Note prepared for the LCS workshop, January 14, 2025

Luigi Pistaferri (Stanford University)

There are many reasons why consumption may be preferred to income or wealth as an overall indicator of well-being. Consumption is the typical argument in the utility functions used by economists; and it may be a better indicator of the long-term resources of a household. On the other hand, consumption is much harder to measure than the other two indicators, and this may perhaps explain why we have many surveys collecting information on income, but very few collecting detailed information on consumption. Similarly, administrative sources of income data are more common than administrative sources of consumption or spending data. Some of the measurement problems are conceptual (i.e., what to include in consumption, how to measure services from durables, how to account for price differentials across space, how to measure the value of home-produced goods, etc.), while others are more practical (related to the collection mode, recall issues, etc.). The "white paper" circulated in advance of the meeting discusses many of these issues. In Pistaferri (2015) I offered some general reflections on consumption data collection issues.

I have used consumption data in several papers. As I describe below, I have experimented with different consumption definitions, with consumption data from different countries, and with different collection/measurement strategies. I have not personally been directly involved in the collection of consumption data, except for a series of projects in which (together with collaborators) we have elicited propensity to consume with respect to hypothetical changes in income or wealth. Below I discuss the challenges I have encountered in working with consumption data and some of the solutions I adopted or have been proposed by others.

CES, PSID and combined CES-PSID data

In some of my work, I have used US consumption data available from two surveys: the Consumer Expenditure Survey (CES) and the Panel Study of Income Dynamics (PSID). In a few papers I have combined information from the two datasets using imputation strategies. Both the CES and the PSID are useful to paint a picture of the spending habits of Americans, but present several issues, some common to both, some more idiosyncratic.

The CES is the only data set with comprehensive and detailed information on household expenditure and its various components. Available on a continuous basis since 1980, it is

used by the Bureau of Labor Statistics (BLS) to form weights that go in the computation of the Consumer Price Index (CPI) (and for other minor matters as well). The CES is composed of two distinct modules, the Interview survey (where spending information is by three-month recall), and the Diary survey (where spending is collected by filling a 2-week diary). Respondents in the Interview survey are sampled every 3 months (for a total of 5 times, although data for the first interview are not released because they are merely preparatory), while those in the Diary survey are sampled only once. The two modules cover different consumption items, with some overlap.

While the CES is in principle the data set of reference for researchers interested in studying consumption behavior, it has several issues. The most relevant one is a detachment of its average consumption from NIPA (National Accounts) aggregates. The detachment is worse for some items and it is less serious for others. Bee, Meyer, and Sullivan (2013) show that well-measured categories include imputed rent on owner-occupied housing, food at home, gasoline, communication, rent and utilities, and vehicle purchase. Service flows from houses and vehicles are also reliable because ownership of these goods appears to be reported accurately. On the other hand, food away from home, purchase of furniture, clothing and alcohol are severely underreported relative to NIPA.

Depending on the questions one is interested in, it is possible to focus on a subset of goods to draw more general inferences. For example, Meyer and Sullivan (2023) use only the well-measured components of consumption, which under some circumstances can be used to proxy for trends in overall consumption inequality (the assumptions are that the well measured goods are a constant share of total consumption and that the prices of the well measured goods have moved in parallel with the price of the whole basket of goods that households consume).

To study consumption behavior, an alternative to the CES is the PSID. The PSID is available on an annual basis from 1968 to 1997, and on a bi-annual basis after 1997. The PSID's initial goal was to allow social scientists to study income dynamics (and poverty) between and across generations. For these reasons, information on consumption was ancillary. Until 1997, the PSID collected information only on a few consumption items, mostly food (at home and away from home), rent paid by renters, and (very occasionally) utility payments. These items are only a share of total consumption of households. Starting with the 1999 wave, however, the PSID has begun collection of information on a much larger range of items (covering about 70%-90% of the items collected in the CES). Respondents

typically report spending for broad categories, with the reference period being (with some exceptions) the previous calendar year.

The PSID new consumption data have proved able to capture some broad aggregate trends, despite their lower scale and scope, but come from a small sample, where representativeness is a problem, and where coverage of the goods that households purchase is less than 100%. Moreover, the coverage has changed over the years, so researchers who are interested in consumption dynamics either use a measure that remains constant over time (but has low coverage of total spending), or use a more complete measure but have to impute missing consumption in the earlier waves.

