

Formulas for Individual-Level Variables in the SYZZ Micro Files that Reproduce PSZ TB2f

Piketty-Saez-Zucman's (2018, hereafter "PSZ") Online Appendix Table TB2f disaggregates top 1% income into nine categories: Compensation of employees, Labor component of mixed income, Labor component of pension income, Capital component of mixed income, Capital component of pension income, S-corporation dividends, C-corporation dividends, C-corporation retained earnings, and Interest and rents. Our analysis of top incomes in Imputed National Income requires individual-level variables that sum to the nine TB2f top 1% categories. However, PSZ's micro files (which they kindly shared) do not contain these nine variables at the individual level. Instead, PSZ use aggregations of their micro-file variables in order to create the TB2f aggregates. We follow their code in order to create the nine variables at the individual level.

This PDF documents our formulas and their fidelity to PSZ's TB2f top 1% aggregates. We closely match the TB2f aggregates in all years. Two deviations are notable. First, we find more top 1% S-corporation dividends 2011-2014 (and correspondingly less C-corporation dividends and C-corporation retained earnings in those years) than PSZ because we use actual wealth aggregates rather than PSZ's extrapolations (see our Online Appendix D). Second, we obtain slightly more Interest and rents and correspondingly less C-corporation dividends and C-corporation retained earnings than do PSZ. That discrepancy has no bearing on our main findings, as we treat all three of those categories as financial capital income.

Our formulas use variable names (e.g. "ptinc" and "peinc") from PSZ's micro files. We use the following subscripts:

- "tot": "aggregate sum".
- "top1pt": "top 1% sum, defining top 1% using ptinc-based rank".
- "top1pe": "top 1% sum, defining top 1% using peinc-based rank".
- "pszAgg": "comes from PSZ's published appendix data file DINA(Aggreg)_QJE.xlsx" rather than from PSZ's micro files.

Relevant Stata file: build_usdinasyzz_20190719.do. The relevant output files are named usdinaYYYYsyzz_es1.dta.

Names of the individual-level variables we create, where i denotes an individual in the PSZ micro files:

- $empcomp_i$: Compensation of employees
- $lmixinc_i$: Labor component of mixed income
- $lpeninc_i$: Labor component of pension income

- $kmixinc_i$: Capital component of mixed income
- $kpeninc_i$: Capital component of pension income
- $sdiv_i$: S-corporation dividends
- $cdiv_i$: C-corporation dividends
- cre_i : C-corporation retained earnings
- $intrents_i$: Interest and rents
- $labinc_i = empcomp_i + lmixinc_i + lpeninc_i$
- $capinc_i = kmixinc_i + kpeninc_i + sdiv_i + cdiv_i + cre_i + intrents_i$
- $allinc_i = labinc_i + capinc_i$

Formulas for the individual-level variables. Note that $sdiv'_i$, $cdiv'_i$, and cre'_i are intermediate variables.

$$empcomp_i = \left[flemp_i + flprl_i \times \left(\frac{flemp_i}{flemp_i + flmil_i} \right) + plcon_i \times \left(\frac{flemp_{top1pt}}{flemp_{top1pt} + flmil_{top1pt}} \right) \right] \times \frac{ptinc_{top1pe}}{ptinc_{top1pt}}$$

$$lmixinc_i = \left[flmil_i + flprl_i \times \left(\frac{flmil_i}{flemp_i + flmil_i} \right) + plcon_i \times \left(\frac{flmil_{top1pt}}{flemp_{top1pt} + flmil_{top1pt}} \right) \right] \times \frac{ptinc_{top1pe}}{ptinc_{top1pt}}$$

$$lpeninc_i = \left[plben_i \times \frac{ptinc_{top1pe}}{ptinc_{top1pt}} + prisupen_i \times \frac{prisupen_{top1pe}}{prisupen_{top1pt}} + invpen_i \times \frac{invpen_{top1pe}}{invpen_{top1pt}} \right] \times \frac{socinlabshare_{pszAgg}}{socinlabshare_{pszAgg} + socinshare_{pszAgg}}$$

$$kmixinc_i = fkbus_i \times \frac{ptinc_{top1pe}}{ptinc_{top1pt}}$$

$$kpeninc_i = \left[plben_i \times \frac{ptinc_{top1pe}}{ptinc_{top1pt}} + prisupen_i \times \frac{prisupen_{top1pe}}{prisupen_{top1pt}} + invpen_i \times \frac{invpen_{top1pe}}{invpen_{top1pt}} \right] \times \frac{socinshare_{pszAgg}}{socinlabshare_{pszAgg} + socinshare_{pszAgg}}$$

