Data Appendix

Table 1:

Annual time series for real group income per tax unit for the top 1% and the three subgroups of the top 1% are based on our calculations using IRS-SOI income data as compiled by Piketty and Saez (2010) (http://elsa.berkeley.edu/~saez/TabFig2008.xls, extending Piketty and Saez (2003)). We obtain aggregate US income pre-tax, pre-transfer and excluding capital gains from 1929 forward from NIPA Table 2.1 as the sum of wage and salary disbursements (line 3), proprietors' income (line 9), personal dividend income (line 15), personal interest income (line 14), and rental income of persons (line 12). Prior to 1929 we obtain aggregate US income pre-tax, pre-transfer and excluding capital gains from Piketty and Saez's data set which for this period is also based on national accounts (we calculate growth rates in each source separately and use the NIPA growth rate from 1930 on and the Piketty and Saez growth rate for 1929 and before). We translate the aggregate US income to real terms and to per tax unit values using the price deflator and number of tax units from Piketty and Saez.

The unit of analysis is a tax unit.

Table 2:

Sources for group incomes and aggregate income for all tax units are as for Table 1.

Ratios reported in Panel B are calculated based on Piketty and Saez's income shares (as group income share/group size).

Median real household income used in the bottom panel is from the Current Population Survey and obtained from http://www.census.gov/hhes/www/income/data/historical/inequality/taba2.pdf

The unemployment rate used in the bottom panel is from the Current Population Survey, for males age 25-54, obtained from www.bls.gov, series LNS14000061.

Table 3, Figure 2, Figure 3, Figure 4:

Sources for group incomes and aggregate income for all tax units, as well as for income composition, are as for Table 1.

Wages and salaries include pensions, bonuses and stock-options exercises. Entrepreneurial income is defined as profits from S-Corporations plus profits from Partnerships plus profits from sole proprietorship businesses (Schedule C income) plus farm income.

Table 4:

The table is based on data from Compustat ExecuComp obtained via Wharton Research Data Services. We merge in information from Compustat on which month a given company's fiscal year ends in and restrict the sample to firms with fiscal years ending in December. We furthermore restrict the sample to observations for which both of the two total compensation variables tdc1 and tdc2 as well as their subcomponents are available. For 1992-2009 this results in 111,881 observations. The average number of executives covered is 6,216 per year, and the average number of firms covered is 1,223 per year. To obtain real values we deflate by the price deflator from Piketty and Saez, updated to 2009 using BLS data posted at http://www.bls.gov/cpi/cpiurs1978_2009.pdf. To extend our series for aggregate US income to 2009, we include NIPA data from 2009, and extend the number of tax units to 2009 by increasing the number of tax units for 2008 in the Piketty and Saez data using the growth rate in the number of tax units from 2008 to 2009 calculated from the 2008 and 2009 American Community Survey, Table S1201.

Following Piketty and Saez we calculate the number of tax units as the sum of married men, divorced and widowed men and women, and single men and women aged 20 and over.

Table 5, Table 7, and CPS references in text

We use the March CPS files for years 1968 to 2009, covering income years from 1967 to 2008. We combine extracts for a subset of years used in Autor, Katz and Kearney (2008) (from David Autor http://econ-www.mit.edu/faculty/dautor/data/autkatkear08) with variables extracted from Unicon (www.unicon.com), which provides extract tools for CPS data (and is the source of the Autor et al. files). We process the data using the files provided by Autor to make variables comparable over time.

We collapse the files to one observation to family. We keep all individuals who are heads (relhd=1) and use the family weight to calculate all statistics. Income is family income in the year prior to the interview year. We confirmed that our calculated income percentiles match those reported by the BLS in several years (http://www.census.gov/hhes/www/cpstables/032009/faminc/new01_001.htm for example). We deflate all amounts and use the same measure of aggregate income as in Table 1.

For Table 5, for industry, we use the variable indmly variable from unicon and for occupation the variable occ. Because the coding of the latter changes prior to 1982 and after 2001, we use this sub-period.

For Table 7, note that there is no data for 1975 because family weights are missing from the CPS for that year, and that we omit 1973 and 1974 because the topcoding covers more than one percent of the income distribution. The internal data was provided by Jeff Larrimore and corresponds to updated series for Figure 3 in Burkhauser et al (2009).