The PSID is typically used by researchers who are interested in studying the joint dynamics of consumption and income, since (unlike the CES) it has a long panel. A long time-series of consumption-income data is required to test theoretical predictions, for example, predictions from the life cycle model or the permanent income hypothesis. However, as said above, before 1999 the only consistent consumption information that is collected in the PSID is on food spending (and the value of food stamps), which many researchers found limiting, since food is a necessity and it constitutes a declining share of total consumption. The solution to this has been, in some cases, to combine data from the CES and the PSID to impute total consumption using information on food and demographics available in both surveys.

Imputation of consumption in the PSID has a long history. One early example is Skinner (1987). The Skinner procedure consists of regressing in the CES:

$$c_{it} = \theta_0 + \theta_1 f_{it} + \theta_2 x_{it} + \eta_{it}$$

where c is total non-durable spending, f is food spending, and x are observable household characteristics. Skinner (1987) then obtains the imputed value of consumption in the PSID as: $\hat{c}_{it} = \hat{\theta}_0 + \hat{\theta}_1 f_{it} + \hat{\theta}_2 x_{it}$. The key element that makes this procedure feasible is that while c is only observed in the CES, the other items (food spending and x) are observed in both datasets.

In Blundell, Pistaferri and Preston (2008) we extended the Skinner's idea by using a theory-consistent imputation procedure. We first posit a demand function for food:

$$f_{it} = \alpha + \beta c_{it} + \gamma p_t^F + \delta p_t^{NF} + \lambda x_{it} + \epsilon_{it}$$

where p^F and p^{NF} are prices of food and non-food items. The imputed value of total consumption in the PSID can be obtained by inversion:

$$\hat{c}_{it} = \hat{\beta}^{-1}\left(f_{it} - \hat{\alpha} - \hat{\gamma}p_t^F - \hat{\delta}p_t^{NF} - \hat{\lambda}x_{it}\right)$$

In a companion unpublished paper, Blundell, Pistaferri and Preston (2008), we discussed the properties of this imputation procedure. A paper by Crossley, Levell and Poupakis (2022) suggests that the estimator proposed by Blundell, Pistaferri and Preston (2008) (which they rename "rescaled-regression-prediction" estimator) has better empirical properties that the Skinner (1987) procedure.

In Attanasio and Pistaferri (2014) we consider a hybrid imputation procedure. Since food is available in the PSID already, so there is no need to impute it, we construct a measure of total spending net of food spending in the CEX (n_{it}), and regress:

$$n_{it} = \mu_0 + \mu_1 f_{it} + \mu_2 x_{it} + v_{it}$$

Finally, we impute in the PSID the measure: $\hat{c}_{it} = (1 + \hat{\mu}_1)f_{it} + (\hat{\mu}_0 + \hat{\mu}_2 x_{it})$. This procedure allows us to impute consumption data back to the late 1960s when the PSID started (of course, we cannot control for time effects since the CES starts in 1980). We show that this technique has good out-of-sample performance (i.e., if used in the years when PSID has both food and total spending, relevant moments of c_{it} and $\hat{c}_{it}$ align closely). Needless to say, while these imputation procedures are creative, they are also plagued by measurement issues and low statistical power.

Both the CES and the PSID (in its current format) obtain a measure of consumption as the sum of many sub-aggregates. In the CES alone there are more that 500 UCCs (universal classification codes). While not all of them are used to construct a measure of consumption, most are. There has been some debate on how much of the detachment from the NIPA aggregates may come from the excess burden imposed on respondents, survey fatigue, and so on. The fact that for wealthy households the value of time is higher now than it was when the CES started may be a possible explanation for why overall non-response, item non-response and measurement error are all larger for people at the top of the income/wealth distribution. One possible solution would be to add a scanner component to the datasets that, at least for some items, would reduce data collection burden. Another possibility is to impute consumption of wealthy people (where measurement is poor) using the part of the distribution where consumption is better measured (assuming some external validity conditions are met). Another possibility is to collect more limited data. The SHIW (discussed next) is one example.

