EXAMINATION OF RELATIONSHIPS BETWEEN ACTUAL AND REPORTED CHANGES IN THE SIPP

Lynn Weidman, U.S. Bureau of the Census[1]

1. INTRODUCTION

One set of variables that is important in SIPP and has been examined closely is the receipt of benefits from various state and federal government programs. Burkhead and Coder (1985) showed that a large proportion of the changes between receipt and nonreceipt (hereafter referred to as transitions) were reported as occurring between the last month of a reference period and the first month of the following one. As an example of this pattern see Table 1. It shows the distribution of transitions and non-transitions in food stamp recipiency for the first 16 months of all rotation groups combined in the SIPP 1984 panel. (Only individuals who responded to the first four interviews and received food stamps at some time during this period are included.) Transitions between R = receipt and N = nonreceipt are recorded in the rows labeled RN and NR. The months noting adjoining reference periods are 4th to 5th, 8th to 9th and 12th to 13th, and they have much larger RN and NR counts than the other months.

Weidman (1986, 1987) explored possible relationships of this pattern to demographic characteristics, interview status and imputed receipt. It was determined that for food stamps and social security there is some difference between the reporting of receipt by self-and proxy-respondents, and, in general, a larger proportion of between reference period transitions occurs when at least one of the two months is imputed. However, these effects are of a scale too small to register the magnitude of the observed pattern. A natural conclusion from these results is that frequently people simply report the same receipt status for all months of a reference period, regardless of when a change actually occurred.

As long as the actual number of transitions remains constant, the difficulty in determining this number is due solely to errors in reporting the months of occurrence. Without the aid of auxiliary data it is impossible to estimate the error structure due to reporting for four consecutive months simultaneously, unless assumptions are made to oversimplify the structure. A discussion of the bias in the estimation of transition levels and an approach to measuring it was given by Hubble and Judkins (1986). Currently an administrative record study is being carried out at the Census Bureau in order to estimate the SIPP error structure in reporting receipt of benefits from nine government transfer programs (see Moore (1986)).

Another item of interest is the period of time a person receives benefits from a given program, which will be called a spell. Agencies responsible for benefit programs and people studying the effectiveness of them want to track changes in spell lengths in order to monitor costs, determine effects of changes in qualifications for benefit eligibility, examine the effect of government policies on benefit receipt, etc. To this point, not much effort has been put into the estimation and examination of spell lengths as reported in SIPP. One of the reasons for this is undoubtedly that they require longitudinal estimation through the matching of several waves of data as compared to the simpler cross-sectional nature of transitions. (A wave is a set of reference periods, usually one per rotation group, whose data are processed together.) Also, the error structure of reported lengths is complicated because both the start and end of spell lengths have their own response errors. Estimating a distribution of spell lengths is more complicated than estimating a number of transitions. The possible patterns of changes in these distributions are more complicated than numbers of transitions increasing or decreasing.

AS for transitions, when the distribution of spell lengths is constant, estimation is complicated solely by errors in reporting the starting and ending months. When the number of transitions and distributions of spell lengths change, the estimation problems become more difficult because each month must be allowed to have a different set of parameters associated with it. Our interest is in being able to identify the types of changes that take place and to recognize them as soon as data will allow.

In this paper we present linear models that attempt to represent the relationships between the observed and actual transitions and spell lengths. The parameter vector of unknown values to be estimated is the mean number of a specified transition actually occurring for a rotation group in a specified set of months, or the mean reported number of spells with specified start and end months. These models are obviously oversimplified because they use only the reported data and no other information source. However, since our purpose is to estimate changes and not actual levels, the estimates may not be unreasonable. In the future we hope to combine this approach with the results of the administrative records study of Moore (1986) and the work of Hubble and Judkins (1986) to get improved estimates of the error structure. Estimation of these models will not be pursued here, but we will take a brief graphical look at the comparison of observed and actual values (Figure 2).

In the next section the relationship between reported and actual transitions is discussed and a basic model for reporting transitions is presented. The reporting probabilities are discussed in section 3. Section 4 examines the error structure for the transition models and section 5 extends the modeling approach to reported spell lengths.

