Consumption Inequality Over the Last Half Century. Some Evidence using the New PSID Consumption Measure

By ORAZIO ATTANASIO AND LUIGI PISTAFERRI*

There is a lively debate in the U.S. regarding trends in inequality over the last 40 years, and about its causes and consequences. An interesting aspect of the debate is that trends in inequality for different measures of household resources (such as wages, earnings, income, consumption or wealth) seem to be different, and sometimes in a non-negliglible way. While there is little doubt that wage inequality has risen substantially (Autor, Katz and Kearney, 2008), the evidence for individual earnings is more nuanced (especially for women, see Gottschalk and Danziger 2005). On the consumption side, the evidence is even more mixed. On one hand, Krueger and Perri (2006) find that consumption inequality grew very little in the 1980-2000 period despite the large increase in wage or income inequality. Their results are confirmed, although less dramatically, by Meyer and Sullivan (2010). On the other hand, Aguiar and Bils (2012), Attanasio, Battistin and Ichimura (2004), Attanasio, Battistin and Padula (2009) and Attanasio, Hurst and Pistaferri (2012) argue that consumption inequality has risen more than initially believed. Finally, trends in wealth inequality (as reported by Kopczuk and Saez, 2005) are surprisingly *downward*.

The goal of this paper is to contribute to the debate regarding trends in consumption inequality. Knowing whether consumption inequality grows as much as income inequality or substantially less has important policy implications. For example, Krueger and Perri's (2006) results have been used to argue that the welfare implications of the growth in income inequality are less worrying than it would appear at first. Some of the differences can be explained by measurement issues (Attanasio, Hurst and Pistaferri, 2012; Attanasio, Battistin and Padula, 2009). For example, the quality of consumption data and the quality of wage data could be very different, although some researchers have argued that for people at the very bottom of the income distribution, consumption is easier to measure and a better measure of well-being than income (Meyer and Sullivan, 2010). In general, however, different patterns of inequality in earnings and consumption could arise simply because they are different concepts and are due to behavioural choices as well as the ability (or lack thereof) that individual might have to smooth out some income shocks. Different patterns of consumption and income inequality can therefore be informative both about the nature of shocks individual households face, about individual preferences and about the ability to absorb shocks. Given these considerations, appropriate measurements of consumption inequality and its trends are extremely important to establish the welfare consequences of income inequality.

Much of the literature on consumption inequality has used the Consumer Expenditure Survey (CEX, which is, despite a number of problems, the only micro-level data set that contains comprehensive information on consumption since the early 1980s. In this paper, we present a new measure of consumption inequality with the goal of providing some fresh evidence on the important issue of how trends in consumption inequality have evolved over a long time period in the US. To this purpose, we use the re-designed PSID data. Starting with the 1999 wave, the PSID has began collection of a larger array of information on consumption components, besides food (one of the few items consistently present in the survey before 1999). Based on the fact that these data appear to match NIPA aggregates well, we propose imputing consumption to the PSID families observed

* Attanasio: Department of Economics, University College London, Gower Street, London WC1E 6BT; IFS and NBER. E-mail: uctpjrt@ucl.ac.uk. Pistaferri: Department of Economics, Stanford University, Stanford CA, 94305, USA; SIEPR, and NBER. Email: pista@stanford.edu. Thanks to Davide Malacrino for invaluable research assistance, Richard Blundell for comments, and participants to the 2013 conference on "New Directions in Consumption Research with the PSID" for comments.

in the years *before* 1999 using the more comprehensive consumption data available from 1999 onward. The main advantage of this procedure is that we have a direct way of testing whether trends in consumption inequality are replicated by the imputation procedure at least for the period in which we have data on both (in-sample verification). Another advantage is that given that the PSID is available since the late 1960s we can provide the longest time series available on consumption inequality.

I. Imputing consumption

We extend an idea that has already been used in the literature (see, for instance, Blundell, Pistaferri and Preston, 2008, Aguiar and Bils, 2012, and Skinner, 1992). The PSID contains information only on some consumption items consistently over a long period of time. One can then think of imputing total consumption, or consumption items that are missing, using information available in the PSID about other variables. To this purpose it is necessary to infer the relationship between these other variables and total consumption using either an alternative data set or a different time period. In what follows, we use the late years of the PSID, when information on a larger number of commodities is available, to estimate the relationship of interest. The extrapolation we perform is theory-consistent in that it is based on a theoretical relationship that links the observed variables to what we want to estimate in a way that is consistent with a model of consumer demand. Within a typical demand system, the allocation of total resources spent in a given period over different commodities depends on relative prices, taste shifters (such as demographic variables), and total expenditure. One can think of inverting this relationships to infer total expenditure from information on individual commodities, relative prices and taste shifters. Our approach does this in an *ad-hoc* fashion, which can be interpreted as an approximation of a demand system.[1]

[1]In the working paper version of this paper (available on our websites), we have also implemented an alternative approach which explicitly estimates a demand system. This is omitted here for lack of space.

