



July 2015

## Frank-Sommeiller-Price Series for Top Income Shares by US States since 1917

Mark Frank, Estelle Sommeiller, Mark Price, and Emmanuel Saez<sup>1</sup>

We have incorporated in the World Top Income Database the top income shares for each US state since 1917 using the series originally developed by Mark Frank, Estelle Sommeiller, and Mark Price. The series have been adjusted to be consistent with the Piketty and Saez US wide series for income including realized capital gains. All income values are reported in current dollars and need to be deflated using the CPI2014 variable (consumer price index, base 2014) to obtain real values. We plan on updating regularly these state series each year.

### Methodology

We have proceeded as follows starting from the Sommeiller-Price and the Frank series. These two sets of series were created independently.

The Sommeiller-Price series cover 1917-2012. They were originally created in Sommeiller's PhD dissertation (Sommeiller, 2006) and later updated and published in Sommeiller and Price (2014). The Sommeiller-Price series can be downloaded online at

<http://www.epi.org/publication/unequal-states/>

The Frank series initially covered the period 1945-2004 and were published in Frank (2009). Frank (2014) later updated the series to 1916-2012. The Frank series can be downloaded online at

[http://www.shsu.edu/eco\\_mwf/inequality.html](http://www.shsu.edu/eco_mwf/inequality.html)

We start from the Sommeiller-Price series since 1917 because they provide more percentiles than the Frank series.<sup>2</sup> The Sommeiller-Frank series also follow more closely the Piketty-Saez methodology before 1942, when only a fairly small fraction of families filed income tax returns.<sup>3</sup> After 1941, when the vast majority of families file tax returns, the Sommeiller-Price series and the Frank series are fairly close (see below).

---

<sup>1</sup> The original state level series were created by Mark Frank, Estelle Sommeiller, and Mark Price. Emmanuel Saez just helped coordinate the minor adjustments to the original Frank and Sommeiller-Price series, needed to make them comparable to the Piketty-Saez series. We thank Yang Chen for outstanding research assistance in preparing all the series for posting on the WTID starting from the original Frank and Sommeiller-Price series.

<sup>2</sup> Frank series include the top 10% and top 1% income shares, as well as a number of other inequality indexes such as the Gini, Theil, or Atkinson indexes.

<sup>3</sup> The Sommeiller-Price series take into account non-filers. The Frank series compute top income shares within the set of tax filers.

**Piketty-Saez series benchmarking.** The Sommeiller-Price series still differ from the Piketty-Saez series because not all the adjustments carried out by Piketty-Saez can be done at the state level (due to lack of information on the composition of incomes, and in particular realized capital gains, by brackets in the state level statistics).<sup>4</sup> Fortunately, Sommeiller-Price also provide US wide series that are created using the same methodology that they apply to each state. Hence, comparing the US wide Sommeiller-Price series to the Piketty-Saez series allows us to assess the discrepancies that arise because of the difference in methodologies. Fortunately, the difference between these two series is fairly modest and both series have the same overall U-shape over the century. Nevertheless, to correct for this difference, we simply re-adjust each state series in each year by the ratio of the Piketty-Saez series (series including realized capital gains) to the US wide Sommeiller-Frank series for the corresponding year. For example, if the US wide Sommeiller-Frank series has a top 1% income share of 20% in year t and the Piketty-Saez series has a top 1% income share of 22% in the year t, we multiply the top 1% state level income shares by  $22/20=1.1$  in year t for all states. As an additional check, we have also applied the same correction strategy to the Frank series (using Frank US wide series). The correct Sommeiller-Price series and the corrected Frank series are pretty close for the post 1941 period.<sup>5</sup>

**Imputations for 1983-1985.** Next, the Sommeiller-Price series do not cover years 1983-1985 but the Frank series include the years 1983-1985 thanks to unpublished tabulations that Frank obtained directly from the Statistics of Income division of the IRS. Therefore, we use the Frank series to impute 1983-1985 values so as to have a complete annual series from 1917 to 2012.

