

I want to perform a two-sample IV procedure in the following way.

1. Run a regression with trust as the dependent variable (LHS) and demographic variables as the independent variables (RHS) similar to the regression used to produce table 1 in the paper "Returns Heterogeneity and Consumption Inequality over the Life Cycle" by Daminato and Pistaferri (2014).

The trust and demographic variables come from the 2020 wave of the Household Retirement Survey (HRS). I made an effort to match the following variables described by the paper: age, education, employment, shares of wealth allocated to different asset classes, amounts of debt to compute total debt to wealth (if relevant for my uses), computed wealth percentiles after constructing a wealth variable.

2. Use the estimated relationship between trust and demographics to impute a measure of trust in the PSID data. I started with the 2019 wave and located the imputed version of each of the variables described above (since it was the initial reference point).

I've had difficulty locating the following variables, which are needed to compute returns to wealth (eq. 1 in the paper): capital gains/losses, payments on debt, estimate of net investment flows. Also, the HRS only started measuring trust in 2020 but DP2024 uses panel data to estimate fixed effects in the regression with returns as the dependent variables.

My first thought was that, if I can compute returns for 2019 and then do the same for more PSID waves, I can use the same imputed trust variable for each wave (implicitly assuming that individuals had the same level of trust across the years).

A final, minor concern is that the years of the survey don't exactly line up (although they are more conducted every two years).