Analytic Uses of the Data

Food Stamp Receipt
Characteristics of Spell Data on the First Longitudinal File
Applications of the SPD Longitudinal Weights for Analyses
Person-Level Analysis
Family-Level Analysis
Household-Level Analysis

Bottom

One attractive feature of the Survey of Program Dynamics (SPD) is that it produces ten years of longitudinal data with welfare reform legislation enacted near the middle of those ten years. This places analysts using the SPD in a unique position to evaluate the impacts of welfare reform. While the construction of these data results in many advantages for the researcher, it also introduces special challenges. The combination of data from the 1992/1993 SIPP, the 1997 SPD Bridge, and the 1998-2002 SPD raises several issues of concern for researchers—including recall periods, missing data and varying levels of aggregation. Although these issues are of concern, they are not crippling to would-be analysts.

This chapter contains information on conducting analysis using the SPD longitudinal data. First, we discuss an example using food stamp receipt, addressing data concerns as well as possible solutions to problems encountered. Then, we describe characteristics of spell data on the 1998 SPD longitudinal file. Finally, we provide instruction for various levels of analysis: person-level, family-level, and household-level. An additional example of using the first longitudinal file to measure the effect of welfare reform (Hess 2001) is available on the Internet at this address: <https://www.census.gov/prod/2001pubs/spd2001-1.pdf>.

Bottom

Top

Food Stamp Receipt

Food stamp receipt has decreased since the passing of welfare reform legislation in 1996. The debate remains open for several issues for which an analysis examining individual receipt patterns over time might offer some insight. Two such issues are "cream-skimming" and the effect of time limits. "Cream-skimming" addresses the question of whether or not the declining food stamp caseload was solely driven by individuals with a briefer history of food stamp receipt. Has welfare reform targeted the "easiest" cases or have individuals with persistent food stamp receipt been equally affected? The effect of time limits for food stamp

receipt remains an open question. Are receipt spells becoming shorter? Are individuals "stockpiling" their eligibility or have receipt patterns remained relatively unchanged? The unique structure and timing of the SPD might offer insight to the answers to these and other policy questions. A discussion of a concrete example, such as food stamp receipt, can also illuminate the general data issues of recall periods, missing data and varying levels of aggregation.

The components comprising the SPD (the 1992/1993 SIPP, the 1997 SPD Bridge, and the 1998-2002 SPD), have different recall periods and levels of aggregation. Respondents in the SIPP panels were interviewed three times per year and, as a result, faced a recall period of four months. Respondents in the 1997-2002 SPD are interviewed only once per year and may face recall periods up to fifteen months. Food stamp receipt is asked at the monthly level for both the SIPP panels and the 1998-2002 SPD. These responses may be aggregated by the data analyst to obtain annual totals. The SPD Bridge file has food stamp receipt information at only the annual level. Receipt is summed for the entire year and specific months of receipt are not available. Data from late 1995 is missing from the SIPP panels, and the amount of missing data depends upon the rotation group of the respondent. For more information regarding rotation groups, see the SIPP Users' Guide.

Consider the example of examining patterns of food stamp receipt before and after welfare reform. Individual analysts may opt to focus on total months of receipt per calendar year (in a sense treating each year as one observation) or look at individual months of food stamp receipt (treating each month as one observation). The SPD will report total number of months of receipt per year from 1992 to 2002 but will not differentiate which months receipt did or did not occur. If this level of analysis is sufficient then the only concern faced by the researcher is how to handle the missing data for a portion of 1995. Several options are available. One might simply use the partial count available for 1995 or treat all of 1995 as a missing observation. If one feels that any adjustment or imputation of food stamp receipt severely compromises the quality of the data, this may be the best option. An alternative is to conjecture that data obtained before and after the missing months sheds light as to the likely receipt for the missing months.

For example, suppose an individual received food stamps for all twelve months in 1994 and 1996. If the 1995 data show nine months of receipt with three months of missing data, it might be reasonable to assume that receipt would have occurred during the missing months. Other cases may involve more ambiguity and may require a greater level of an analyst's judgement.

In general, one can think of the following structure:

Let X = total number of months of food stamp receipt in 1994

Let Y = total number of months of food stamp receipt in 1995 (with missing data)

Let Z = total number of months of food stamp receipt in 1996

There are three possible cases:

1. 1. X = Z. If assigning receipt to any, all or none of the missing months can result in X = Y = Z, then adjust the data to make all three equal.

2. 2. X < Z. If assigning receipt to any, all or none of the missing can result in X < Y < Z, then adjust the data to fit that range. Whether the adjusted value of Y is closer to X or Z is left to the discretion of the analyst. One might consider examining receipt totals from 1993 and 1997 to better establish consistent patterns.