To study the change in same-family income across years, we merge the March CPS files using code based on code from the NBER (http://www.nber.org/data/cps_match.html). From 1993-2004, two variables, _hhid and lineno, uniquely identify a record. There are some duplicates but these constitute a relatively small fraction. From 2006-2009, _hhid hhid2 and lineno uniquely identify a record. Prior to 1993, we check for consistency of matches by comparing genders of "matches," as the NBER code does.

Table 6, Figure 1:

Group level consumption data are constructed based on micro data from the Consumer Expenditure Survey (CE), using consumption data from January 1982 to February 2009.

The unit of analysis in the CEX is a consumer unit. A consumer unit comprises either: (1) all members of a particular household who are related by blood, marriage, adoption, or other legal arrangements; (2) a person who is financially independent of others (regardless of living arrangements), or (3) two or more persons living together who use their income to make joint expenditure decisions.

Our consumption measure is expenditure on non-durable goods plus some services, constructed from the detailed CE universal classification code (UCC) consumption categories. The main categories of excluded services are health care, education and housing (except the nondurable and service components of household operations), which may all have a substantial investment or durable component. The definition of what is a durable contains a subjective element. Our UCC code processing files are available on request. All consumption values are deflated to be in real terms, see section I.D.

The CE has a panel dimension with each consumer unit providing data for up to four interviews (with less than four interviews available in case of a consumer unit missing one or more interviews). For a given consumer unit, interviews are conducted three months apart. In a given interview, the consumer unit reports consumption for the prior three calendar months. While a given consumer unit is interviewed three

months apart, interviews are spread over the year, with some consumer units interviewed months 1, 4, 7, and 10 and others interviewed months 2, 5, 8, and 11 etc.

To reduce the impact of measurement error we apply the following drop criteria: (a) Drop consumer units with less than the full four interviews of data. (b) Drop consumer units that are incomplete income reporters. (c) Drop consumption for 1980 and 1981 due to a change in the questionnaire regarding food consumption in 1982. The food question changed again in 1987. Dropping growth rates straddling the 1987 change does not materially affect our results. (d) Drop consumer unit-interviews for which the ratio of (our measure of) consumption this interview to that the prior interview is above 5 or below 1/5. (e) Drop consumer units who live in student housing, or who report a change in age of household head between any two interviews different from zero or one year. (f) Drop consumer unit-interviews with zero reported consumption (of our consumption measure) for the three months prior to the month of the interview, non-zero consumption in the interview month, consumption values reported for a different number of months than 3, or for whom consumption or consumer unit characteristics are missing.

The final sample has 274,654 "quarterly" log growth rate observations (where the quotes indicate that these are based on consumption values for one 3-month period relative to the prior 3-month period, but for some households these do not line up with calendar quarters). For a given household the maximum number of growth rate observations is three.

Definitions for Table 6 are as follows:

We define the top 5% of households based on consumption in the prior interview (and using the survey weights). The first row of Table 6 is then based on consumption in the current interview.

When calculating growth rates from one three-month period to the next, we always use the same set of households for the two periods to avoid any impact of households entering or exiting the sample. For row 2 to 4 of Table 6, we calculate the left-hand side variable as follows: We first calculate "quarterly" log growth rates for each consumer unit. In each month, we average these log growth rates across consumer units (using survey weights). We then sum four "quarterly" average log changes to obtain annual growth rates (available at monthly frequency), defined as the log growth rates from a three-month period to the same three-month period the following year. We construct the NIPA-based right-hand side variables (starting from monthly NIPA data) to also be log changes from a given three-month period to the same three-month period the following year.

For row 5 of Table 6, we construct the right-hand side variable as the sum of four values of log(weighted average CE consumption in this three-month period/weighted average CE consumption in the previous three-month period), for the same set of households in the two periods. For consistency we construct the left-hand side variable the same way. Results in row 2 to 5 are similar whether the left-hand side variable is based on weighted averages of log consumption growth rates or the log growth rates of weighted average consumption values.

As explained in Parker and Vissing-Jorgensen (2009) (and the on-line appendix to that paper), sorting on initial consumption leads to nonstandard measurement error when analyzing consumption growth rates of percentile groups, but the use of log growth rates in rows 2 to 5 Table 6 ensures consistency of the estimators.

In row 6, the fraction of aggregate consumption fluctuations borne by group is the coefficient in a regression of (Change in group consumption per household) $\times$ (Group share of population)/ (Lagged aggregate consumption per household) on the growth rate of aggregate consumption per household, with aggregate consumption being CE aggregate consumption.

Since all estimations are based on changes from a given three-month period to the same three-month period the following year, with data available at the monthly frequency, all regressions are estimated using Newey-West standard errors allowing for autocorrelation up to lag 11 (and including monthly dummies, which are never significant).