2. MODELING TRANSITIONS

A transition between receipt states for "source A" is said to occur in a given month when the answer to the question about whether or not a person received benefits from "source A" is different than it was in the previous month. It is now well-known that a large proportion of these transitions are reported as occurring in the first month of a reference period for each rotation group and smaller proportions are reported in each of the other three months. If the number of transitions is about constant over a period of months, and we sum the numbers of transitions for any single month over the four rotation groups, we get a fairly constant number reported over these months because of the rotation pattern. In particular, Judkins (1986) has compared some SIPP and Food and Nutrition Service data on food stamp recipiency. This indicates that estimates of transitions from all the rotation groups combined follow general trends, but data from the rotation groups shouldn't be used separately.

What can we conclude from this previous work? Most of the transitions are probably reported, but they are often reported in months other than the one in which they actually occur. Let p_{ij} = probability(transition occurs in month j of a reference period and is reported in month i), i,j=1,2,3,4. (Note that a transition can be reported in a reference period other than the one in which it occurs if we let i be < 1 or > 4. However, for simplicity we will assume these probabilities of reporting are 0 and not allow such cases.) Let O_{gm} be the number of transitions of a specified type reported by rotation group g in month m and A_m be the mean actual number occurring for a rotation group in month m. (Assume that all rotation groups exhibit the same behavior.) If we let m=1 be the first month in some wave, then assume that rotation group 1 reports for months 1 to 4, rotation group 2 for months 2 to 5, etc. (See wave I in Figure 1.) Then a simple linear model for the relationship between actual and reported transitions in months 1 to 4 for rotation group 1 is

$$\begin{bmatrix} O_{11} \\ O_{12} \\ O_{13} \\ O_{14} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \end{bmatrix} \quad (1)$$

or $\underline{O}_{1,1,4,} = P_{1,1,4}\, \underline{A}_{1,4} + \underline{e}_{1,1,4}$

where the subscripts on $\underline{R}$, P and $\underline{e}$ represent the rotation group, first month of range of observations being modeled and last month of range, and the subscripts on $\underline{A}$ represent the first and last months of the range. We can look at models of this form for any set of months and any combination of rotation groups. In order to be able to estimate a set of unknown A_m's, we must have observations on more than one rotation group. (Because of the number of parameters in these models.) To represent combined data from several rotation groups define the direct sum $\oplus$ of models of the form (1) as the sums of terms for the same months. E.g., for rotation groups 1 and 2 in wave I we have the probability matrix, say $Q^{I}_{1,2}$,

$$Q^{I}_{1,2} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & 0 \\ p_{21} & p_{22}+p_{11} & p_{23}+p_{12} & p_{24}+p_{13} & p_{14} \\ p_{31} & p_{32}+p_{21} & p_{33}+p_{22} & p_{34}+p_{23} & p_{24} \\ p_{41} & p_{42}+p_{31} & p_{43}+p_{32} & p_{44}+p_{33} & p_{34} \\ 0 & p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

and $\underline{O}_{1,1,4} \oplus \underline{O}_{2,2,5}$ is

$$\begin{bmatrix} O_{1,1} \\ O_{1,2}+O_{2,2} \\ O_{1,3}+O_{2,3} \\ O_{1,4}+O_{2,4} \\ O_{2,5} \end{bmatrix} = Q^{I}_{1,2} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12}+e_{22} \\ e_{13}+e_{23} \\ e_{14}+e_{24} \\ e_{25} \end{bmatrix} \quad (2).$$

This model also represents a particular type of transition -- either R to N or N to R (R = receipt, N = nonreceipt) -- for a particular income source. In general, if we have a model for transitions from S to T ((S,T)=(R,N) or (N,R)), then the population of interest consists of all people who were in state S in the first month making up the pair of observations that determine O_{g1}. Although this population changes over time, there is no reason to expect that the response error pattern changes. This means that the covariance structure between entries of $\underline{e}$ remains the same over time, in the sense that the variance for the ith month of a reference period is the same for each rotation group and reference period, and the same holds for the covariance between the errors for the ith and jth months. This error structure will be discussed more fully after we take a look at the response probabilities p_{ij}.