$$sdiv'_i = scorpinc_i \times \frac{nipadiv_{pszAgg}}{scorpinc2_{tot}}$$

$$cdiv'_i = \left(\frac{divinc_i + divest_i + kginc_i + kgest_i}{divinc_{tot} + divest_{tot} + kginc_{tot} + kgest_{tot}} \right) \\ \times \left(\frac{ttdiv_{pszAgg}}{corpequitywealth_{pszAgg} + corppensionwealth_{pszAgg} - Sequitywealth_{pszAgg}} \right) \\ \times (netcorpdiv_{pszAgg} - nipasdiv_{pszAgg} + netforeigndiv_{pszAgg})$$

$$cre'_i = \left(\frac{divinc_i + divest_i + kginc_i + kgest_i}{divinc_{tot} + divest_{tot} + kginc_{tot} + kgest_{tot}} \right) \\ \times \left(\frac{ttdiv_{pszAgg}}{corpequitywealth_{pszAgg} + corppensionwealth_{pszAgg} - Sequitywealth_{pszAgg}} \right) \\ \times (corpretaineearnings_{pszAgg})$$

$$\begin{aligned}
& \times \left(\frac{scorpinc2_{top1pt}}{scorpinc2_{tot}} \times \frac{ttscorw_{pszAgg}}{ttdivw_{pszAgg} + ttscorw_{pszAgg}} + \frac{divinc_{top1pt} + divest_{top1pt} + kginc_{top1pt} + kgest_{top1pt}}{divinc_{tot} + divest_{tot} + kginc_{tot} + kgest_{tot}} \times \frac{ttdivw_{pszAgg}}{ttdivw_{pszAgg} + ttscorw_{pszAgg}} \right) \\
& \quad \times \left(\frac{sdiv_i = sdiv'_i}{sdiv'_{top1pt} + cdiv'_{top1pt} + cre'_{top1pt}} \right) \\
& \quad \times \left(\frac{ttdivw_{pszAgg} + ttscorw_{pszAgg}}{ttdivw_{pszAgg} + ttscorw_{pszAgg} + corppensionwealth_{pszAgg}} \right) \\
& \quad \times (divshareofPFI_{pszAgg} + reshareofPFI_{pszAgg} + corpbizproptaxshareofPFI_{pszAgg}) \times personalfactorinc_{pszAgg}
\end{aligned}$$

$$\begin{aligned}
& \times \left(\frac{scorpinc2_{top1pt}}{scorpinc2_{tot}} \times \frac{ttscorw_{pszAgg}}{ttdivw_{pszAgg} + ttscorw_{pszAgg}} + \frac{divinc_{top1pt} + divest_{top1pt} + kginc_{top1pt} + kgest_{top1pt}}{divinc_{tot} + divest_{tot} + kginc_{tot} + kgest_{tot}} \times \frac{ttdivw_{pszAgg}}{ttdivw_{pszAgg} + ttscorw_{pszAgg}} \right) \\
& \quad \times \left(\frac{cdiv_i = cdiv'_i}{sdiv'_{top1pt} + cdiv'_{top1pt} + cre'_{top1pt}} \right) \\
& \quad \times \left(\frac{ttdivw_{pszAgg} + ttscorw_{pszAgg}}{ttdivw_{pszAgg} + ttscorw_{pszAgg} + corppensionwealth_{pszAgg}} \right) \\
& \quad \times (divshareofPFI_{pszAgg} + reshareofPFI_{pszAgg} + corpbizproptaxshareofPFI_{pszAgg}) \times personalfactorinc_{pszAgg}
\end{aligned}$$

