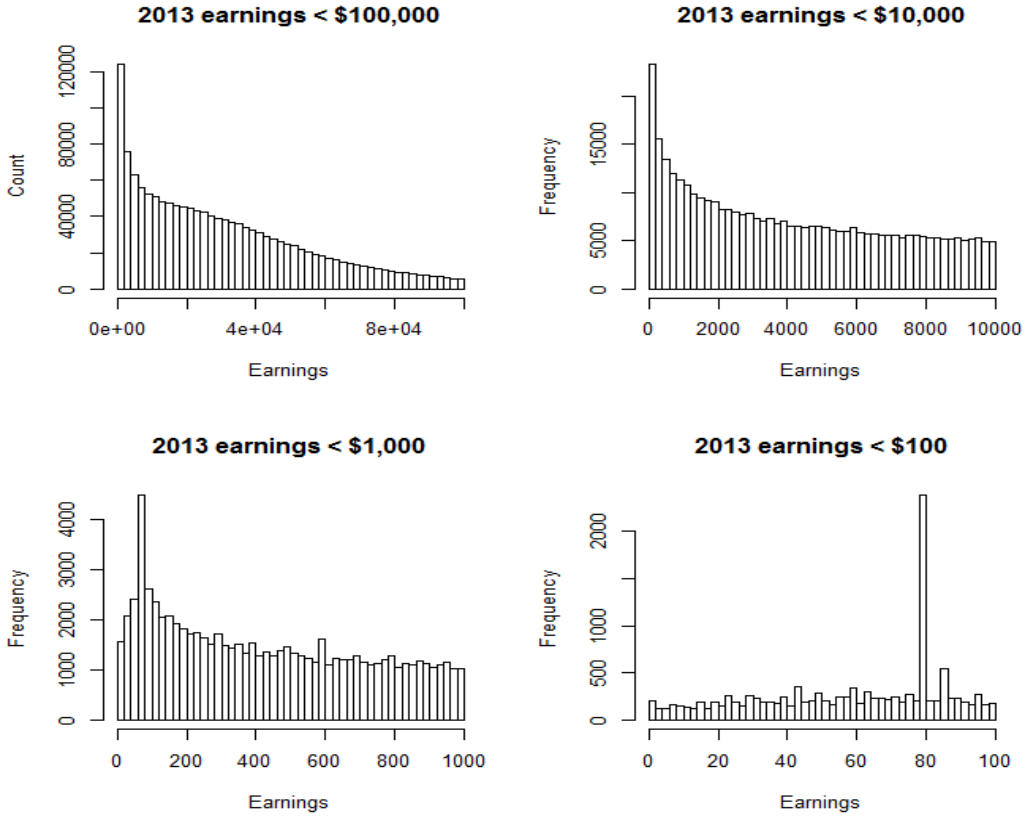
The study of changes in the earnings, both growth in average earnings and growth in the variance of earnings and changes in the dynamics of individual earnings, is important in Social Security analysis.

- Shape and scale of the earnings distribution
  - Social Security benefits and taxes rise in proportion with earnings only if the shape of the distribution does not change
- Earnings dynamics
  - The Social Security benefit formula was developed in part with an eye to the dynamics of individual earnings histories
  - Replacement rate definitions: how to deal with decline before retirement?
- The transition to retirement
  - Not all workers go to zero immediately. Many have low or sporadic earnings for many years
  - What is the threshold below which a worker can be considered retired?

Many earnings analyses first trim the lowest earnings in the distribution, removing them from the statistical sample. This study explores that practice.