A. Data description and selection

Collection of expenditure data in the PSID has often been of secondary interest. The component that is consistently present throughout the survey for all households (with the exception of 1973, 1988, and 1989) is spending on food (at home and away from home). The survey also collects information on the monetary value of food stamps (with the exception of 1973). Another component that is also available consistently over the sample period is rent payments (again, with the exception of 1988 and 1989), but there is no information on rent equivalents for non-homeowners. Before 1999, the survey occasionally collected information on other consumption components, such as home insurance, utilities, and child care. Starting with the 1999 wave, however, the PSID began collection of information on a larger number of consumption components, now covering about 70% of non-durable spending from the national accounts. They include health expenditures, utilities, gasoline, car maintenance, transportation, education, and child care. A few additional consumption categories were added in 2005 (such as clothing and entertainment). The new PSID data are very valuable because they seem much better aligned with NIPA than the CEX, see Blundell, Pistaferri and Saporta (2012) (but see also Meyer and Sullivan, 2012, for a view that this match is good even in the CEX; i.e., the CEX matches well NIPA for items that are also in the PSID, and matches very poorly the items that are not in the PSID).

We use data from the 1968-2011 Panel Study of Income and Dynamics (PSID), and for reasons of space refer the interested reader to the PSID website for details about the survey sampling scheme. The PSID data were collected annually until 1996 and biennially starting in 1997.

We started by appending all the PSID family data files. The resulting data set (after dropping the SEO, Latino, Immigrant subsamples) includes 145,199 observations. We replace topcoded values for family income with Pareto estimates. We also set to missing some observations for food (at home and away from home) that we judged to be outliers (7 observations) or that are topcoded (116 observations). Next, we drop households where the head is younger than 25 or older than 65 (36,629 observations) and

those where the head or spouse (if present) have implausible hourly wages (below half of the federal minimum wage). This gives our final sample of 105,047 observations.

B. Methodology

To impute the the logarithm of net consumption[2] (defined below), $\ln n$ we use the following equation:

$$\ln n_{it} = Z'_{it}\beta + p'_t\gamma + g(f_{it};\theta) + u_{it} \quad (1)$$

where Z are socioeconomic variables, p are prices, f includes consumption components that are consistently collected throughout the years (i.e., food at home, food away from home, and food stamps), $g(.)$ a polynomial function, and u an error term.[3] Equation (1) can be interpreted as an (approximated) demand system that relates f to n. Relative prices are included to control for the fact their changes will induce changes in the allocation of total expenditures among different commodities.

Our imputation equation is estimated using data for the period in which we have a more comprehensive measure of consumption (the 1999-2011 waves, or 1998-2010 calendar years given that data are retrospective). Our net consumption measure n_{it} (the sum of home insurance, electricity, heating, water, miscellaneous utilities, car insurance, car repairs, gasoline, parking, bus fares, taxi fares, other transportation, school tuition, other school expenses, child care, health insurance, out-of-pocket health, rent) excludes food consumption and the consumption categories added in the 2005 wave (clothing and entertainment) in order to deal with a homogenous consumption series. Our net consumption measure also includes rent, which are equal to actual payments for renters and an estimate of the rent equivalent for homeowners. Since homeowners report the value of their house, we impute a rent equivalent measure equal to 6% of the self-reported house value (Flavin and Yamashita, 2010). The pooled OLS regression is run on 26,815 person-year individuals surveyed in 1999 or later.

We then construct a measure of imputed total consumption as given by:

$$\widehat{c}_{it} = f_{it} + \exp\left\{Z'_{it}\widehat{\beta} + p'_t\widehat{\gamma} + g\left(f_{it};\widehat{\theta}\right)\right\} \quad (2)$$

The measure so obtained is transformed in real, adult-equivalent terms by dividing it by the overall CPI index, and then by the OECD adult equivalence scale.[4] This is the variable whose distributional features we discuss below.

C. Results

The results of estimating the imputation regression (1) are reported in the working paper version and only summarized here. In particular, for the years after 1999 we regress the observed total consumption on a third-degree polynomial in total food consumption (the sum of spending on food at home, food away from home, and the monetary value of food stamps - results are similar if we use a linear function in food), socio-economic variables (dummies for age, education, marital status, race, state, employment status, self-employment, head's hours worked, homeownership, disability, family size, the number of children in the household), and relative prices (the overall CPI and the CPIs for food at home, food away from home, and rent). As we mentioned above, this regression could be interpreted as being an approximation to a relationship derived from a demand system. The adjusted R^2 of the regression is just above 0.5.