3. 3. X > Z. If assigning receipt to any, all or none of the missing can result in X > Y > Z, then adjust the data to fit that range. Whether the adjusted value of Y is closer to X or Z is left to the discretion of the analyst. One might consider examining receipt totals from 1993 and 1997 to better establish consistent patterns.

In cases 2 and 3, if adjustments to Y cannot result in fitting into the desired range of values, then one might consider using the total for 1995 without making adjustments to 1995. Finally, one could probabilistically estimate receipt in any given month and then determine how many missing months are "likely" to have food stamp receipt.

Bottom ▼

Top ▲

Characteristics of Spell Data on the First Longitudinal File

As a longitudinal survey, one of the strong attributes of the SPD is to provide a collection of data that renders itself to the estimation of spell durations for participation in various transfer programs and unemployment. The methodology for spell duration estimates using the SPD data generally entails the following three components:

- Non-sampling errors—particularly the bias induced by the seam phenomenon.
- Definitions of a spell.
- The statistical approaches used for the spell duration estimates.

This section does not discuss the methodology for spell duration estimates per se. The objective of this section is to discuss the characteristics of the spell data on the first longitudinal file associated with the spell duration estimates. The relationship between the spell data on the first longitudinal file and those on the SIPP Panel 1992 and 1993 longitudinal files is also included in the discussion.

All the time-varying data on the first longitudinal file are yearly instead of monthly like those on the SIPP Panel 1992 and 1993 longitudinal files. The yearly data on the first longitudinal file generally cover 1992, 1993, 1994, 1996 (the SPD Bridge), and 1997 (SPD 1998). For example, on the first longitudinal file the variable PAWMONE7 represents the number of months in 1996 in which a sample person received public assistance payments; the variable LKWKSE4 represents the number of weeks in 1994 in which a sample person was looking for work or on layoff from a job.

For the 1992, 1993, and 1994 data, the user can decompose the yearly data on the first longitudinal file into monthly data by linking the sample people back to the SIPP Panel 1992 and 1993 longitudinal files. For 1997 data, the yearly data on the first longitudinal file

can be decomposed into monthly data; however, at present, these monthly data are available to the public only by special request to the Census Bureau. For the 1996 data, the yearly data cannot be directly decomposed into monthly data because the SPD Bridge did not ask the respondents for month by month recalls. Therefore, if needed, the user has to use an analytical approach to decompose the 1996 yearly data into the monthly data based on the monthly data for 1992, 1993, and 1994 on the SIPP Panel 1992 and 1993 Longitudinal Files, and 1997 monthly data from the SPD 1998 available for the cohort of sample people under consideration.

On the basis of the above discussion, if the first longitudinal file is used alone for spell duration estimates, the time unit of a spell duration may be more advantageously expressed in years and then the spell duration treated as a continuous yearly random variable instead of a discrete weekly or monthly variable. For example, "a sample person receiving 23 weeks of public assistance in 1997" will be converted to "a person receiving 23÷52 = 0.4423 years of public assistance in 1997." Similar to the SIPP, the SPD sample data were subject to the preselected starting and ending points for data collection and recall period specified by the sample design. Consequently, the spells reported in the SPD panel (including the SIPP Panels 1992 and 1993) will generally cover the following four situations:

- A spell may start and end during the panel (an uncensored spell—a spell observed at its entirety).
- A spell may start during the panel and be still ongoing at the end of the panel (a right censored spell).
- A spell may start before the beginning of the panel and end during the panel (a left censored spell).
- A spell may start before the beginning of the panel and be still ongoing at the end of the panel (a doubly censored spell).

Since the SPD data collected prior to the SPD Bridge were extracted from the SIPP Panels 1992 and 1993, the SPD spell data inherently carried over a type of non-sampling error commonly referred to as "the seam effect." In the SIPP, the seam is the boundary between the four-month reference periods for interviews in successive waves of the panel. Namely, for participation in various programs, the number of spell starts or stops reported for the four-month recall (reference month one) was substantially higher than those reported for the one, two, or three month recalls (reference months four, three, and two). This is contrary to the expectation that, after the first wave, the distribution for reported spell starts or stops by month of recall is a uniform one—with approximately 25% of spell starts or stops being reported at each month of recall. As indicated in the *SIPP Quality Profile* (1998), the bias in the spell data due to the seam effect is significant in the SIPP panels and cannot be ignored in the spell duration estimates. In the SIPP, the cause of the seam bias in the spell data has not been identified with certainty, but it has been commonly suggested that questionnaire wording and design, length of recall, and the interaction between them play an important role. For the SPD, the seam effect between the combined SIPP Panels 1992 and 1993 and the SPD Bridge, and the SPD Bridge and the SPD 1998 on the spell data have not been studied.