**Wyoming 2012 typo.** Finally, as statistics for the State of Wyoming for the top \$1m+ bracket have a typo in 2012 (the average income in this bracket is over \$10m, which a totally implausible value leading to an absurdly high top 1% income share that year), we correct for this typo by assuming that the average income in top \$1m+ in 2012 in Wyoming is equal to its 2011 value adjusted for the US wide growth in the average income in the top \$1m+ bracket from 2011 to 2012.

---

<sup>4</sup> In particular, the micro-level annual public use files available since 1960 do not have state information for high income returns. Hence, they cannot be used for the estimation of state level top income shares. Therefore, all the estimations have to rely on state level tabulations created by the Statistics of Income of the IRS using internal data. Such tabulations for recent years are posted at <http://www.irs.gov/uac/SOI-Tax-Stats-Historic-Table-2>

<sup>5</sup> Some significant differences arise between the two corrected series in the period 2002-2009 when the IRS Statistics of Income only published State level income tax statistics up to a top AGI bracket of \$200,000 and above. Before 2002 and after 2009, the top bracket is \$1,000,000 and above. See Internal Revenue Service, Statistics of Income Division, "Individual Income and Tax Data, by State and Size of Adjusted Gross Income", online at <http://www.irs.gov/uac/SOI-Tax-Stats-Historic-Table-2> Frank uses a split sample imputation while Sommeiller-Price use a Pareto interpolation, explaining the differences in the two series for these years when the top threshold of \$200,000 is below the top 1% threshold. For imputations above the top bracket, the Pareto interpolation is preferable and hence we decided to follow the Sommeiller-Price series.

## **Dataset description**

We provide the dataset in two versions.

Version 1: A Stata format file (US\_state\_level.dta) containing the following variables:

Year (year), State (state in full name), st (state ID from 1 to 51 (0 for the United States), alphabetical order), Top10\_adj, Top5\_adj, Top1\_adj, Top05\_adj, Top01\_adj, Top001\_adj (top income shares for the top 10%, 5%, 1%, .5%, .1%, .01% respectively, including capital gains, expressed in percent), N\_TaxReturn (number of tax returns filed in the state) in thousand, N\_TaxUnit (number of tax units in the State, if everybody had been required to file a tax return) in thousand, AvgInc (average income per tax unit, current US Dollars), TotalInc (total income including imputed income of non-filers, current million US Dollars), AGI (total Adjusted Gross Income reported by tax filers, current million US Dollars), CPI14 (Consumer price index, CPI-U-RS series, base 1 in 2014). From these variables, it is straightforward to recover real values for average income in each group.

The variables are provided for each state from 1917 to 2012, as well as for the full United States (Piketty-Saez series including realized capital gains).

Version 2: An Excel format file (US\_state\_level.xlsx) containing the variables described above (which correspond both to the country and state levels), as well as all the other variables available in the *World Top Incomes Database* at the country level.

## References

Frank, Mark. W. 2009 "Inequality and Growth in the United States: Evidence from a New State-Level Panel of Income Inequality Measure" *Economic Inquiry*, Volume 47, Issue 1, Pages 55-68.

Frank, Mark. W. 2014 "A New State-Level Panel of Annual Inequality Measures over the Period 1916 - 2005" *Journal of Business Strategies*, vol. 31, no. 1, pages 241-263.

Internal Revenue Service, Statistics of Income Division, "Individual Income and Tax Data, by State and Size of Adjusted Gross Income", online at <http://www.irs.gov/uac/SOI-Tax-Stats-Historic-Table-2>

Piketty, Thomas and Emmanuel Saez "Income Inequality in the United States, 1913-1998", *Quarterly Journal of Economics*, 118(1), 2003, 1-39, series updated to 2014 in June 2015.

Sommeiller, Estelle. *Regional Income Inequality in the United States, 1913–2003*. PhD. dissertation, University of Delaware, 2006.

Sommeiller, Estelle, and Mark Price "The Increasingly Unequal States of America Income Inequality by State, 1917 to 2011", *Economic Analysis and Research Network Report*, February 2014.