The Technical Appendix for the Economic Insecurity and the Great Recession

November 2011
This appendix focuses on changes in the construction of the Economic Security Index (ESI) made since the 2010 initial release, particularly the adoption of the March supplement to the Current Population Survey (CPS) as the primary source of data regarding year-to-year income changes. The ESI also draws on three additional sources: the Survey of Income and Program Participation (SIPP), which was the primary data source for the prior ESI series released in 2010; the Panel Study of Income Dynamics (PSID); and the Consumer Expenditure Survey (CEX) of the Bureau of Labor Statistics. For more detail on the design and methodology of the ESI, see our previous technical report at http://economicsecurityindex.org.

The remainder of this report is structured as follows. Section I explains the process used to match households in the CPS from year to year. Unlike traditional panel surveys, which follow households or individuals over time, the CPS is a survey of geographic residences. A matching algorithm is therefore needed to determine which individuals are the same from one year to the next. Section II details known issues related to the quality of income measurement in the March CPS, namely imputations and topcoding, as well as the ESI’s treatment of the relevant variables. Unlike the SIPP, the CPS lacks a measure of liquid financial wealth used to decide who is capable of buffering against an economic loss. The ESI therefore requires an imputed measure of wealth based on data from the SIPP. Section III describes this process and presents data on the reliability of these imputations. Section IV briefly discusses the medical out-of-pocket (MOOP) imputation.

I. CPS Matching

The new ESI uses income data from the CPS March supplement, an annual survey that asks detailed labor force and income questions. It is a large sample of approximately 70,000 households per year and serves as the source for estimating official poverty rates.

Unlike the SIPP, the March CPS is not a traditional panel survey, in which a set of respondents are consistently followed over time. Instead, geographic residences are sampled and interviewed on a rotating basis over a period of about a year and a half, regardless of the current occupant. Because the survey is repeated twice in March, however, it is possible to trace a subset of individuals from one year to the next if the individuals are living in the same housing unit in March of both years.
A developed literature exists regarding how to match adjacent years of the CPS March Supplement (see Katz, Teuter and Sidel, 1984; Welch, 1993; Madrian and Lefgren, 2000; Feng 2001; Feng, 2008). Yet there is no method designed to comprehensively and uniformly produce a complete series of matches back to the 1960s. Furthermore, as noted by Welch (1993) every study should use its own matching criteria depending on the parameters to be measured. Matching algorithms typically take the form of identifying all matches based on anonymous survey identifiers and then validating or invalidating these “naïve” matches (Madrian and Lefgren, 2000) based on observable characteristics. Because we are interested in measuring income instability, it would be problematic to condition potential matches on characteristics highly associated with income volatility—for example, having persistent labor force status or occupational classification.

Our matching method is motivated by two goals: (1) maximizing potential matches and (2) minimizing any bias introduced by the matching process. The key yardstick for determining whether matches are invalid is the so-called migration flag, which indicates whether someone has moved in the last 12 months—in which case, the individual should not be able to be matched in the CPS. However, while limiting the number of mismatches is important, it is at least as important for our analysis that mismatch rates be consistent so as not to bias our measure of trends in income instability over a quarter century.

At the same time, the CPS data have undergone many changes over the 1986-2010 period. In particular, the sample size dramatically increased in the early 2000s to allow researchers to study the expansion of children’s health insurance under the State Children’s Health Insurance Program (SCHIP). For the first 3 years of the SCHIP expansion, household identifiers do not uniquely identify households across surveys. Starting in 2005, an additional household identifier is included in the data, which improves the reliability of matches.

Given the foregoing, we developed a matching method that uses a distance-minimization approach to match individuals within households. The household roster in year 1 is joined with the roster in year 2 to produce every possible comparison for individuals from both years. Rather than performing a naïve match based on all available identifiers, in other words, we compare all individuals who share a household ID and find those who are most similar. The distance measure used is a weighted score based on a comparison of state, sex, age, race, the individual's identifying “line number,” and marital status (in order of highest to lowest weight). Line numbers, which are within-household person identifiers (1-39), and marital status are given the lowest weight because (1) we are interested in remedying a problem with line numbers and (2) we do not want to unnecessarily exclude those who experience volatility because of change in family composition.

Using this weighted score, we consider two individuals a valid match if both are each other’s closest “neighbor” in the household. We then perform an additional validation procedure using the magnitude of the distance measure to ensure that the match is actually a close one and not just the best possible within the
household – for example, a one person dwelling will match a 30 year old to a 70 year old unless we put some restrictions on the absolute distance allowed between potential matches. The method of matching has several advantages for our purposes.

