

# Distributional Dynamics

Data guide: what we provide, in plain English

Christian Bayer   Luis Calderon   Moritz Kuhn

October 2026 vintage   ■   [distributionaldynamics.com](https://distributionaldynamics.com)

## 1 What this is

Household surveys measure consumption, income and wealth, but no single survey measures all three, every quarter, over decades. Wealth is surveyed every three years (SCF), consumption in another survey (CEX), income in several others (CPS, SIPP, PSID). Our method combines these five surveys with national accounts and estimates the **joint distribution of consumption, income and wealth of U.S. households, every quarter from 1962 to 2024**. “Joint” means we know not only how unequal each variable is, but also how they line up: for example, how much the income-rich consume, or how much wealth the consumption-poor hold.

Everything described here is publicly available under CC BY 4.0. Please cite: Bayer, C., L. Calderon and M. Kuhn, “Distributional Dynamics”, CEPR Discussion Paper 19829.

## 2 Basics you need for every file

- **Unit:** the household. Consumption and income are annual amounts per household (measured every quarter); wealth is net worth (assets minus debts). Income is before taxes and includes transfers.
- **Dollars:** 2019Q4 U.S. dollars (deflated with the CPI-U). Totals are anchored to the national accounts, so the means per household match the national-accounts per-household values.
- **Time:** quarterly, written YYYY-Qq (for example 2019-Q4).
- **Survey views.** Surveys measure the same households differently; the SCF, for example, captures the very rich much better. The model therefore reports the estimated distribution *as each survey would measure it*: one view per survey file (CEX, CPS, CPS2, PSID, SCF, SIPP1–3). All eight are published as coefficients (product 7). The ready-made statistics and micro data are for the PSID and SCF views; income also comes in two CPS views:
  - **PSID view** (`psid`): the only survey that measures consumption, income and wealth together. Income from 1968, wealth from 1984-Q2, consumption from 1999-Q2.
  - **SCF view** (`scf`): income and wealth, 1962-Q3 to 2024-Q1. Best for the top tail.
  - **CPS views**, income only: `cps1967` (1967-Q4 to 1993-Q3) and `cps` (from 1993-Q4). They are separate because the CPS changed how it measures income in 1993; the top-10% income share is about 2 points higher under the new measurement.

The SCF view has higher top shares than the PSID view; neither is “wrong”.

- **Groups.** Households are ranked within each quarter. Deciles (`d1–d10`), quintiles (`0–20 ... 80–100`) and the wealth groups the Federal Reserve uses (`bottom50`, `next40`, `top10`).
- **Between survey waves** the estimates are model-based: the model uses all surveys and the aggregate data to fill in the quarters a survey does not cover.

### 3 The products

#### 1. Group statistics: inequality over time

|                   |                                                                                                                                                                                                                                                                   |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>What</b>       | Mean per household and share of the total, for consumption, income and wealth, by decile, quintile and bottom 50 / next 40 / top 10.                                                                                                                              |
| <b>Coverage</b>   | 1962-Q3 to 2024-Q1 (SCF view); PSID and CPS views as in Section 2.                                                                                                                                                                                                |
| <b>Files</b>      | <code>dd_tidy_long.csv</code> : one row per quarter, survey view, variable and group (easiest to use).<br><code>dd_series.csv</code> / <code>.dta</code> : one row per quarter, with decile columns (means, totals, shares), also detrended ( <code>_dt</code> ). |
| <b>Use it for</b> | The top-10% wealth share since 1962; how the bottom half's consumption moved in the Great Recession.                                                                                                                                                              |

#### 2. Cross-tabulations: one variable by another's groups

|                   |                                                                                                                                                             |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>What</b>       | Each variable within groups ranked by another variable: consumption by income quintile, wealth of the top 10% by income, income by wealth group, and so on. |
| <b>Coverage</b>   | PSID view: all directions, quintiles and bottom 50 / next 40 / top 10. SCF view: income by wealth group and wealth by income group.                         |
| <b>Files</b>      | <code>dd_tidy_long.csv</code> , rows where <code>ranked_by</code> differs from <code>variable</code> .                                                      |
| <b>Use it for</b> | The consumption gap between the top and bottom income quintile over time.                                                                                   |

They come from the estimated joint distribution, so they are consistent with the totals every quarter.