The most interesting aspect of this imputation procedure approach is that we have an indirect way of checking how accurate this procedure is in predicting trends in consumption inequality (which in the end is what we are interested in) by simply looking at in-sample forecasts. That is, we can compare $st.dev.(\log\widehat{c})$

[2] We use net consumption to avoid bias from correlated errors (in food).

[3] An initial specification also included a household fixed effect in the attempt to capture unobserved time-invariant tastes for consumption. This procedure, however, would impute consumption backward only for the households observed in the 1999-2011 period (for whom a fixed effect can be estimated). Given attrition, small sample sizes, and age effects, the results were very unstable, and thus we decided to rely on a simpler statistical procedure and try to capture heterogeneity using a rich control function.

[4] The OECD scale is $(1 + 0.7(A-1) + 0.5K)$, where A is the number of adults and K the number of children in the household unit.

and $st.dev.(\log c)$ for the 1998-2010 period, where both measures are available. Of course, $E(\log \widehat{c}) = E(\log c)$ by construction, but this is clearly not true for the variance (or, more precisely, for the trends in the variance).

Panel A of Figure 1 plots the standard deviation of the log of actual consumption over the 1998-2010 period (the dashed line) and the standard deviation of the log of imputed consumption (obtained from the regression above, the solid line). The two series are normalized to their value in the initial year (1998). Clearly, the in-sample forecasts appear able to reproduce trends in the true standard deviation quite well: consumption inequality increases by about 7 points up to 2006, before declining in the last two waves (with some evidence of a decline associated with the Great Recession and already documented by many researchers, among others Petev, Pistaferri and Saporta-Eksten, 2012).

Panel B shows the effect of the imputation procedures farther back in time (to 1967). Again, the series are normalized to the value of 1998. After a period of relative stability (if anyting, declinining inequality) that lasts until the mid-1970s, dispersion in the distribution of log consumption appears to rise significantly. In particular, between 1978 and 2010, the growth is of about 13 points.

While this imputation procedure does well in replicating trends in the standard deviation of log consumption, one may wonder whether it can replicate trends in other inequality measures that are less subject to extreme values. To this purpose, in Panel C we plot the Gini coefficient (we get similar results if we use the interquartile range). Three things can be noted. First, our imputation procedure works also for this different inequality measure. Second, the rising trend in consumption inequality after 1980 is confirmed (together with the declining level of inequality of the 1970s). Between 1978 and 2006 the Gini coefficient rises by 0.05 points per adult equivalent and in real terms. Finally, there is evidence of a substantial slowing down of consumption inequality during the Great Recession.

How does the growth in consumption inequality compares to the growth in income inequality? In the final Panel D of Figure 1 we plot our consumption inequality series (mixing actual and estimated standard deviation of logs) against the standard deviation of the log of family income (also transformed into real, adult-equivalent terms). During the 1970s the two series are synchronized and relatively stable, a well known fact. Between 1978 and 1993, both series grow, but income inequality rises faster than consumption inequality (15 vs. 6 points). However, income inequality after 1993 slows down, while consumption inequality keeps rising. In the last 10 years of the sample period, however, income inequality again rises faster than consumption inequality, and during the Great Recession income inequality keeps rising while the rise in consumption inequality comes to a complete halt.[5]

II. Conclusions

This paper contributes to the debate regarding trends in consumption inequality in the US. We present a new measure of consumption inequality based on the redesigned 1999-2011 PSID. We impute consumption to the families observed before 1999 using the more comprehensive consumption data available from 1999 onward. One advantage of this procedure is in-sample verification of the quality of the imputation procedure; another is that it yields a long time series (1967-2010). Consumption inequality was stable in the 1970s, as was income inequality. It increased significantly after 1980. The Great Recession was associated with a decline in consumption inequality. Trends in income inequality are similar, but not totally synchronized. For example, income inequality slowed down in the second half of the 1990s and went up during the Great Recession. Understanding why these trends differ is an important question left for future research.

REFERENCES

Aguiar, Mark A. and Mark Bils. 2011. "Has Consumption Inequality Mirrored Income Inequality?," NBER Working Papers 16807.

Attanasio, Orazio, Eric Battistin and Hide Ichimura. 2004. "What Really Happened to

[5]The trends in consumption inequality are similar if we use a larger sample (heads aged 18-80), if we exclude health and education from our consumption measure, or if we impute backward using the more comprehensive 2004-2010 consumption measure (which includes clothing and entertainment) .