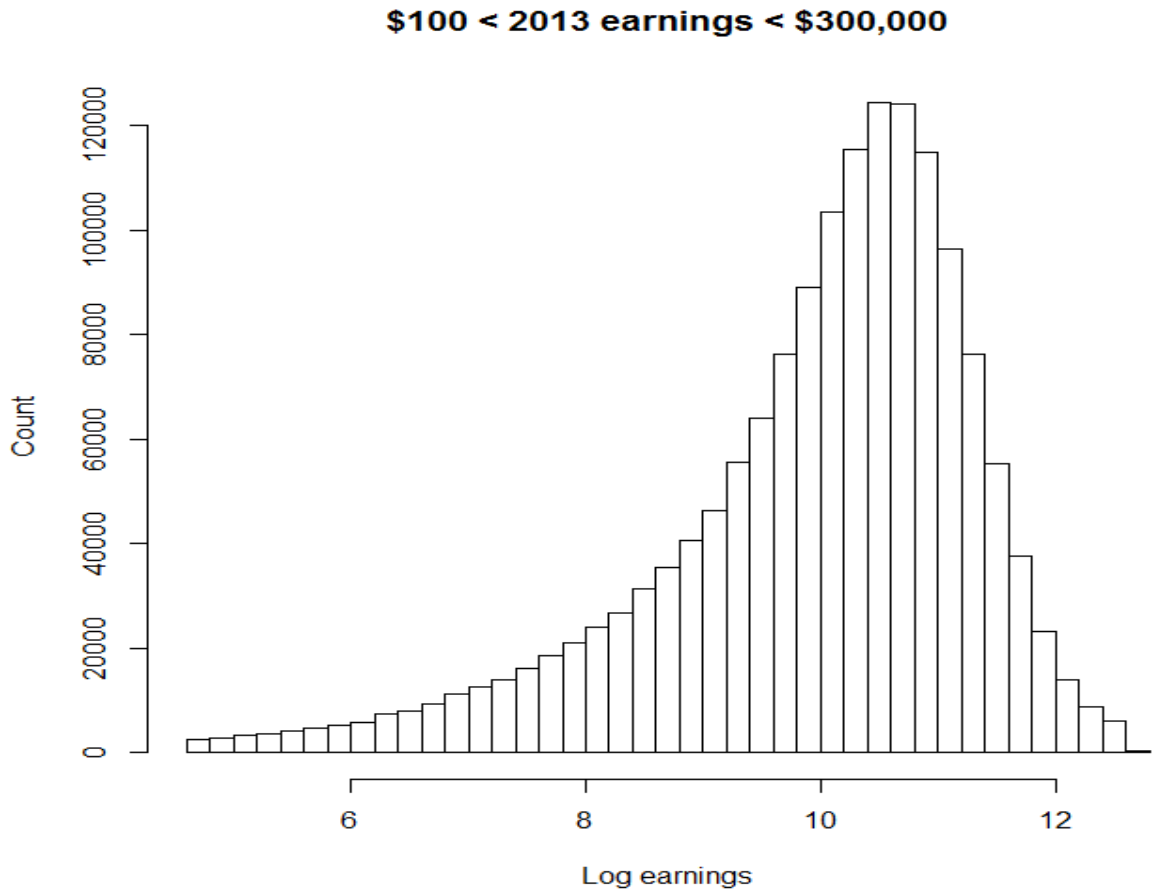
The problem of very low earnings

- The modal $100 earnings interval in the Social Security data is the bottom one
- Given this shape of the distribution, there is no easily defensible threshold for filtering out very low earnings



Log earnings

- The log earnings distribution is closer to the familiar bell shape
- But the long lower tail dominates variance measures



Why trim the lower earnings?

- Low earnings are not representative of most workers
- Lowest earnings intervals are noisy
  - From year to year, or from one year of age to the next, these extreme values add noise
- For studies using log earnings, the low earnings (and the noise) dominate the measures of variance
- For studies of the timing of retirement, there is a need to distinguish partial retirements

Trimming does not necessarily affect measured growth rates

- Trimming low earnings raises all the location measures (mean, median, percentiles), but as long as it raises them by the same proportion in every year, growth rates will not be affected
- For studies over long time spans, the threshold must be adjusted to keep pace with earnings distribution as it shifts upward
  - If the threshold is set with an index that grows more slowly than wages, (CPI, PCE), the location measures will also grow more slowly
  - If set with an index that grows faster than the median wage, location measures will also grow faster than the median wage
- In addition, the youngest and the oldest workers, with greater proportions of low earners, will be more affected by trimming
  - Even if earnings distribution rises proportionately at all ages, and even if threshold is accurately indexed to this rise, there can be an interaction with the changing age-sex composition