3. PROBABILITY STRUCTURE

Before we can estimate a set of A_m's, the entries p_{ij} of the appropriate $P_{g,k,\ell}$ must be estimated. In addition to the 16 p_{ij}'s in (1), there are four probabilities of not reporting transitions, denoted by p_{oj}, j=1,2,3,4, where $p_{oj} = 1 - \sum_i p_{ij}$. Without information from sources other than SIPP, the only way we can individually identify the values of these 20 p_{ij}'s is to make assumptions about their relationships with each other. These assumptions can be that the probabilities follow some model that depends on an estimable set of parameters, or that certain restrictions on the p_{ij} hold. Models of the former type for recall loss have been presented in the literature, but they are generally for continuous data such as earnings.

Here we will consider only simple restrictions that lead to reasonable estimates. The first thing to note is that there is a general pattern for all reported transitions. The number for a specific type is about the same in months 2, 3 and 4 of a reference period, but the number in month 1 is more than twice as large as for each of the other months. This makes it fairly safe to assume that the p_{1j}'s and p_{kk}'s are the dominant probabilities. If we further assume that the remaining p_{ij}'s are about equal, we can take them to be zero without much distortion of the reporting patterns. Also, since the reported values are about the same for months 2, 3 and 4, it is reasonable to assume that $p_{22}=p_{33}=p_{44}$ and that $p_{1k}=1-p_{kk}-p_{0k}$, k=2,3,4. What about p_{11} and the p_{0j}? If we know approximate values of the p_{0j}, possibly all the same, from some other source, then we can use them in the estimation process. Without benefit of additional information we must assume that $p_{11}=1$ and $p_{0j}=0$, j=1,2,3,4. This leaves us with only one unknown probability, p_{kk}, to estimate. Under these restrictions, the basic probability matrix for a single reference period as in (1) is

$$P_{1,1,4} = \begin{bmatrix} 1 & 1-p_{kk} & 1-p_{kk} & 1-p_{kk} \\ 0 & p_{kk} & 0 & 0 \\ 0 & 0 & p_{kk} & 0 \\ 0 & 0 & 0 & p_{kk} \end{bmatrix} \quad (3)$$

If there is a wave or set of waves for which the transition numbers (proportions) remain constant, then the unknown value of p_{kk} can easily be estimated. Given the pattern of transition reporting discussed above, look at estimation from a single wave. Let n_i be the total weighted number of transitions reported in the ith month of a wave for all rotation groups. Letting m_a be the mean actual number occurring in each month for a rotation group, $4[1+3(1-p_{kk})]m_a=E(n_1)$ and $4(3p_{kk})m_a=E(n_2+n_3+n_4)$. Replacing expected values by observed frequencies we have the estimates

$$p_{kk}=\frac{4(n_2+n_3+n_4)}{3(n_1+n_2+n_3+n_4)} \text{ and } m_a=\frac{n_1+n_2+n_3+n_4}{16} .$$

(Note that $p_{kk} \leq 1$ if $n_1 \geq \frac{n_2+n_3+n_4}{3}$, which holds for the data of interest.)

Figure 2 shows the reported and actual transitions of combined rotation groups for an example of increase in transitions. For this example we assume that the probability relationships given in (3) hold exactly with no variance in reporting and p_{kk}=.6. On the horizontal scale month 1 is the first month of a wave. The actual number of transitions prior to month 2 is constant at the level denoted by 1 on the vertical scale. There is an increase of .05 of that number in each of months 2 to 5, so that the actual number in month 5 is 1.2 times the original number. The reported number increases in month -1, because for one rotation group a proportion $(1-p_{kk})$ of the increase in month 2 is reported as occurring in month -1. Under the assumed probability relationships between actual and observed transitions, similar reporting happens in all months, causing changes to be reported ahead of the actual occurrence.