First, because we do not require that the line numbers match exactly, we are able to reliably match individuals even in years for which line numbers are absent. By considering all possible comparisons, the distance-based approach allows us to find matches without unique person identifiers. Applying this method consistently allows us to ensure that any bias created by this method is carried through the series and does not only affect years in which line numbers are missing.

Second, the use of distance matching provides an elegant solution to the problem of non-unique household IDs. By comparing everyone across households by household ID, the distance-procedure effectively sorts households based on who is most similar. For continuity, we use this method even when the additional household ID becomes available in 2005. Previous research has attempted to match households in the March supplement to the monthly basic file, as the basic file does not include households who are sampled as part of the expansion. We prefer the distance approach because (1) it is unclear that individuals included because of the sample expansion are truly unmatchable across years, and (2) to the extent that we do not trust the household IDs, we do not wish to use them to match to the basic file.
Third, our matching method produces consistent match and mismatch rates. Figures 1 shows the share of accepted matches (individuals) who indicate moving in the last 12 months. This is the benchmark used by Madrian and Lefgren (2000) as an indicator of mismatch. Since only non-movers should be matchable in the CPS, this should happen only when staggered interviews mean a household is interviewed slightly less than 12 months apart or due to measurement error. As can be seen, our mismatch rate is both low and stable.

Figure 2 shows the share of individuals in year 1 who are matchable to year 2 for each match pair. We typically match about 60-70 percent of individuals who are in year 1 (MIS 1-4) to a verified year 2 (MIS 5-8) record, and like the mismatch rate, the match rate is stable over time. While it may appear that match rates have increased, this is merely a result of declining migration, rather than increasing success of the match procedure over time. The matching algorithm matches a strikingly stable proportion of those who indicate they have maintained the same residence for at least 12 months (see Figure 3). Note that the sharp drop off in match rates beginning in the early 2000s is the result of the SCHIP sample expansion and not any change in the performance of the matching algorithm.

**Fig. 2 Proportion of Year 1 Individuals Matched to Year 2**
II. Income Measurement in the CPS

The ESI’s measure of income is household gross money income. Money income includes earned income (wage and salary income from employment), property and asset income, cash transfer payments (e.g. AFDC/TANF, SSI, Social Security, unemployment benefits, and veterans payments), and self-employment income. It also includes lump-sum and one-time payments, such as catch-up payments from Disability Insurance, and settlements and distributions from retirement accounts, to the extent that respondents report these as income.

As was true previously, the ESI measure of available income is equivalized to reflect the distribution of resources over multiple persons in the household. The ESI uses the equivalence scale recommended by the National Academy of Sciences for poverty calculations, which is designed to account for the sharing of expenses that typically occurs within households and families with different compositions (Citro and Michael, 1995).

Treatment of Imputed Income

In large household surveys like the CPS and the SIPP, direct responses regarding key variables such as income and wealth often go unreported. Under these circumstances, it is common to apply standardized procedures for filling
imputations.” In the case of non-response to income questions, Census Bureau imputation procedures match individuals with missing data (“recipients”) to “donors” with similar characteristics. If no match is found, another match is attempted, requiring fewer characteristics, and so on until all missing items have been filled. In addition, the CPS uses imputation to handle full survey non-response, if the household responded to the CPS March basic survey, by imputing whole records based on characteristics collected in the basic survey.

Because we cannot know whether either of these methods produces reliable estimates of income for these individuals, we exclude these individuals from our analysis. Previous research has dealt with this issue in a variety of ways. Some studies have set thresholds and excluded observations for which more than a certain proportion of income is imputed (see Hertz, 2007; Winship, 2011), while others have excluded observations with any imputations at all (Ziliak, Hardy and Bollinger, 2011). Cameron and Tracy (1998) adopt the approach of excluding only those with imputed earnings, as this is the main focus of their analysis.

The ESI excludes only those individuals whose household heads or spouses have imputed earnings (wage and salary, self-employment, or farm earnings). The rationale behind this choice is that a large proportion of the sample has at least some income imputed, even if it is from a minor source that is unlikely to induce artificial volatility of the magnitude necessary to affect the index. Furthermore, wage and salary earnings are the main source of income for the vast majority of the sample. Finally, it turns out that the index is not especially sensitive to the method of excluding imputed values. The trend and levels of income volatility are relatively similar regardless of the imputation exclusion adopted (see Figure 4).

After dropping households where the head or spouse were subject to earnings or whole imputation, we reweight the sample using propensity scores by state age, race and sex to correct for the fact that imputed individuals have slightly different characteristics from the full sample. We reweight each year of non-imputed matches to match the characteristics of the full year sample, thus simultaneously correcting for imputation exclusions and, to a small extent, the bias introduced by the exclusion of movers from the sample.