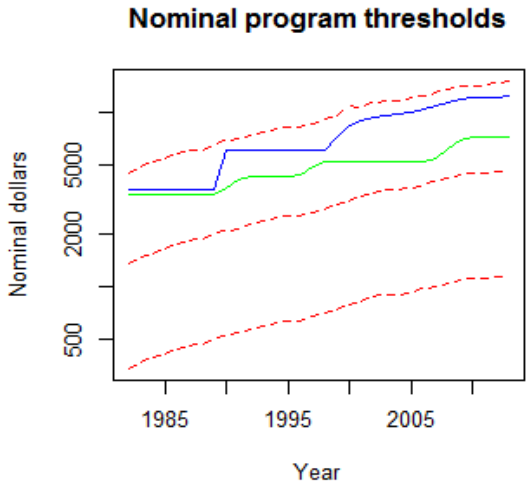
Trimming might also affect measured variances

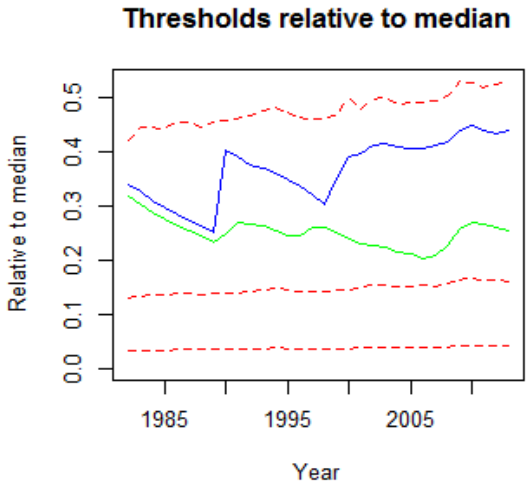
- Trimming low earnings *reduces* measured variance of log earnings
- Trimming low earnings also tends to reduce another common measure of variance: the ratio of 90th percentile to 10th percentile
  - Trimming raises the low percentiles more than the high percentiles, so the ratio of high to low declines
- As with location measures, the growth of variance is not necessarily affected
- The same issues arise: does the threshold indexing accurately reflect the rise in earnings, and is there an age-sex composition effect?
- Decomposition of earnings growth into variance within age-sex groups and variation between age-sex groups might be particularly sensitive to these issues