4. ERROR STRUCTURE

For each rotation group and reference period assume that the variance structure of $\underline{e}_{g,b,b+3}$ is the same, for each rotation group and reference period where b denotes the beginning month of the reference period for rotation group g, and that $\underline{e}_{g,b,b+3}$ and $\underline{e}_{h,d,d+3}$ are uncorrelated if $g \neq h$ or $d \geq b+3$. From available data for any set of reference periods with constant transition levels we can estimate $\text{Var}(\underline{e}_{g,b,b+3}) = V$. It may be appropriate to simplify the structure of V somewhat by making certain assumptions such as those used to determine the p_{ij}. For example, the variances for the last three months of a reference period are the same, the covariances between month 1 and the other months are the same, or the covariances among months 2, 3 and 4 are all the same. If all these assumptions hold, V has only three distinct parameters.

Regardless of what form of V is appropriate, the A_m's can be estimated using generalized least squares and $\hat{V}$. This requires a set of data separate from that being modeled from which $\hat{V}$ is obtained, or an iterative procedure that estimates $\hat{V}$ simultaneously with the A_m's.

5. MODELING SPELL LENGTHS

As an example, let month 1 be the earliest month reported in wave I and 13 the earliest in wave I+3 one year later. Suppose that the most recent available data is for wave I+3 and that we want to estimate the number of spells of different lengths that begin in wave I and end in wave I+3. Refer to Figure 1 to see the months of interest.

The models are similar to (1), but there are 16 spell periods that can occur between any two reference periods for each rotation group. Let $O_{g,k}^{\ell}$ denote the number of reported spells from rotation group g that start in month k and end in month ℓ. The vector of these values for rotation group 1 in the period we are examining is

$$\underline{O}_{1,1,4}^{13,16} = (O_{1,1}^{13}, O_{1,1}^{14}, O_{1,1}^{15}, O_{1,1}^{16}, O_{1,2}^{13}, \ldots O_{1,4}^{16})' .$$

The corresponding actual values for all rotation groups are A_k^{ℓ} and

$$A_{1,4}^{13,16} = (A_1^{13}, A_1^{14}, A_1^{15}, A_1^{16}, A_2^{13}, \ldots A_4^{16})' .$$

For the probability matrix in (1) add a subscript: S if the probabilities are for reporting the start of a spell and E if reporting the end. Then the probability structure for observing the start and end of spells in the given reference periods for rotation group 1 is obtained by using Kronecker products,

$$P_{1,1,4S} \otimes P_{1,13,16E} = P_{1,1,4}^{13,16} .$$

The resultant model for reporting spells is

$$\underline{O}^{13,16}_{1,1,4} = P^{13,16}_{1,1,4} \underline{A}^{13,16}_{1,4} + \underline{e}^{13,16}_{1,1,4} \quad (4)$$

where $\underline{e}^{13,16}_{1,1,4}$ is the error vector for rotation group 1 corresponding to $\underline{A}^{13,16}_{1,4}$.

Using the form of the probability matrix in (3) with $Q=P_{1,13,16E}$,

$$P^{13,16}_{1,1,4} = \begin{bmatrix} Q & Q(1-p_{kks}) & Q(1-p_{kks}) & Q(1-p_{kks}) \\ 0_4 & Qp_{kks} & 0_4 & 0_4 \\ 0_4 & 0_4 & Qp_{kks} & 0_4 \\ 0_4 & 0_4 & 0_4 & Qp_{kks} \end{bmatrix}$$

where 0_4 is a 4X4 matrix of 0's.

The entry that corresponds to the row for $O^k_{1,i}$ and the column for A^{ℓ}_j is $p_{ijS}p_{k\ell E}$ which is the probability of reporting a spell starting in month i and ending in month k when it actually starts in month j and ends in month ℓ.

So far we have only presented a model for rotation group 1. Using the direct sum over the four rotation groups we get the overall model for spells starting in wave I and ending in wave I+3. This modeling procedure can be extended in the obvious way to cover any set of starting and ending spell months.

