

EXPLORING CROSS-DOMAIN INSTABILITY IN FAMILIES WITH CHILDREN APPENDIX A

DATA AND SAMPLE CONSTRUCTION

Survey of Income and Program Participation

The Survey of Income and Program Participation (SIPP) is a nationally representative, household-based survey used to document monthly changes in a variety of important domains, including economic well-being, family dynamics, and education. Households are interviewed over multiple years and the SIPP provides the most extensive set of data on how the nation's economic well-being evolves over time.

We use the 2008 SIPP panel throughout this analysis. Interviews for Survey Wave 1 of the 2008 SIPP were conducted between September 2008 and December 2008, whereas surveys for the last wave—Panel Wave 16—were administered between the months of September 2013 and December 2013. During each interview, participants were asked questions about the four months preceding the interview. Thus, with each panel wave consisting of four months, these 16 waves of the 2008 SIPP panel provide us with up to 64 months of data with which to track the economic well-being of families with children.

Wave 1 started with a universe of approximately 65,500 housing units. However, after excluding units ineligible for the surveys—for reasons such as vacancies—and non-response, field representatives successfully interviewed about 42,000 housing units at baseline. For subsequent interviews, SIPP interviewers followed individuals identified from the Wave 1 households, and collected data on all the people living with them (typically in person and/or by phone). Accordingly, within this survey design, the SIPP allows us to identify a set of children at the survey baseline (Survey Wave 1) and follow their transitions into and out of various households over the 2008 to 2013 period. Moreover, we use the 2008 SIPP sample weights to produce descriptive statistics that are nationally representative of the children central to our analysis.

Finally, the data are representative of the civilian, noninstitutionalized population living in the United States. The excluded institutionalized population consists primarily of individuals in correctional institutions and nursing homes. For more technical documentation on the design of the SIPP, see the Census Bureau's documentation page: <https://www.census.gov/content/dam/Census/programs-surveys/sipp/tech-documentation/complete-documents/2008/SIPP%202008%20Panel%20Wave%201%20-%20Core%20File.pdf>

Sample Construction

We used several steps to define our analysis sample. We started by using the baseline of the 2008 SIPP—i.e., Wave 1—to identify all households with children less than 13 years old. With the SIPP covering five years, we chose this age threshold to avoid confounding issues associated with children entering young adulthood (reaching the age of 18). With this constraint, we were left with 19,675 children under age 13 in Wave 1 of the 2008 SIPP.

For our descriptive analysis, it was important to track children over time because instability is measured *by changes in status for households with children*. Thus, from our initial restriction, we further limited the sample to households with children less than 13 years old at baseline that also appear in at least eight waves—or 32 months of data. Stated another way, we use all children who remained in the SIPP for at least half of data collection. We chose this threshold to provide a sufficient window for us to capture a wide range of instability measures, which can potentially negatively affect child development.

Using the SIPP children at baseline, the table below shows how they were distributed across a maximum of 16 waves:

Number of Observations				
Child_Panel_Obs	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	660	3.35	660	3.35
2	711	3.61	1371	6.97
3	761	3.87	2132	10.84
4	671	3.41	2803	14.25
5	691	3.51	3494	17.76
6	714	3.63	4208	21.39
7	682	3.47	4890	24.85
8	720	3.66	5610	28.51
9	691	3.51	6301	32.03
10	748	3.80	7049	35.83
11	813	4.13	7862	39.96
12	770	3.91	8632	43.87
13	1108	5.63	9740	49.50
14	1474	7.49	11214	57.00
15	3562	18.10	14776	75.10
16	4899	24.90	19675	100.00

The table can be read as follows: 660 in the first cell of the “Frequency” column indicates that 660 of the 19,675 children identified at baseline—or 3.35 percent—appeared in only one wave. Seven hundred and eleven children appear in only two waves, and so on. Thus, by limiting the sample to children observed in eight or more waves the number of child observations drops from 19,675 to 14,785. Moreover, when further restricting the data to children appearing in 32 or more months (recall, the maximum is 64 months), the child sample falls again to 14,767. We dismiss 25 percent of the sample children in the SIPP by adopting these exclusions; the remaining 14,767 observations form our *Child Panel* for this analysis. Moreover, these exclusions ensure a sufficient period of time for us to monitor the instability measures central to this analysis.