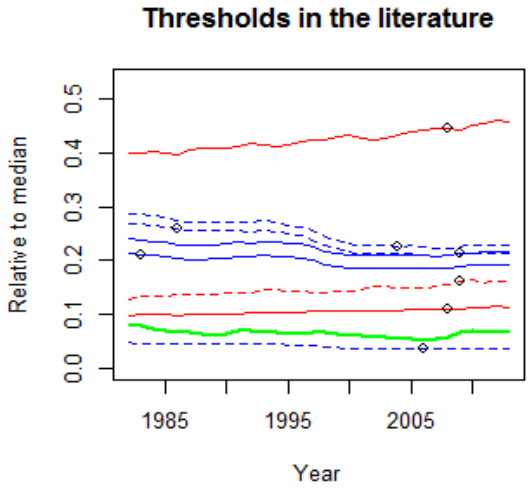
Possible thresholds

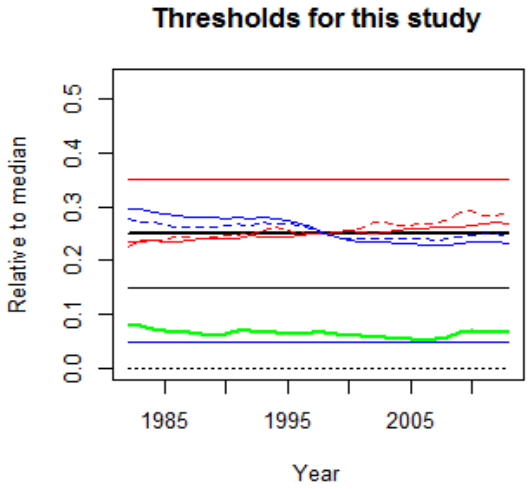
- Federal minimum wage (FMW)
  - ~20-30 percent of median earnings
- OASDI-defined thresholds
  - One quarter of coverage (QC1)
    - ~ 3 percent of median earnings
  - Four quarters of coverage (QC4)
    - ~12 percent of median
  - Retirement earnings test applicable at age 62 (RET62)
    - ~ 40 to 50 percent of median earnings
- None satisfactory in themselves
  - Often subject to legislative changes
  - OASDI thresholds are indexed to National Average Wage (NAW) index
    - Has been growing faster than the bottom half of the distribution
- Literature tends to refer to FMW
  - Level set by reference to FMW in a particular year
  - Then indexed: threshold over study span does not follow FMW

Thresholds









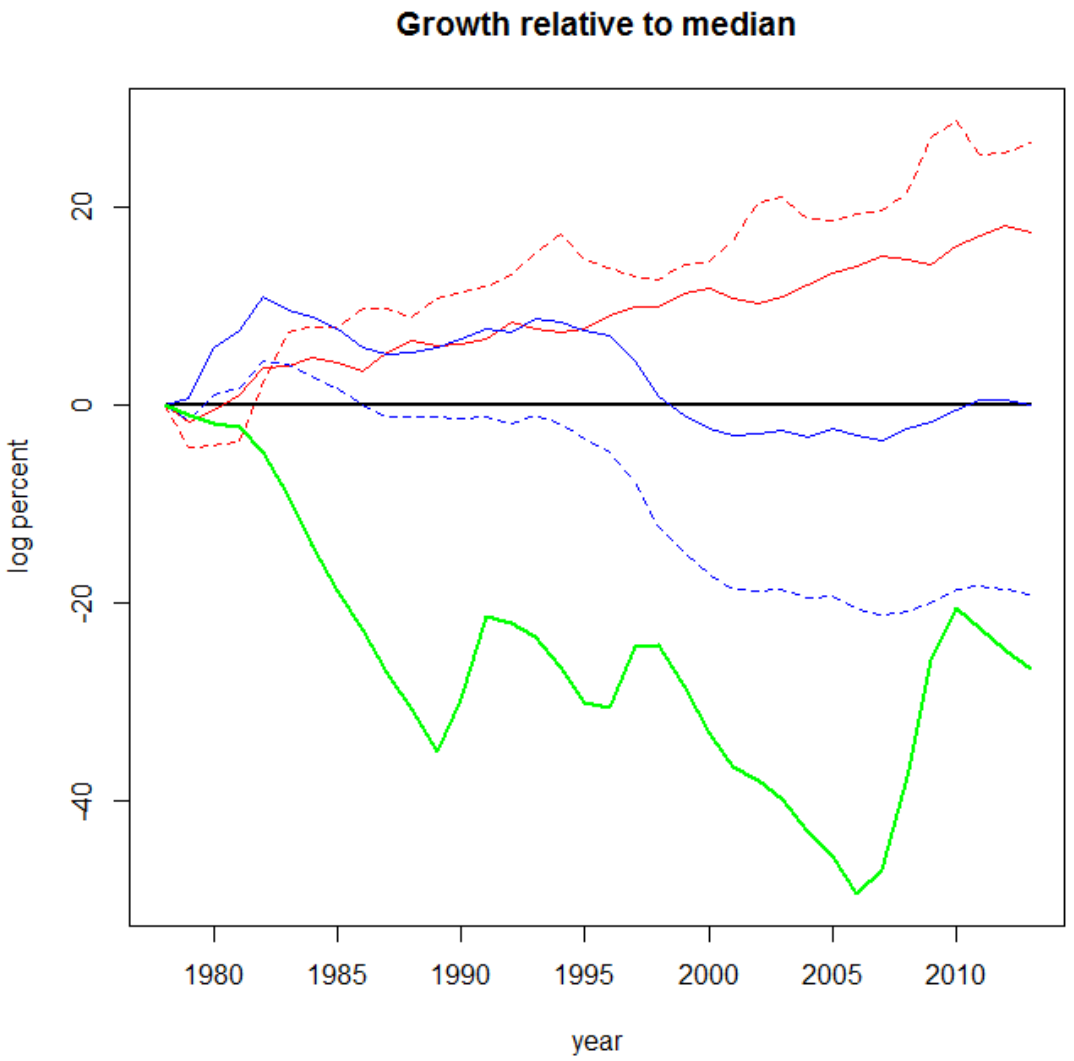
Top panels:
Dashed red, low to high:
QC1, QC4, RET62
Blue: SGA
Green: FMW