Bottom ▼

Top ▲

Applications of the SPD Longitudinal Weights for Analyses

Each SPD sample person was assigned four weights: two are crude longitudinal panel weights (*LGTPERWT* on the 1997 SPD file, and *LGTPERW8* on the 1998 SPD file); the other two are the refined longitudinal panel weight (*SPDLNWGT*) and the longitudinal annual weight (*ANNUALWT*) on the SPD first longitudinal file. A sample person on the 1997 SPD file, the 1998 SPD file, and the SPD first longitudinal file will have either a positive weight or a zero weight assigned to LGTPERWT, *LGTPERWT*, *SPDLNWGT*, and *ANNUALWT* according to his or her longitudinal interview status (as described in Chapter 5). The SPD first longitudinal file contains annual data for 1992, 1993, 1994, 1996, and 1997 while the 1998 calendar year file contains only annual data for 1997. Therefore, by using the first longitudinal file to obtain data for longitudinal analyses, analysts can avoid the burden of linking files.

On the 1997 SPD file, the original sample members with positive longitudinal panel weights (*LGTPERWT* > 0) collectively provide a crude representation of the characteristics of the noninstitutionalized civilian population in March 1993 (the SPD panel universe) for the time span between 1992 and 1996. Similarly, the original sample members with *LGTPERW8* > 0 on the 1998 SPD file collectively provide a crude representation of the characteristics of the noninstitutionalized civilian population in March 1993 for the time span between 1992 and 1997. The weight, *LGTPERWT* or *LGTPERW8*, of a sample person quantitatively represents the number of people in the survey universe who have the demographic and economic characteristics similar to those of the sample person. To use the *LGTPERWT* or *LGTPERW8* for any estimates covering multiple years requires matching the sample persons on the 1997 SPD file or the 1998 SPD file back to the 1992/1993 SIPP longitudinal files. The crude longitudinal panel weights, *LGTPERWT* and *LGTPERW8*, were produced to be used for preliminary estimates and research at the early stage of the SPD when the SPD first longitudinal file and the refined longitudinal panel weight, *SPDLNWGT* were not available. However, the *LGTPERWT* and *LGTPERW8* are superseded by the *SPDLNWGT* on the SPD first longitudinal file.

On the SPD first longitudinal file, the longitudinal panel weight, *SPDLNWGT*, should be used for any estimates covering multiple years within 1992 to 1997, and the longitudinal annual weight, *ANNUALWT*, should be just for any annual or calendar year estimates. However, the *SPDLNWGT* is also recommended to be used for any annual or calendar year estimates if the estimates do not concern the characteristics of the children born after the first interview of the 1992/1993 SIPP panels. Some caution should be taken when using the *ANNUALWT* for estimating the characteristics of children aged six and less. Because of the approach used to assign the weights to the sample children born after the first SIPP interview, the estimates for the children in this age group are generally 2.2 percent higher than the corresponding 1998 benchmark estimates. By race for the children in this age group, the estimates are 3.6 percent higher than the benchmark estimates for non-Black, and 5.4 percent lower the benchmark estimates for Black children. Since the 1992 data are available only for the sample units from the SIPP Panel 1992 (which is approximately half of the SPD sample size), the weights used any 1992 estimates must be twice the longitudinal weights on the file (i.e., 2 × *SPDLNWGT* or 2 × *ANNUALWT*.) The variances

of the estimates for this year will need to be inflated by two as well. All the weights, *LGTPERWT*, *LGTPERW8*, *SPDLNWGT*, and *ANNUALWT* can be used for the following three levels of analyses:

- Person-level analysis
- Family-level analysis
- Household-level analysis

Since all the four weights can be used in the same manner for above three levels of analyses; without the loss of generality, the discussion of the levels of analysis provided below will be made based only on the longitudinal panel weight, SPDLNWGT on the SPD first longitudinal file.