SHIW data

In some of my work I have used consumption data based on a limited set of retrospective questions, in which households are asked to recall their total spending of non-durables during a given reference period. An example is Italian consumption data available from the Survey of Household Income and Wealth (SHIW), a representative survey of the Italian population that is conducted biannually by the Bank of Italy. The survey collects an overall measure of non-durable consumption, as well as information on food spending, the purchase of selected durable goods, and the estimated value of selected durables owned. To give an example, in the 2010 wave of the survey, household non-durable spending was obtained from the answer to the following question:

"*How much did your household spend on average per month in 2010 in cash, by credit card, cheque or debit card, on all items? Include all spending, for both food and non-food, and exclude only the following items: purchases of jewelry, cars, etc., maintenance, alimony, allowances, gifts; extraordinary home maintenance; rental of dwelling; regular mortgage payments; life insurance premiums; contributions to supplementary pension schemes*."

Information on actual rent paid (for renters) and imputed rent (for homeowners) are collected separately.[1] Households are also asked to report their food spending,[2] and the value of purchases in three major durable categories, valuables, vehicles, and furniture.[3] For the same three durable categories, households were also asked to provide an estimate of the value of all the goods owned at the end of the calendar year, and prompted to think of what they would have received in 2010 if they had sold them.

There are pros and cons from using a measure of consumption that is based on such limited set of questions. The most obvious advantage is that these questions are cheap to include in a comprehensive, non-consumption focused survey as the SHIW. In large scale surveys, there may be some cognitive burden and respondent fatigue if long consumption modules are asked, while a limited set of questions may attenuate the burden on the respondents. Moreover, the questions on the purchases and value of owned durables allow

---

[1] "*What was your monthly rent in 2010, excluding condominium charges, heating and other expenses?*" and "*Imagine you wanted to let your house/flat, what monthly rent do you or the household think could be charged? Do not include condominium charges, heating or other expenses*", respectively.
[2] "*What, instead, is the average monthly expenditure on food alone? This includes spending on food in supermarkets and the like and spending on meals eaten regularly outside the home.*"
[3] *Valuables* include: jewelry, ancient or gold coins, works of art, antiques including furniture; *Vehicles* include: cars, motorcycles, caravans, motor boats, boats, bicycles; *Furniture* include: furniture, furnishings, rugs and carpets, lamps, small household appliances, washing machine, dishwasher, vacuum cleaner, floor polisher, TV, PC, fridge, cooker, heater, air conditioner, radio, video-recorder, CD player, HI-FI equipment, mobile phone, fax machine, camera, camcorder.

the construction of some rough measures of services from durables. On the other hand, recall questions have many disadvantages. People may genuinely forget purchases made during the reference period (especially if such purchases are small, infrequent, irregular, or stigmatizing). They may also misinterpret what to include, confusing personal and business expenses, counting gifts received and omitting gifts made, etc.

The problem of a survey like the SHIW is that it has the opposite problem of the CES – we're asking consumers to report information on a huge aggregate that may be quite hard to recall. One solution (which the SHIW has adopted) is to add a component of non-durable consumption that is relatively easier to recall (food) to benchmark for errors. Indeed, food spending at home it appears to track NIPA aggregates well. An alternative, inexpensive way to obtain a measure of total consumption would be to ask two questions: (1) how much is spent on food at home, and (2) what is the share of total consumption that is spent on food at home (i.e., the variable $\omega_i = f_i/c_i$). If information is collected on ω_i and f_i, it becomes immediate to obtain a measure of total consumption c_i.

Administrative-based measures of consumption

There is currently a "Big Data" revolution taking place in many fields, including economics. Researchers are increasingly using complex, large, and sophisticated administrative databases to answer important questions in Labor Economics, Public Finance, International trade, etc. For example, people have managed to access micro-level IRS data, Social Security records and matched employer-employee information. Despite its quantitative relevance, consumption is lagging behind in administrative data collection effort and study. Four exceptions can be cited: (1) Spending data available from credit/debit card providers; (2) Use of longitudinal administrative tax record information on income and wealth to create consumption starting from the budget identity; (3) data on spending, income, assets, etc., for consumers using online financial aggregators; (4) scanner data information coming from Nielsen Homescan-style data sets.

In principle, administrative data reduce the extent of measurement error that plagues most survey data. However, the creative data collection efforts cited above come with their own problems. Scanner data are the closest equivalent to administrative data we can get on spending. However, the data are either not representative of all households or not representative of all the goods that people buy (most data sources refer primarily to food/grocery store items). Similar representativeness issues apply to spending data originating from credit/debit card providers or online financial aggregators.