Bottom left panel
Index is color coded -
NAW red
Lagged NAW dashed red
CPI blue
PCE dashed blue
FMW (260 hours) green

Bottom right panel:
Level thresholds at 0 (dotted), 5 percent (blue), 15 percent (black), 25 percent (black) and 35 percent of median earnings.
Indexed thresholds based on 25 percent of median in 1994, indexed by CPI (blue), PCE (dashed blue), NAW (red), and lagged NAW (dashed red).
260 hours of FMW in green.

Growth in various indexes

NAW (red)
Lagged NAW (dashed red)
CPI (blue)
PCE (dashed blue
FMW (green)



Procedure

- Explore effects of a range of thresholds
  - Level: 0, 5, 15, 25, and 35 percent of median earnings
    - Index is effectively the growth in median wages
  - Index: CPI, PCE, NAW, NAW with two-year lag
    - Level set by 25 percent of median wage in 1998
- Data: “Detail” segment of Master Earnings File (MEF)
  - 1982-2013 wage and salary earnings plus deferred compensation
  - Slightly edited to remove implausible extreme amounts
  - 1% sample, ages 16-80, both sexes
    - 81 million person-year observations
    - 49 million positive earnings
    - 37 million above 35% of median
- For each threshold, restrict the sample each year to those with earnings above that year’s threshold
  - Tabulate mean, mean log earnings, percentiles, etc.
    - Whole sample, and by single-year age and sex groups
    - Age-sex tabulations allow analysis of changing composition