We considered other approaches to constructing the child panel, ranging from a fixed panel of children appearing in all 16 survey waves, to children appearing in at least two waves (the minimum required for any changes to be calculated). Alternative methodologies did not affect the study’s general conclusions.

Finally, it is important to note that the study’s estimates most likely represent a lower bound of instability within the most vulnerable populations—which are largely the lower-education households in this analysis. This is because unstable households are most likely to drop out from the sample due to the volatility of their life circumstances, yet we cannot observe their data after survey exit. If the survey was somehow able to retain these households, we should expect the instability measures the study highlights to be higher than reported. For more information, see the Nonresponse Bias study conducted by the Census Bureau (<https://www2.census.gov/programs-surveys/sipp/tech-documentation/complete-documents/2008/sipp-2008-panel-waves-03-16-nonresponse-bias-analysis-alyis-15.pdf>).

Education as a Proxy for Socioeconomic Status

In defining the baseline categories for analysis, it is vital they are relatively stable over the duration of the data panel. Otherwise, we could introduce a large amount of mismeasurement into our estimations as category designations would be valid for only a portion of the time. In this analysis, we used the maximum educational attainment in the household at baseline as a proxy for the earnings capacity—and socioeconomic status (SES)—of the household. This section outlines the rationale behind our decision.

In the SIPP data, there are two logical measures for SES: educational attainment and household income. However, household income is much more volatile than the maximum educational attainment for the household, as the following table reveals (for children appearing in eight or more SIPP survey waves):

Changes from Baseline | Household Income versus Education

Household Income	Education		Total
	No Change	Change	
No Change	607 4% 55% 7%	496 3% 45% 10%	1,103 7.7%
Change	8,894 60% 65% 93%	4,770 32% 35% 90%	13,664 92.3%
Total	9,501 64%	5,266 36%	14,767

Layout

Frequency

Percent

Row Percent

Column Percent

As displayed above, slightly more than 92 percent of households with children experienced a change in their household income group after baseline—measured as a function of the federal poverty line (FPL)—at some point during SIPP data collection. The results for education were much more stable: the

maximum education in the household remained stable for roughly two-thirds (64 percent) of the sample examined when comparing the value reported at baseline (Wave 1) to those reported in all other waves.

Since education was the more stable option—and strongly correlated with earning capacity and economic security—we used it to define categories of family types to examine in our descriptive analysis. To confirm that different education levels were highly correlated with different income levels in our sample, we also examined this in the SIPP data, where income is defined as a percentage of the FPL and education is defined by four basic groupings: Less Than High School, High School Only, Some College, and College Plus. Findings from this exercise show that higher education levels are strongly correlated with higher levels of household income, and confirm this study design decision:

Correlation of Household Income and Education Conditional Row Distributions

Household Income Group (% FPL)	Household Educational Attainment			
	< High School	High School	Some College	College Plus
< 50%	28.0%	26.4%	34.9%	10.7%
50 to 100	29.2%	27.9%	33.1%	9.9%
100 to 150	19.0%	26.7%	42.0%	12.4%
150 to 200	11.0%	23.1%	45.5%	20.4%
200 to 300	4.8%	15.9%	49.8%	29.5%
300 to 500	1.3%	8.1%	36.6%	54.0%
500% +	0.7%	2.9%	16.2%	80.3%

Note: Percentages greater than 20% are highlighted.

The values reported above are the percentage of the row that fall across the four education groupings. In other words, the first cell of 28.0 percent indicates that, for the Household Income Groups at or below 50 percent of the FPL, 28 percent of the households had a maximum educational attainment level of less than high school. Equivalently, only 10.7 percent of households with income at or below 50 percent of FPL had a household education level of College Plus. Thus, by both examining the distribution of educational attainment conditional on the household income at baseline, and examining where the bulk of the observations reside, we see that income and education are very highly correlated.