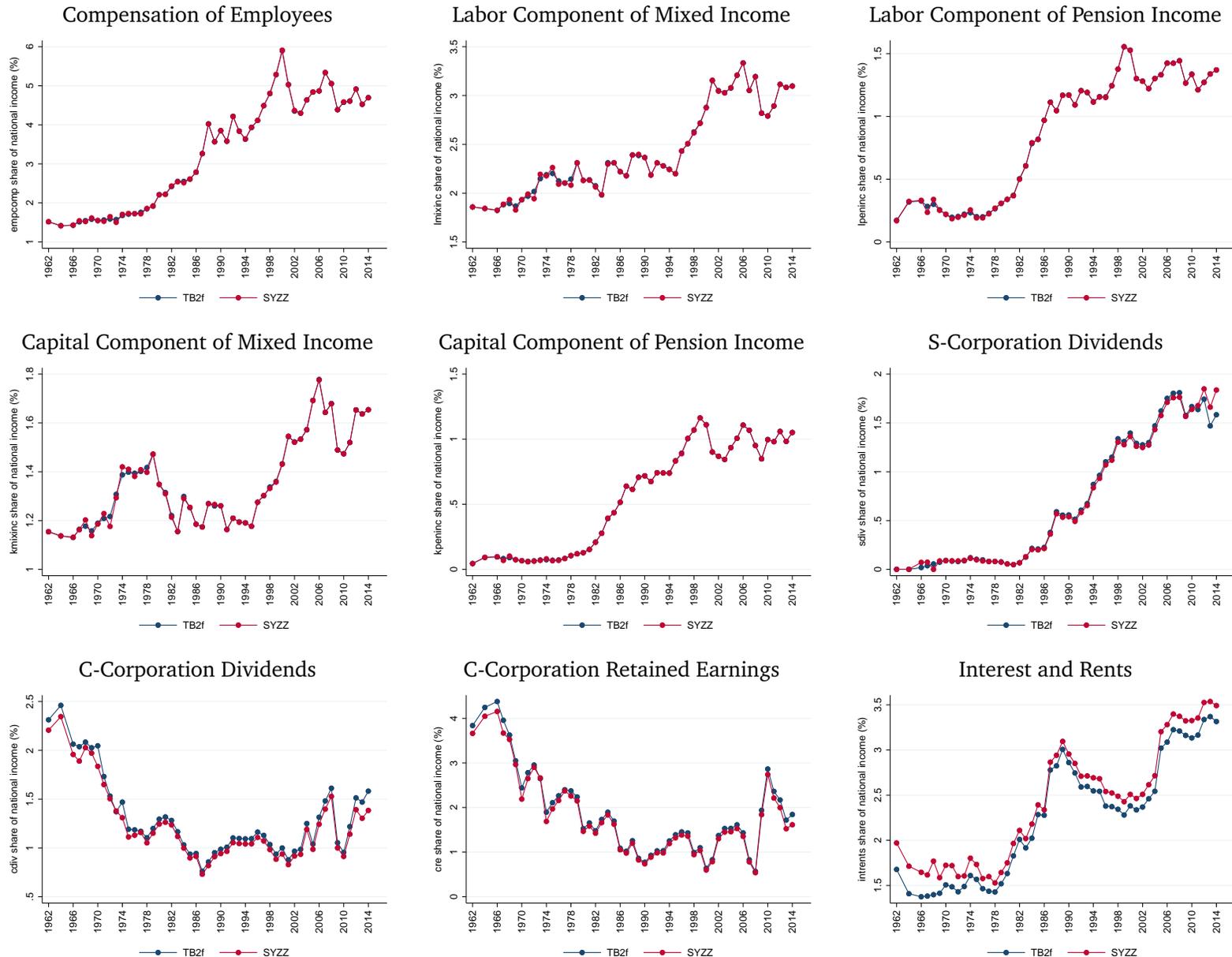
$$\begin{aligned}
& cre_i = cre'_i \\
& \times \left(\frac{1}{sdiv'_{top1pt} + cdiv'_{top1pt} + cre'_{top1pt}} \right) \\
& \times \left(\frac{scorpinc2_{top1pt}}{scorpinc2_{tot}} \times \frac{ttscorw_{pszAgg}}{ttdivw_{pszAgg} + ttscorw_{pszAgg}} + \frac{divinc_{top1pt} + divest_{top1pt} + kginc_{top1pt} + kgest_{top1pt}}{divinc_{tot} + divest_{tot} + kginc_{tot} + kgest_{tot}} \times \frac{ttdivw_{pszAgg}}{ttdivw_{pszAgg} + ttscorw_{pszAgg}} \right) \\
& \times \left(\frac{ttdivw_{pszAgg} + ttscorw_{pszAgg}}{ttdivw_{pszAgg} + ttscorw_{pszAgg} + corppensionwealth_{pszAgg}} \right) \\
& \times (divshareofPFI_{pszAgg} + reshareofPFI_{pszAgg} + corpbizproptaxshareofPFI_{pszAgg}) \times personalfactorinc_{pszAgg}
\end{aligned}$$

$$allinc_i = peinc_i \times \frac{peinc_{top1pt}}{peinc_{tot}}$$

$$intrents_i = allinc_i - empcomp_i - lmixinc_i - lpeninc_i - kmixinc_i - kpeninc_i - sdiv_i - cdiv_i - cre_i$$