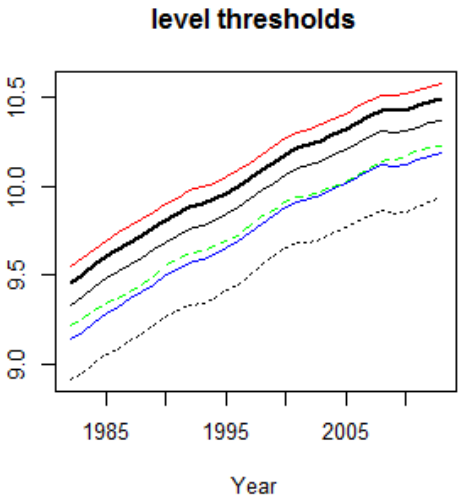
Results

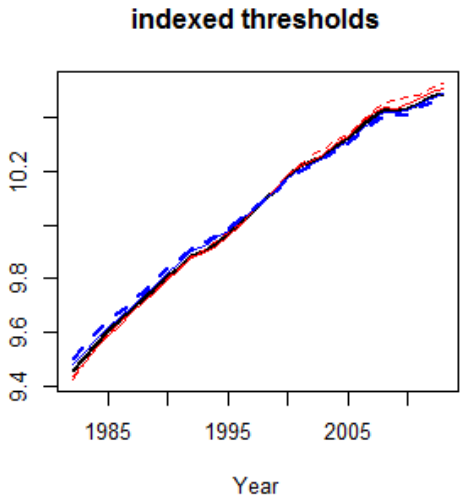
- For growth in mean log earnings:
  - Age-sex adjustment reduces the cumulative growth over the 32-year study span by somewhat more than 10 percentage points
  - Changing the level of the threshold has very little effect. Increasing from 0 percent of median to 35 percent of median added on the order of half a percent
  - Changing the index had larger effects, a range in the cumulative growth difference over 1 percent, but still smaller than the age-sex adjustment
- There were somewhat similar results for the employment rate, the percent of the workforce with earnings above the threshold
  - The zero threshold employment rate was slightly lower by 2013 than it had been in 1982.
  - Age-sex adjustment turns this into a slight increase
  - Changing the level of the threshold had smaller effects than age-sex adjustment
  - On the other hand, the choice of index had effects of the same magnitude as the age-sex adjustment

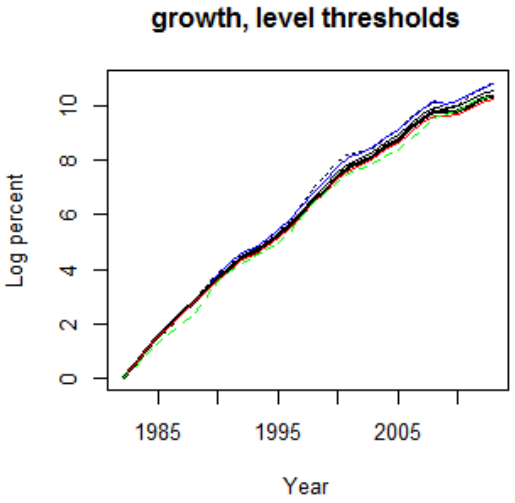
Effect of thresholds on mean log earnings

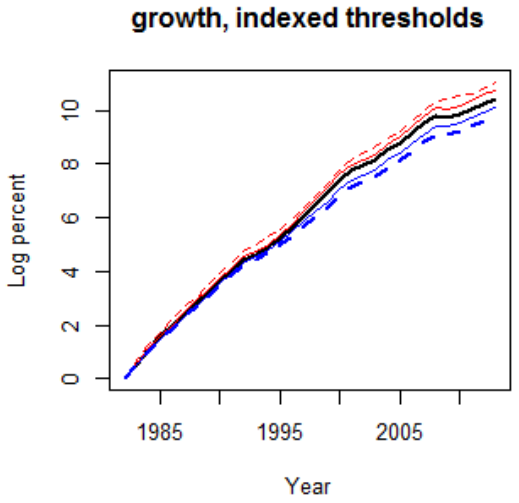
Level thresholds:
0 (dotted)
5% (blue)
15% (black)
25% (wide black)
35% red

Indexed thresholds:
NAW (red)
NAW lagged (dashed red)
CPI (blue)
PCE (dashed blue)