DATA LIMITATIONS

In addition to the analysis limitations implicit in descriptive analyses (e.g. the inability to draw causal conclusions), there are three main limitations to the SIPP that may introduce further bias into our estimates.¹ This section draws heavily upon Census Bureau documentation (found here: <https://www.census.gov/content/dam/Census/programs-surveys/sipp/tech-documentation/complete-documents/2008/SIPP%202008%20Panel%20Wave%2001%20-%20Core%20File.pdf>) and the SIPP Quality Profile (found here: <https://www.census.gov/sipp/workpapr/wp230.pdf>).

Sampling Errors

SIPP estimates are based upon samples. Typically, samples are drawn from a much larger population because it is infeasible to conduct a complete census of all residents. The difference between the true population estimate and the sample is known as sampling error.

By comparing the sample estimates produced within the SIPP sampling frame to other external resources, the Census Bureau creates sampling weights to help minimize sampling error. We use these sample weights in all the calculations performed in our analysis.

Non-Sampling Errors

While sampling error focuses on samples that may not accurately reflect the overall population, non-sampling errors pertain to the quality of the data obtained. For example, as outlined by the Census Bureau, non-sampling errors can stem from:

- Missing information, including sample attrition
- Differences in interpretation of questions
- Inability or unwillingness of the respondents to provide accurate information
- Data collection and processing problems
- Biases resulting from memory recall (e.g. seam bias)

The Census Bureau uses a variety of quality control and edit procedures to minimize non-sampling errors. Moreover, the relatively short survey window of the SIPP (four months, whereas most nationally representative surveys are administered annually) helps to mitigate memory recall bias. Unfortunately, missing information cannot be recovered and is a fundamental limitation of all survey data, especially when attrition is non-random.

In addition, it was difficult to differentiate potentially positive shocks in some measures used for this study. Variables with this limitation include:

- Income Decreases: In some cases, parents may decrease work effort to spend more time with their children, or to invest in further education, which could ultimately benefit the family. We could not distinguish earnings losses for these types of reasons.
- Moves: The child residential moves measure may reflect either a child's move to a different household, or a whole household's move to a different residence (in this there may be some overlap with change in family/household composition). While we can tell that the child is moving, we cannot differentiate moves to better neighborhoods or situations that otherwise provide more opportunities for child development.
- Composition Changes: Change in family or household composition entails a change in the number of people in the family unit (related by birth, marriage, or adoption residing together) or in the household (any member). In theory, some compositional changes—such as a grandparent

who brings resources into the household—could positively impact the children central to this analysis. However, we lack sufficient data to say definitively which compositional changes are strictly positive (detailed examination of household rosters and the characteristics of the individuals entering and exiting the households of the SIPP children would provide more information but this was beyond the scope of this study). Thus, we default to the literature finding that any large compositional change is, on average, detrimental to child development and leave the untangling of this complex relationship for future research.

Definitional Limitations

We were also constrained by the data collected by the SIPP, which may be accurate but are not perfectly suited for the measures we created. Two examples illustrate:

- **Compositional Changes:** A second issue related to compositional changes is the difficulty in identifying changes in household and family membership if they occur in rapid succession. For example, suppose a cohabitating spouse leaves and is immediately replaced by another member in the next month. Under the approach we use in our analysis—i.e., the total number of household or family members—we would miss that the composition has changed, while the overall number stayed the same. Without a detailed analysis of household rosters, we cannot estimate how frequently this type of household churning occurs.
- **Paycheck Variation:** For individuals not paid at a monthly level, we should expect that there will be some months with more paychecks than others. For example, a two-week pay cycle implies that there are two months with three paychecks, and 10 with two. Unfortunately, there is no obvious way for us to account for this variation and it could drive a portion of the income volatility in our analysis.

¹ For an additional discussion of limitations and strengths of the SIPP for capturing income volatility among low-income households in particular, see Wolf et al. 2014, p. 401.