Note: Much data is needed to estimate the spell length parameters of these models, more than SIPP could supply due to the low proportion of transitions. Therefore, we should look at modeling groups of spell lengths, e.g., 5-8 months, 9-12 months, etc.

6. FUTURE WORK

The next step in this investigation is to construct data files to use for estimation of the models presented. In addition, we will look at the validity of the assumptions made. This includes examining the sensitivity of the probability assumptions to small changes in probabilities and the suggested variance assumptions for transitions. For spell lengths we must look at the variance structure and possible simplifications. After we are satisfied with the model formulation we will examine ways to combine this approach with others previously undertaken or currently being explored, as well as looking for ways to incorporate auxiliary data in order to develop improved models.

REFERENCES

1. Burkhead, Dan and John Coder (1985). "Gross Changes in Income Recipiency From the Survey of Income and Program Participation." American Statistical Association, Proceedings of the Survey Research Methods Section.
2. Hubble, David and David Judkins (1986). "Measuring the Bias in Gross Flows in the Presence of Auto-Correlated Response Errors." American Statistical Association, Proceedings of the Survey Research Methods Section.
3. Judkins, David. "SIPP Gross Flows: Validation of Food Stamp Turnover." Internal Census Bureau memorandum from David Judkins to Rajendra Singh, October 16, 1986.
4. Moore, Jeff (1986). Background Paper for the Meeting of the ASA/SSRC Joint Working Group on the Technical Aspects of SIPP. U.S. Bureau of the Census, Washington, D.C. October 1986.
5. Weidman, Lynn (1986). "Investigation of Gross Changes in Income Recipiency from the Survey of Income and Program Participation." American Statistical Association, Proceedings of the Survey Research Methods Section.
6. Weidman, Lynn (1987). "Investigation of Possible Causes of Transition Patterns from SIPP." Internal Census Bureau Report, February 1987.

FOOTNOTE

1. This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

TABLE 1

Month-to-Month Transitions: Food Stamps

Receipt Status	1st to 2nd	2nd to 3rd	3rd to 4th	4th to 5th	5th to 6th	6th to 7th	7th to 8th	8th to 9th	9th to 10th	10th to 11th	11th to 12th	12th to 13th	13th to 14th	14th to 15th	15th to 16th
RR	1240	1255	1274	1159	1270	1278	1287	1161	1260	1261	1265	1135	1216	1205	1219
RN	40	47	35	174	26	38	42	167	33	36	29	157	25	44	40
NR	62	54	61	129	46	51	51	123	37	33	40	97	33	54	43
NN	653	639	625	517	652	627	614	519	659	659	655	572	713	684	685

Figure 1

Rotation Group Structure for Waves One Year Apart

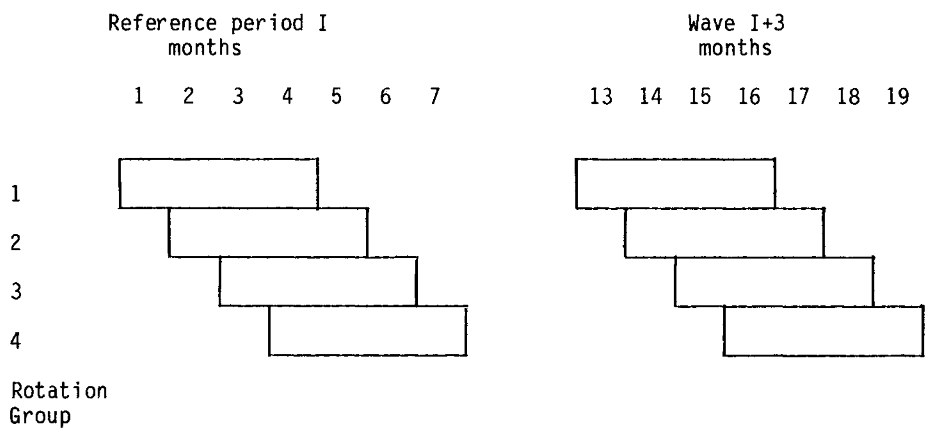

Figure 2

Reported and Actual Transitions