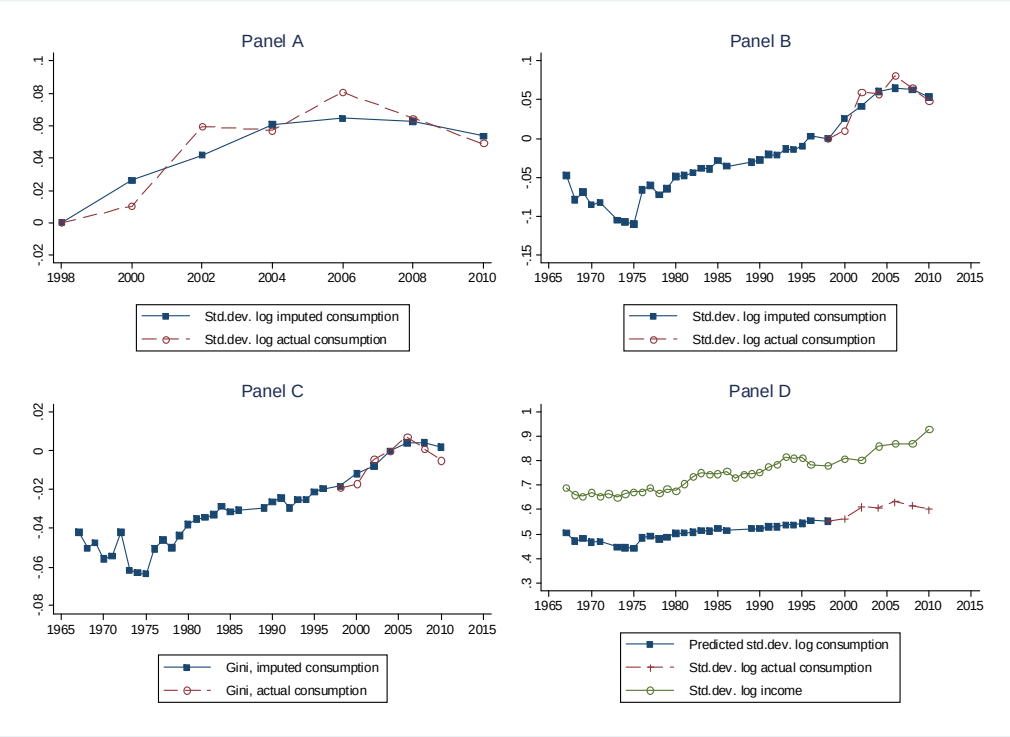

FIGURE 1. CONSUMPTION INEQUALITY AND INCOME INEQUALITY.

Consumption Inequality in the US?," NBER Working Papers 10338.

Attanasio, Orazio, Eric Battistin and Mario Padula. 2010. *Inequality in Living Standards since 1980: Income Tells Only a Small Part of the Story*, Packt Publishing.

Attanasio, Orazio, Erik Hurst and Luigi Pistaferri. 2012. "The evolution of income, consumption, and leisure inequality in the US, 1980-2010", NBER Working Papers 17982.

Autor, David, Lawrence Katz, and Melissa Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists". *Review of Economics and Statistics*. 90(2): 300-323

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption Inequality and Partial Insurance." *American Economic Review*, 98(5): 1887-1921

Blundell, Richard, Luigi Pistaferri and Itay Saporta-Eksten. 2012. "Consumption Inequality and Family Labor Supply", NBER Working Papers 18445.

Flavin, Marjorie, and Takashi Yamashita. 2002. "Owner-occupied housing and the composition of the household portfolio", *American Economic Review*, 92: 345–362.

Gottschalk, Peter and Sheldon Danziger. 2005. "Inequality Of Wage Rates, Earnings And Family Income In The United States, 1975-2002 ," *Review of Income and Wealth*, 51(2): 231-254.

Kopczuk, Wojciech and Emmanuel Saez. 2004. "Top Wealth Shares in the United States: 1916-2000: Evidence from Estate Tax Returns," National Tax Journal, 57(2): 445-87.

Krueger, Dirk and Fabrizio Perri. 2006. "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *Review of Economic Studies*, 73(1): 163-193.

Meyer, Bruce and James Sullivan. 2009. "Five Decades of Consumption and Income Poverty," NBER Working Papers 14827.

Petev, Ivaylo, Luigi Pistaferri and Itay Saporta-Eksten. 2012. "Consumption and the Great Recession: An Analysis of Trends, Perceptions and Distributional Effects", in *Analyses of the Great Recession*, D. Grusky, B. Western and C. Wimer (eds.)

Skinner, Jonathan. 1987. "A Superior Measure of Consumption from the Panel Study of Income Dynamics." *Economics Letters*, 23: 213-216.