**Top-coded Income**

To protect anonymity, income sources in the CPS are top-coded—that is, income levels above some determined amount are censored in the public use data files. Because top-codes occur at the level of income components rather than individuals or households, they impact households at all parts of the income distribution.

The treatment of top-codes in the CPS varies over time. High incomes were given the level of the top-code threshold prior to 1996, after which the Census Bureau imputed a cell mean for the top-coded income sources. Using the cell
means allows us to better estimate the total level of income for individuals thus allowing us to measure more accurately the proportional instability caused by fluctuations in other sources. For the official series, we use cell means generously provided by Jeff Larrimore et. al. (2008). Again, the approach taken has little effect on the index. Figure 5 shows that excluding those whose ESI status changes between the cell mean and truncation approach depresses the trend in insecurity, but only to a very small degree. Similarly, neither the level nor the trend in the ESI was materially changed by top-coding our measure of total household income at the 98th percentile (the highest point in the distribution that could be theoretically affected by the component top-codes in the year in which they have the largest potential effect).
Even substantial drops in income may not result in material hardship if a household has sufficient precautionary savings to buffer the decline. In measuring whether precautionary savings are sufficient—the “adequate financial safety net” of the ESI definition—the ESI focuses on “liquid financial wealth,” that is, wealth that can be easily liquidated to replace lost income. In practice, this is all wealth holdings besides the primary home, personal vehicles, and earmarked retirement savings, including cash, stocks, mutual funds, bonds, and other financial assets, as well as vacation homes and other real estate besides one’s home.

The ESI defines an “adequate financial safety net” as liquid financial wealth sufficient to replace lost income for the typical duration and magnitude of loss experienced prior to a return to pre-drop income. Thus, individuals who experience a 25 percent or greater household income loss are not counted as “insecure” if they have liquid financial wealth equal to or greater than the cumulative loss for a typical individual sharing their socio-demographic characteristics who also experience such a loss. To calculate this cumulative income loss, we use PSID data from 1981 to 2009. Because the PSID switched to biannual data collection after 1997, we use data from odd years only. We calculate the median time until full recovery from an income drop of 25
percent or greater, for groups defined by the size of the income drop, pre-drop income levels, and age. Finally, we calculate the average sum of the losses (the difference between pre-drop income and actual income) for each group, based on observations with median recovery length by group (4 to 16 years). Individuals whose incomes drop by 25 percent or more who have net financial wealth in excess of the threshold amounts for their characteristics (drop size, pre-drop income level, and age) are treated as secure.

Fig. 6 Observed and Imputed Liquid Financial Wealth in the SIPP
The CPS does not ask household members about financial wealth, but we use data on wealth from the SIPP to impute net liquid wealth based on reported household total income, asset income (interest, dividends, and other property income), household size, and the age and race of the household head. Within groups defined by income quintiles, size topcoded at 8 people, race and age measured in 15-year groups, we measure the distribution of wealth as the mean and standard deviation of the 7th root of negative and positive liquid net wealth (which are approximately truncated normal distributions), plus the probability of holding zero, negative, and positive liquid net wealth. We then impute via a cold deck imputation method by drawing a normal variate for each household and imputing liquid net wealth from the appropriate distribution based on household characteristics in the CPS. Applying this procedure to the original data in the SIPP, we can see that the imputed distribution matches the original data quite well, illustrated in Figure 6 for several representative years.

IV. Accounting for Medical Out-of-Pocket Costs

The ESI treats medical out-of-pocket spending (MOOP) as a constraint on alternative spending that reduces available family income. The CPS, like the SIPP sample on which the earlier ESI was based, does not include medical expenditure data necessary to calculate the reduction in available family income caused by MOOP. For this reason, we repeat our earlier imputation procedure to estimate the MOOP burden at the household level. This procedure is described in full detail in our earlier technical report and is briefly summarized here.

To impute MOOP expenditures for the CPS sample, we use two donor datasets: The CEX and the SIPP. The CEX provides us with a long-running estimate of the relationship between medical spending and income, age, and family size. We use this dataset to generate static imputed family-level MOOP for the first year in which each family appears in our dataset.

Additionally, the small number of SIPP years for which MOOP data are available for the same families across years then allows us to assess the dynamic relationship between MOOP in year t-1 and MOOP in year t. This model of dynamic MOOP spending accounts for the effect of prior MOOP spending, as well as changes in both family size and income. Thus, we use the CEX-based imputation procedure to estimate a family’s MOOP in year t-1, and then the SIPP-based dynamic imputation procedure to predict the family’s MOOP in year t after accounting both for the persistence of MOOP and changes in both income and family size.

We have conducted extensive benchmarking of this two-part imputation procedure. Results of those analyses appear in our earlier technical report.
References