#### 3. Synthetic micro data: compute your own statistics

|                   |                                                                                                                                                                                                                                                                                         |
|-------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>What</b>       | A household-level stand-in for the PSID view. Each quarter, households are grouped into 1,000 cells by their consumption, income and wealth decile. Each cell has a population weight ( <code>cop_share</code> , summing to 1 per quarter) and its mean consumption, income and wealth. |
| <b>Coverage</b>   | 1984-Q2 to 2024-Q1; consumption from 1999-Q2 (before that, only the income-by-wealth cells have weight).                                                                                                                                                                                |
| <b>Files</b>      | <code>dd_microdata.csv</code> / <code>.dta</code> .                                                                                                                                                                                                                                     |
| <b>Use it for</b> | Any weighted statistic or your own grouping. To merge with your data, rank your households into deciles and join on the quarter and the three deciles.                                                                                                                                  |

Always use the weight. There is no household identifier: the cells are population groups, not people.

#### 4. Uncertainty: standard errors for anything you compute

|                   |                                                                                                                                                                                                                                                       |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>What</b>       | <b>Implicates:</b> the micro data repeated at five draws from the model's posterior.<br><b>Factor draws:</b> 200 posterior draws of the factors (below) and 5/50/95% bands.                                                                           |
| <b>Files</b>      | <code>implicates/PSID_synthetic_microdata_imp1-5.csv</code> ;<br><code>smoothed_factor_draws.csv</code> , <code>smoothed_factors_bands.csv</code> .                                                                                                   |
| <b>Use it for</b> | Compute your statistic on each implicate. The estimate is the average; the standard error is the square root of $1.2 \times$ the variance across the five (as with the SCF implicates). For precise bands, use the 200 factor draws with our scripts. |

## 5. Factors: the quarterly drivers of the distribution

|                   |                                                                                                                                                                                                                           |
|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>What</b>       | Eight latent distributional factors ( $x_1$ – $x_8$ ) that move the whole joint distribution, with their lags ( $x_9$ – $x_{32}$ ) and 11 aggregate factors ( $x_{33}$ – $x_{43}$ ). The factors have no economic labels. |
| <b>Coverage</b>   | 1962-Q3 to 2024-Q1: posterior mode, 200 draws, 5/50/95% bands.                                                                                                                                                            |
| <b>Files</b>      | <code>smoothed_factors.csv</code> , <code>smoothed_factor_draws.csv</code> ,<br><code>smoothed_factors_bands.csv</code> .                                                                                                 |
| <b>Use it for</b> | Feed a factor path (ours, or a counterfactual you construct) to our scripts to get the distribution it implies; use the factors in your own time-series work.                                                             |

## 6. Household balance sheets

|                   |                                                                                                                                                                                                                         |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>What</b>       | Real estate, mortgages, consumer credit, pensions and stocks: mean per household and share of the total, by wealth group (as in the Fed's Distributional Financial Accounts), income quintile and own decile. SCF view. |
| <b>Coverage</b>   | 1962-Q3 to 2024-Q1; pensions from 1983-Q3. Stocks are published for the next 40% and top 10% of wealth only.                                                                                                            |
| <b>Files</b>      | <code>items/dd_items_SCF.csv</code> / <code>.dta</code> (all items); <code>items/&lt;item&gt;_SCF.csv</code> (one each).                                                                                                |
| <b>Use it for</b> | Who holds the mortgage debt; how housing wealth is spread across wealth groups.                                                                                                                                         |

Each item is estimated jointly with income and wealth. Households without the item are modelled explicitly, as a group with zero holdings.