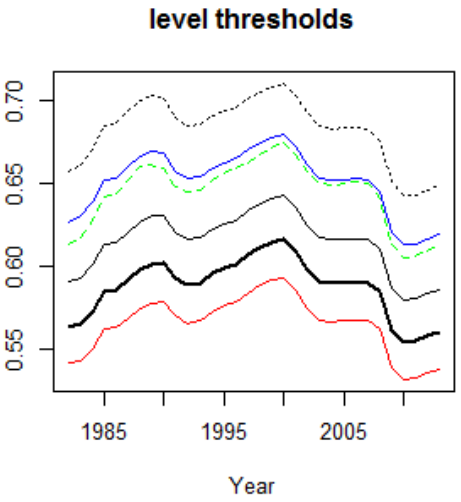


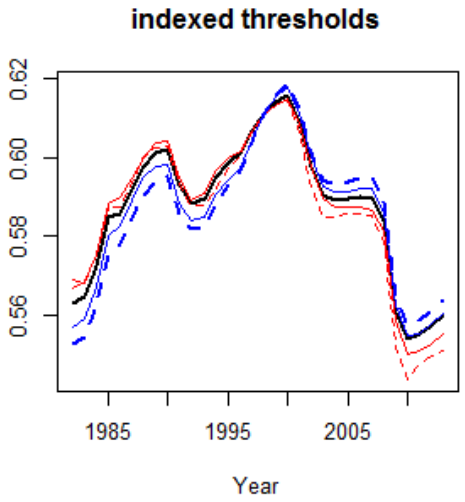


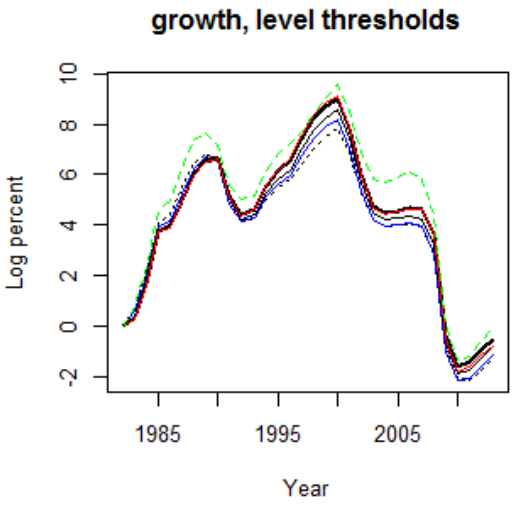
Effects of thresholds on the employment rate

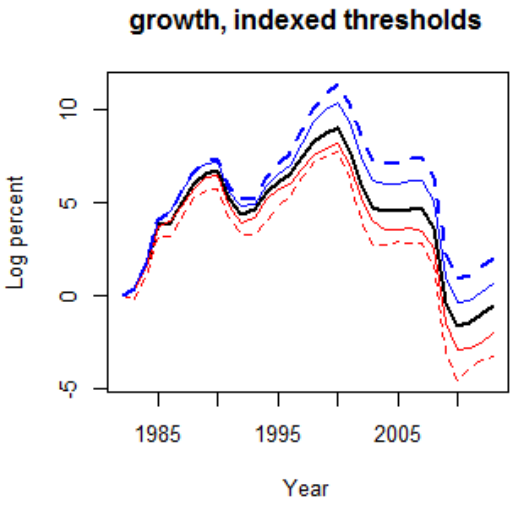
Level thresholds:
0 (dotted)
5% (blue)
15% (black)
25% (wide black)
35% red

Indexed thresholds:
NAW (red)
NAW lagged (dashed red)
CPI (blue)
PCE (dashed blue)