Definitions for Figure 1:

We define the top 5% households based on consumption in the current interview. For each month, we then calculate the survey weighted average consumption for households interviewed in that month (reporting consumption for the prior three months). We do this for the top 5% and for all CE households. We then average a group's values across months (and multiply by 4) to get annual values and plot the log annual consumption in a given year minus the log annual consumption in 2005. A given consumer-unit observation is assigned to a year based on the last month for which consumption is reported. Note that this implies that 2009-values are based on consumption for households interviewed in February 2009 (reporting consumption for November 2008 to January 2009) and in March 2009 (reporting consumption for December 2008 to February 2009). The 2009 values are roughly similar if consumptions values are seasonally adjusted.

Table 8:

Information about the construction of the SOI-CPS data by the Congressional Budget Office is available at http://www.cbo.gov/publications/collections/tax/2009/methodology.pdf. The data is available at http://www.cbo.gov/publications/collections/taxdistribution.cfm (Additional Data on Sources of Income and High-Income Households, 1979-2005). The income data in the SOI-CPS data are from the Internal Revenue Service's Statistics of Income (SOI), except that (a) for non-filers, income from the Current Population Survey (CPS) is used, (b) some income categories are not covered in IRS data (e.g. in-kind benefits such as health insurance) and information on these are merged in from CPS based on finding an observation in CPS with similar income and demographics as a given SOI observation.

The CBO income measure used to define groups is pre-tax, post-transfer including capital gains defined as the sum of wages and salaries (including employee contributions to 401(k) retirement plans), pensions (including taxable withdrawals from individual retirement accounts and deferred compensation plans), proprietors' income, other business income (partnership income, income from S corporations, and rental income), taxable and nontaxable interest, dividends (paid by C corporations), in-kind income (Medicare, Medicaid, employer-paid health insurance premiums, food stamps, school lunches and breakfasts, housing assistance, and energy assistance), imputed taxes, cash transfer payments, realized capital gains, and a small category of other income. Imputed taxes are defined as taxes paid by businesses, i.e. corporate income taxes and the employer's share of Social Security, Medicare, and federal unemployment insurance payroll taxes. Taxes paid by businesses are distributed to households according to their share of capital income. It is assumed that income would have been higher in absence of those taxes.

The unit of analysis is a household. A household consists of the people who share a housing unit, regardless of their relationships. Households with negative income are excluded from the lowest income category but are included in totals. Groups are defined based on income adjusted for household size (income/square root of household size).

Table 9: .

We purchased an extract from the Canadian Longitudinal Administrative Data (http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=13C0019&lang=eng). The data tracks individuals over time and we construct family income by merging to the family level. We exclude households in the

cross-sectional analysis where the individual sampled died during the year, and in the panel data where the individual sampled died during either year (we do not drop those with family member that die). Groups are defined based on percentile in family income distribution based on total pre-tax, pre-transfer family income, excluding capital gains in year t. We had Statistics Canada construct income groups and compute average income in each group for repeated cross-sections and for families in the database in two consecutive years. Income subcategories that sum to our total income measures are defined as follows with variable names. Wage and salary income: earnings from all T4 slips (T4E__) and other employment income (OEI__). Pensions: Pension and superannuation income (SOP4A). Professional income (PFNET). Business income: Business income, net (BNET_), Farming income (FMNET), Fishing income (FSNET), and Commission income (CMNET). Dividend income (XDIV_   ). Interest income (INVI_). Other investment income: Partnership income (LTPI__), Other income (OI___), Rental income (RNET_), RRSP income/withdrawals (T4RSP), RRSP income of 65+ (RRSPO), and Alimony income (ALMI_). Transfer income: (TRPIN) includes Old age security pension, Canada/Quebec pension plan benefits, Employment insurance, Goods and services tax credit, Provincial refundable tax credits, Family benefits, Non-taxable income, Child tax credit   , Child tax benefit, and Universal child care benefit. Capital gains (CLKGX). Taxes: Tax, net federal tax (NFTXC) and Tax, net provincial tax (NPTXC).

For aggregate income, we use CANSIM Table 202-0202 from Statistics Canada: Average market income, by economic family type; Canada; All family units (dollars). We deflate all series by the expenditure price deflator from CANSIM Table 1.6.

Figure 5: Based on data from Atkinson, Piketty and Saez (2010), available at http://elsa.berkeley.edu/~saez/.