Bottom ▼

Top ▲

Person-Level Analysis

For longitudinal analysis at the person level, the sample person weights (*SPDLNWGT*) provided on the first longitudinal file can be used directly, as shown in the following illustration. Suppose you want to assess the poverty levels of the people in the SPD panel universe (the 1993 population) before and after welfare reform. The assessment can begin by constructing a transition matrix classifying how many people in the SPD panel universe retained or changed their original (1993) poverty status in 1997:

Poverty Status of People in the SPD Panel Universe (the 1993 population) in 1993 (before welfare reform) and 1997 (after welfare reform).			
		1993 Poverty Status	
		Not in Poverty (denoted by 0_)	In Poverty (denoted by 1_)
1997 Poverty Status	Not in Poverty (denoted by _0)	Cohort 00—People who were not in poverty in both 1993 and 1997 (i.e., stayed out of poverty).	Cohort 10—People who were in poverty in 1993 but were not in poverty in 1997 (i.e., left poverty).
	In Poverty (denoted by _1)	Cohort 01—People who were not in poverty in 1993 but were in poverty in 1997 (i.e., enter poverty).	Cohort 11—People who were in poverty in both 1993 and 1997 (i.e., stayed in poverty).

As indicated in the table above, the people in the SPD panel universe are classified into four cohorts:

- Cohort 00 consists of the people in the SPD panel universe who were not in poverty in 1993 and were also not in poverty in 1997 (i.e., stayed out of poverty).
- Cohort 10 consists of the people in the SPD panel universe who were in poverty in 1993 but were not in poverty in 1997 (i.e., left poverty).

- Cohort 01 consists of the people in the SPD panel universe who were not in poverty in 1993 but were in poverty in 1997 (i.e., entered poverty).
- Cohort 11 consists of the people in the SPD panel universe who were in poverty in 1993 and were also in poverty in 1997 (i.e., stayed in poverty).

Since the panel universe is adequately represented by the original sample persons on the SPD first longitudinal file who have a positive longitudinal panel weight (*SPDLNWGT* > 0), only these sample people need to be considered in estimating the numbers of people in Cohorts 00, 10, 01, and 11. To estimate the numbers of the people in each cohort, identify the family poverty status of the original sample persons (with positive *SPDLNWGT*) in 1993 and 1997 based on the family poverty status indicators (*FAMLISE3* and *FAMLISE7*, respectively). Suppose you define a low income family as "a family with the total family income below the low income threshold." Then, *FAMLISE3*=1 would imply the family is a low income family in 1992, and *FAMLISE7*=1 would imply the family is a low income family in 1993. A person living in a low income family in a given year is in poverty for that year, and not in poverty for that year otherwise. Assign the poverty status of a person as 1 if in poverty and 0 if not in poverty. Based on the above definition of the poverty status of a person, classify the original sample persons (with positive *SPDLNWGT*) as belonging to Cohorts 00, 10, 01, 11—in accordance with their poverty statuses in 1993 and 1997. The estimate of the number of the people in each of the four cohorts in the SPD panel universe can be calculated by summing the weights (*SPDLNWGT*) of the original sample people in the same cohort.

The poverty levels of the people in the SPD panel universe (the 1993 population) before and after welfare reform can be assessed using the estimates of the number of the people in Cohorts 00, 10, 01, and 11. For example, if the estimate of the number of people in Cohort 10 (in poverty in 1993 but not in 1997) is statistically significantly larger than the estimate of the number of people in Cohort 01 (not in poverty in 1993 but in poverty in 1997), then you can infer that more people left poverty than entered poverty after the welfare reform. This suggests that the welfare reform has a positive effect in reducing the poverty level in the pre-welfare-reform population. (The statistical significance test for the comparison can be made using the procedure provided in Chapter 6.)

Bottom ▼

Top ▲

Family-Level Analysis

While families are not defined longitudinally in the SPD, it is feasible to create a time series of family estimates based on these data. For analyses at the family level, the weight (SPDLNWGT) of the sample person who is the reference person of her/his family can be used to represent the weight of that sample family on the first longitudinal file. An illustration would be to suppose that a user wants to estimate the proportions of the low income families in 1994 and 1997 in the SPD panel universe. Based on the above discussion, the user can calculate the estimates based on the six step procedure provided below.

Step 1. Let F94 denote the 1994 estimate of the number of all the families in the SPD panel universe. As discussed above, the weight of a sample family is represented by the weight of

the reference person of that sample family on the first longitudinal file. Therefore, F94 can be expressed as the sum of the weights (*SPDLNWGT*) of all the original sample members with positive weights who were the family reference people in 1994. A family reference person on the first longitudinal file can be identified by the categorical value of the variable *FAMRELE4* equal to one.

Step 2. Let F97 denote the 1997 estimate of the number of all families in the SPD panel universe. In the same manner as Step 1, F97 can be calculated as the sum of the weights (*SPDLNWGT*) of all the original sample members with positive weights who were the family reference people in 1997. A family reference person on the first longitudinal file can be identified by the categorical value of the variable *FAMRELE7* equal to one.