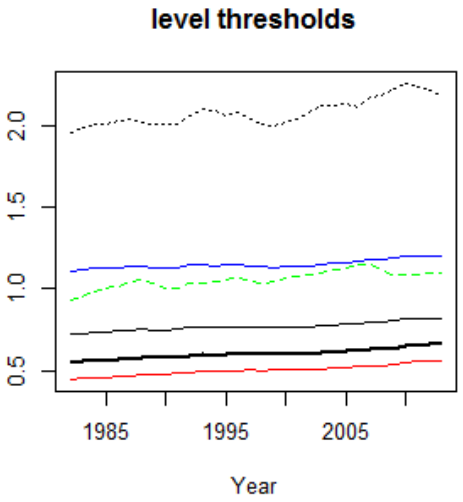
Results: Variance

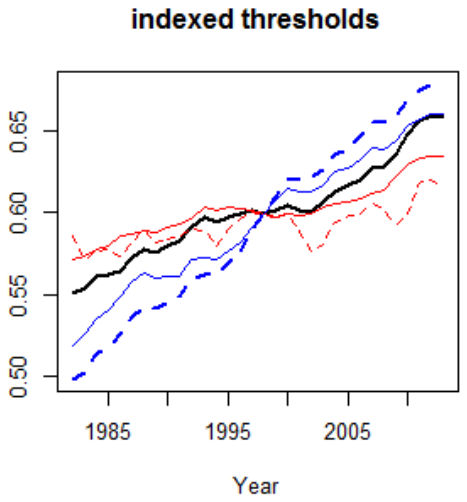
- For variance of log earnings:
  - The *level* of the threshold has large effects on the measured variance, with, as expected, high thresholds reducing the variance.
  - Higher thresholds also tend to be associated with faster growth in the variance, although there are exceptions
  - Indexes that grow more slowly than the median push up the rate of growth of the variance. This effect is not small. Variance using 25 percent of the median grew by about 15 percent over the study span. With a threshold indexed to the PCE, the variance grew by about 30 percent.
  - Faster growing indexes, similarly, reduce the growth in variance. With the lagged NAW index, the variance grew by less than 5 percent.
- The log ratio of the 90th percentile to the 10th percentile of earnings behaves similarly to the variance of log earnings
  - Log(p90/p10) appears to be slightly less sensitive to the threshold indexing rate

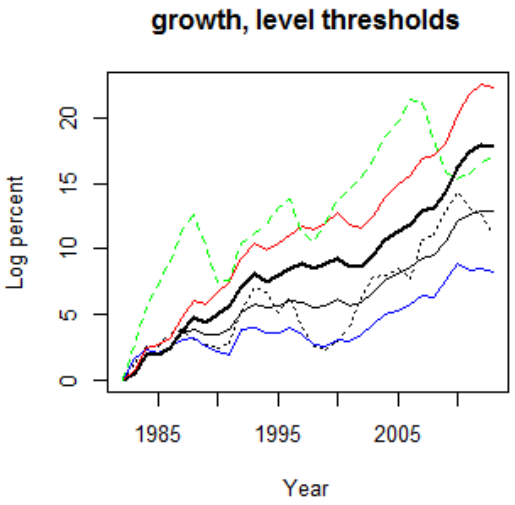
Effects of thresholds on variance of log earnings

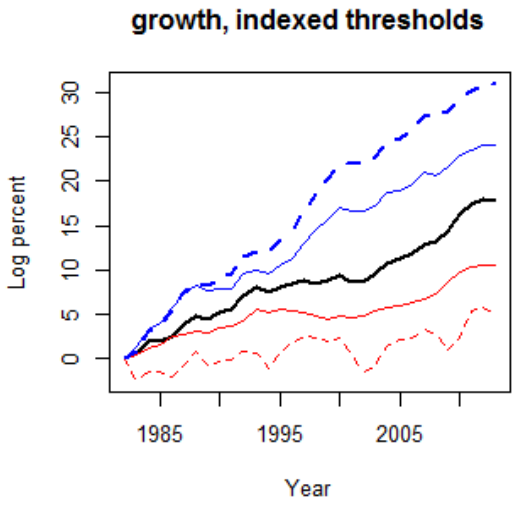
Level thresholds:
0 (dotted)
5% (blue)
15% (black)
25% (wide black)
35% red

Indexed thresholds:
NAW (red)
NAW lagged (dashed red)
CPI (blue)
PCE (dashed blue)









Results: Within and between components of variance

- For variance decomposed by age-sex groups into *within* and *between* components:
  - Most of the variance is *within* variance, which behaves similarly to the total variance
  - The much smaller *between* variance declines slightly over the study span, with higher thresholds or faster growing indexes magnifying the rate of decline, although the absolute decline remains small
  - Adjustment for changes in age-sex composition is problematic for within- and between decompositions, but the indication is that the constant composition within variance would increase slightly, rather then decline, with the growth in total variance also increasing.

Conclusions

- Measured trends in the growth in earnings appear to be relatively immune to the level and indexing trimming thresholds. It might be more important to check for the effects of changing age-sex composition.
- Measured trends in the variance of earnings, however, appear to be sensitive to the indexing and even the level of the thresholds.
- Some researchers carry out sensitivity tests for the *level* of the threshold. It is equally or more important to test the choice of index as well.