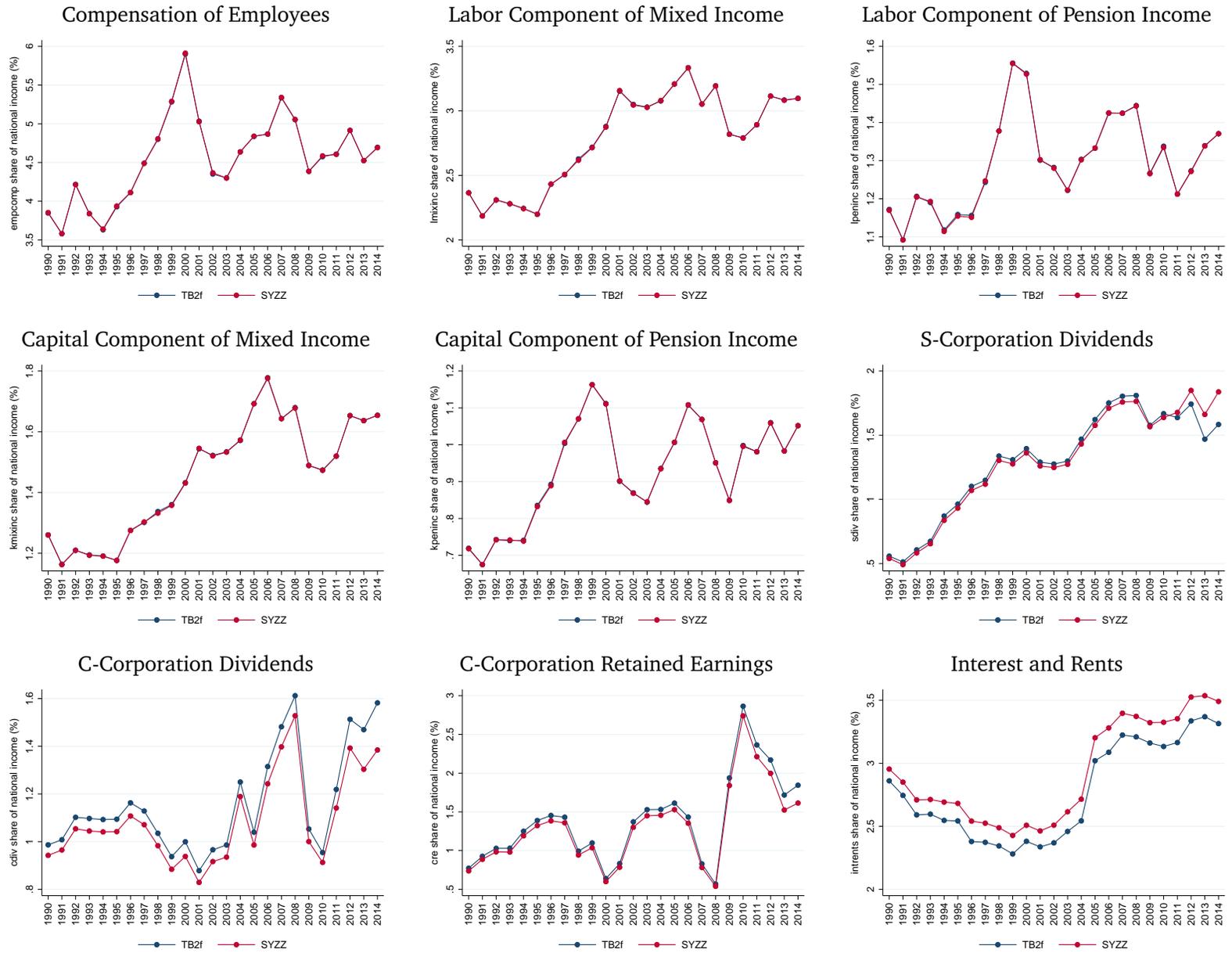
$$\text{Top 1\% } var_i \text{ share of NI} = \frac{var_{top1pt}}{peinc_{tot}}$$

Figure 1: Comparing Top 1% Shares from SYZZ Micro Files to PSZ's Top 1% Shares in TB2f - 1962-2014



Note: These graphs aggregate the individual-level variables named in the y-title from the usdinaYYYYsyzz_es1.dta files. The graphs are created by -reproduce_tb2f in draft_usdinasyzz_20190719.do.

Figure 2: Comparing Top 1% Shares from SYZZ Micro Files to PSZ's Top 1% Shares in TB2f - 1990-2014



Note: These graphs aggregate the individual-level variables named in the y-title from the usdinaYYYYsyzz_es1.dta files. The graphs are created by -reproduce_tb2f- in draft_usdinasyzz_20190719.do.