In my own work, I have used empirical strategy (2). From the household budget constraint, we know that consumption can be written as:

$$c_{it} = a_{it+1} - a_{it} + (ra_{it} + y_{it})$$

i.e., the change in the value of one's net worth plus total income (the sum of income from labor and income from capital). Some administrative surveys from Scandinavian countries (Denmark, Sweden, Norway) allow this. In principle, all components of the left-hand side are measured with minimal error since most income and asset records come from third-parties (firms, banks, etc.); sample sizes are enormous since the data refer to the whole population; and (unless people die or migrate) records are longitudinal by nature. However, there are important issues. First, the budget constraint can be used to compute total spending, which doesn't necessarily coincide with total consumption (the sum of nondurables and services from durables). For example, purchasing a vehicle is treated as consumption in the budget constraint of the year of the purchase but in no other years, even though a car purchased today provides services for many years after the initial purchase. Changes in assets reflect partly active saving (or dissaving) and partly capital gains (or losses); moreover, some capital gains are realized, while others exist only on paper.

Some authors have benchmarked the results they obtain against (small) survey data. Some authors have also been quite entrepreneurial in adding data on inheritances and gifts, car registry, etc. Others have used special registries to measure the increase in value of specific asset components (securities, housing, etc.).

Ideas for data collection

The recent advances in AI technologies could be leveraged to improve measurement of consumption data. There are two areas of potential improvement: (a) reducing respondent burden (through automation, smart prompts, etc.), and (b) improving accuracy of reported data (through bias correction methods, data validation, multi-source data merging, etc.).

Thoughts about comparative analysis from the LCS database

I applaud the initiative of creating the LCS database. I think it would serve the research community greatly.

Here are some thoughts.

I do agree that one challenge for comparability is that consumption may refer to different reference periods, and this may even change across countries. It is important to account for seasonal effects. It is also important to account for potential country-specific patterns. For example, in some countries workers receive extra salary payments for summer and Christmas vacations that affect the seasonality of their consumption behavior. See Browning and Collado (2001). There are country-specific holidays that generate similar idiosyncrasies.

The value of housing services is notoriously hard to estimate, and I believe the proposed approach is based on best practices. Still in the light of comparability across countries, I would add that countries differ substantially in the extent of government-related or employer-provided housing subsidies ("fair rent" agreements, housing vouchers, etc.). This is also due to tax considerations. Asking people how much they pay to rent their dwelling may understate the housing services received. In the SHIW, for example, this is accounted for by asking some follow-up questions: the type of rental contract (i.e., whether rent-controlled, informal, etc.), whether the respondent believes the rent she pays is below market, and if so, the survey asks: "*If your rent were at the market rate, how much do you think you would pay a month excluding condominium charges, heating and other expenses?*". This is an attempt to compute the housing services received as renter.

The notion that health expenditures do not increase people's welfare, and so should not be included in the consumption measure (with the exception of health insurance premiums), is not uncontroversial. For example, one may think that things like cataract surgery, joint replacement, dental restoration, hearing aids, etc., all raise quality of life and the marginal utility from non-durable consumption (like travel or entertainment). There are other examples (such as prenatal care). However, I do agree that health systems across countries are enormously heterogeneous and imputing health spending when services are provided by the national health system is challenging.

I liked the idea of collecting information on major durable goods (ownership, purchase price, year of purchase) to try and measure the services from them. Information on quality, even if rudimentary, would be very valuable.

References

Attanasio O. and L. Pistaferri, 2014. "Consumption Inequality over the Last Half Century: Some Evidence Using the New PSID Consumption Measure," American Economic Review, vol. 104(5), pages 122-126, May.

Bee, A., B.D. Meyer, and J.X. Sullivan, "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?", in Carroll C., T. Crossley and J. Sabelhaus (Eds.), *Improving the Measurement of Consumer Expenditures* (University of Chicago Press), chap. 7, 2013.

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

Browning, M., and M. D. Collado. 2001. "The Response of Expenditures to Anticipated Income Changes: Panel Data Estimates." *American Economic Review* 91(3): 681–692.

Meyer, B.D. and J.X. Sullivan, "Consumption and Income Inequality in the United States since the 1960s", Journal of Political Economy 131:2, 247-284, 2023.

Pistaferri, L. "Household Consumption: Research Questions, Measurement Issues, and Data Collection Strategies", Journal of Economics and Social Measurement 40(1-4), 123-49, 2015.

Skinner, J.. 1987. "A Superior Measure of Consumption from the Panel Study of Income Dynamics." *Economics Letters* 23(2): 213–216.