Step 3. Let FL_{94} denote the 1994 estimate of the number of low income families in the SPD panel universe. On the first longitudinal file, a low income family can be identified by the categorical value of the variable *FAMLISE4* equal to one. In the same token as Step 1, the weight of a low income family is represented by the weight of the reference person of that family. Thus, FL_{94} can be expressed as the sum of the weights (*SPDLNWGT*) of all the original sample members with positive weights who were the reference people (*FAMRELE4* = 1) of a low income family (*FAMLISE4 = 1*) in 1994.

Step 4. Let FL_{97} denote the 1997 estimate of the number of low income families in the SPD panel universe. In the same manner as Step 3, FL_{97} can be expressed as the sum of the weights (*SPDLNWGT*) of all the original sample members with positive weights who were the reference people (*FAMRELE7 = 1*) of a low income family (*FAMLISE7 = 1*) in 1997.

Step 5. Let PL_{94} and PL_{97} be the 1994 and 1997 estimates of the proportions of the low income families among all the people in the SPD panel universe, respectively. By definition, PL_{94} and PL_{97} can be expressed in terms of F_{94}, F_{97}, FL_{94}, and FL_{97} (calculated in Steps 1 to 4) as follows.

$$P_{L94} = \frac{F_{L94}}{F_{94}} \qquad P_{L97} = \frac{F_{L97}}{F_{97}}$$

Step 6. A methodology for estimating the standard errors of the estimates F_{94}, F_{97}, FL_{94}, FL_{97}, PL_{94} and PL_{97}, and a methodology for testing the statistically significant difference between PL_{94} and PL_{97} are provided in Chapter 6.

Bottom ▼

Top ▲

Household-Level Analysis

Although households are not defined longitudinally in the SPD, it is feasible to create a time series of household estimates based on these data. For analyses at the household level, the weight (*SPDLNWGT*) of the sample person who is the reference person of the household can be used to represent the weight of that sample household on the first longitudinal file. An illustration would be to suppose that an analyst wants to estimate the 1994 and 1997 proportions of households headed by females with their own children, but

with no spouse present—in a cohort of all the households headed by householders living with relatives in the SPD panel universe. The analyst can calculate the estimates based on the six steps below.

Step 1. Let H_{94} denote the 1994 estimate of the number of all the households headed by householders living with relatives in the SPD panel universe. As discussed above, the weight of a sample household is represented by the weight of the household reference person on the first longitudinal file. Thus, H_{94} can be expressed as the sum of the weights (*SPDLNWGT*) of all the original sample members with positive weights who were the household reference people living with relatives on the first longitudinal file in 1994. A household reference person living with relatives on the first longitudinal file can be identified by the categorical value of the variable RRPE4 equal to one.

Step 2. Let H_{97} denote the 1997 estimate of the number of all the households headed by householders living with relatives in the SPD panel universe. In the same manner as Step 1, H97 can be calculated as the sum of the weights (*SPDLNWGT*) of all the original sample members with positive weights who were the household reference people living with relatives (RRPE7=1) on the first longitudinal file in 1997.

Step 3. Let HF_{94} denote the 1994 estimate of the number of the households headed by female householders with own children but with no spouse present. On the first longitudinal file, a female can be identified by the categorical value of the variable SEX equal to two, a householder (reference person) living with relatives in 1994 can be identified by the categorical value of the variable $RRPE_4$ equal to one, no spouse present in 1994 can be identified by the categorical value of the variable *MARITLE4* not equal to one or two, and having own children in 1994 can be identified by the categorical value of the variable RRPE4 for someone in her household equal to five. Thus, HF94 can be expressed as the sum of the weights of all the original sample members with positive weights who were a female household reference person living with relatives but no spouse present and had own children in 1994.

Step 4. Let HF_{97} denote the 1997 estimate of the number of the households headed by female householders with their own children but with no spouse present. In the same manner as Step 3, HF_{97} can be calculated as the sum of the weights of all the original sample members with positive weights who were a female (SEX=2) household reference person living with relatives (RRPE7=1) but no spouse present (*MARITLE7* 1 or 2) and had their own children (RRPE7=5 for someone in the household) in 1997.

Step 5. Let PH_{94} and PH_{97} be the 1994 and 1997 estimates of the proportions of the households headed by female householder with own children but with no spouse present in a cohort of all the households headed by householders living with relatives in the SPD panel universe, respectively. By definition, PH94 and PH97 can be expressed in terms of H94, H97, HF94, and HF97 (calculated in Steps 1 to 4) as follows:

Step 6. A methodology for estimating the standard errors of the estimates H94, H97, HF94, HF97, PH94 and PH97, and a methodology for testing the statistically significant difference between PH94 and PH97 are provided in Chapter 6.

Top▲