## 7. Coefficients and scripts: any other statistic

|                   |                                                                                                                                                                                                                                                                      |
|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>What</b>       | Each quarter's distribution, stored as polynomial coefficients, one file per survey view (CEX, CPS, CPS2, PSID, SCF, SIPP1–3). Short Python and Julia scripts turn them into shares, means, quantiles, group means across variables, micro data, or posterior bands. |
| <b>Files</b>      | <code>&lt;survey&gt;_coefficients_normal.csv</code> (in sample) and <code>_average.csv</code> (time-averaged trend, for counterfactuals); scripts in the <code>python/</code> and <code>julia/</code> folders.                                                       |
| <b>Use it for</b> | A percentile, a Gini coefficient, groups of your own choosing, or a counterfactual distribution.                                                                                                                                                                     |

## 8. Other joint distributions (products)

|                   |                                                                                                                                                                                                                                                                                                                                                                                                |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>What</b>       | Four more joint distributions, each estimated separately with three measures: consumption with gross liquid assets and consumer debt; with total assets and total debt; with gross liquid assets and mortgage debt; with income and mortgage debt. Group means and shares (own-ranked and cross groups), the quarterly share of households with no debt, and coefficients. SCF and PSID views. |
| <b>Coverage</b>   | 1962-Q3 to 2024-Q1 (SCF view); PSID view as in Section 2.                                                                                                                                                                                                                                                                                                                                      |
| <b>Files</b>      | <code>products/&lt;product&gt;/&lt;product&gt;_tidy_long.csv</code> ,<br><code>&lt;product&gt;_share_zero.csv</code> , coefficients.                                                                                                                                                                                                                                                           |
| <b>Use it for</b> | Debt and liquidity across the consumption distribution; who has no mortgage or no debt, quarter by quarter.                                                                                                                                                                                                                                                                                    |

## 9. Aggregate anchors

|              |                                                                                                                                                   |
|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>What</b>  | Consumption, income and wealth per household from the national accounts, and the number of households. They set the dollar levels of all outputs. |
| <b>Files</b> | <code>aggregate_anchors.csv</code> .                                                                                                              |

## 10. Inputs and replication

|              |                                                                                                                                                                                                                                         |
|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>What</b>  | The cleaned survey files the model reads (CPS, PSID, CEX, SCF, SIPP), the estimation code and the estimated parameters. Reproducing the baseline results takes under 30 minutes; re-estimating the baseline model takes up to 24 hours. |
| <b>Where</b> | The <a href="#">DD_replication</a> repository on GitHub. The survey files are derived from the original surveys, whose providers' terms apply.                                                                                          |

## 4 Loading the data

All files are at <https://www.distributionaldynamics.com/data/> and in the [DD\\_data](#) repository.

### Python

```
import pandas as pd
t = pd.read_csv("https://www.distributionaldynamics.com/data/dd_tidy_long.csv")
t.query("survey == 'scf' and variable == 'wealth' and ranked_by == 'wealth' and group == 'top10'")
```

### Stata

```
use "https://www.distributionaldynamics.com/data/dd_series.dta", clear
```

### R

```
t <- read.csv("https://www.distributionaldynamics.com/data/dd_tidy_long.csv")
```

### Julia

```
using CSV, DataFrames, Downloads
t = CSV.read(Downloads.download("https://www.distributionaldynamics.com/data/dd_tidy_long.csv"), DataFrame)
```

The [codebook](#) lists every column of every file.

## 5 Caveats

- These are **model-based estimates**, not survey tabulations. Between survey waves, and especially before 1989 when there are few SCF waves, the series are interpolations by the model. They step when a new survey wave enters.
- The **survey views differ** by design (Section 2). Compare like with like.
- **Stocks** are published for the next 40% and top 10% of wealth only. **Pensions** start in 1983.
- **Survey definitions** differ somewhat across surveys (for example, which assets count as wealth); details are in the paper and the replication package.

## 6 License, citation, contact

Data and documentation: CC BY 4.0, publicly available. Code: MIT.

Cite: Bayer, C., L. Calderon and M. Kuhn, "Distributional Dynamics", CEPR Discussion Paper 19829.

Questions or requests: [GitHub issues](#) or [luis.calderon@uni-bonn.de](mailto:luis.calderon@uni-bonn.de).