LONGITUDINAL IMPUTATION OF SIPP FOOD STAMP BENEFITS

Antoinette Tremblay, U.S. Bureau of the Census
U.S. Bureau of the Census, Washington, D.C. 20233

Key Words: Item Nonresponse, Panel Survey

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.

INTRODUCTION

"As analysis, good imputation methods can improve naive estimates based on complete or available cases - but bad imputation methods can make matters worse." (Little and Su 1989)

The primary product of the Survey of Income and Program Participation (SIPP) of the Bureau of the Census is longitudinal data. Basically, SIPP collects detailed information on income and wealth, providing a tool for managing and evaluating government transfer and service programs. Data collection results in over a thousand items relating to the economic situation of persons, families, and households in the United States. The survey uses a rotating panel design, with a new panel of sample households being introduced at the start of each calendar year. The sample persons of these households are interviewed eight times, at four month intervals, over a two and a half year period. Each round of interviews for a panel (i.e., four months) is designated a wave (Census 1991). When item nonresponse occurs on a given wave(s), due to answer refusal, data editing, or omission during the interview, it is desirable to make use of data from other waves in the imputation process. This data could come from surrounding waves of the same household/person(s) that exhibit the missing value or from other households. Since this longitudinal information may be highly correlated with the missing value, a reduction of nonresponse bias and an improvement of precision may be realized.

This research examines the current and three alternative longitudinal imputation methods as applied to the SIPP item of food stamp income. The basis of the research is the first four waves of the 1990 SIPP panel. This panel consists of approximately 21,900 interviewed households. Although the most current wave item nonresponse rates (1988 SIPP panel) indicate a relatively low range of 5-8% for food stamp income, it is deemed worthy to be the continuous variable of interest for several reasons. Eight percent of the 1990 total population received food stamp benefits (i.e., 20 million participants) amounting to almost 18 billion dollars of federal, state and local expenditures (U.S. House of Rep. 1992). Also, computer programming for this research is eased considerably since food stamp income can be treated as a household-level variable, as is done operationally.

The primary evaluation criterion is the accuracy of the imputations; that is, the comparison of the actual and the imputed values. A framework is provided for the application of the methodologies of this research to other SIPP items, with higher nonresponse rates.

EVOLUTION OF DATA SETS

The first four waves of the 1990 SIPP Longitudinal Microdata File provide the analysis data, and the person records are accumulated to the household level, as appropriate. These cases are further restricted to only those households where, for the reference person, (a) the interview status for each month is an interview, (b) the interview status for the relevant interview is an interview by self or proxy, (c) the weight assigned for calendar year 1990 is positive, and (d) food stamp income is positive for at least one month across the four waves. Restriction (d) is responsible for the majority of household eliminations. At this stage of data set creation, there are 1,294 household records. The size of the data matrix is further reduced by excluding those variables that, deemed by subject matter specialists, could have little, if any, effect on nonresponse or on the presence or absence of receipt of food stamps. (This step is undertaken primarily for implementation of the flexible matching imputation method.)

The basic steps in the creation of ten simulation data sets consist of reviewing the nonresponse patterns of these households across the four waves, identifying groups suitable for applying a 'missing-at-random' assumption, and then the designation of pseudo-missing values within replicates.

In the review of the nonresponse patterns, 1,265 of the 1,294 households contain either no missing waves of food stamp income data, or merely a single wave of missing data. The remaining 29 exhibit other nonresponse patterns. (Five households have less than one full wave of missing data and 24 households have more than one wave of missing data.) Only full, single wave imputation is addressed in this research. The distribution of the 1,265 households is shown in the first data column of the top portion of Table 1. Of these, 1,164 exhibit no missing waves of food stamp income data, and the remainder are missing one of the four waves. The 1,164 households are used as truth; henceforth, these original households will constitute the actual data set. Since they are the complete cases that are turned into incomplete cases, the ten simulation data sets created from them must mimic the nonresponse pattern of the 1,265 households.

Table 1. Number of HHs Allocated to each Simulation Data Set, by Nonresponse Pattern

Nonresponse Pattern	Truth	Simulation Data Sets										
		1	2	3	4	5	6	7	8	9	10	Σ
No Miss. Wvs	1164	1074	1079	1076	1078	1061	1065	1076	1066	1066	1072	N/A
Miss. Wv 1	20	19	18	24	19	14	12	22	21	16	18	183
Miss. Wv 2	27	21	20	25	24	21	31	24	22	24	21	233
Miss. Wv 3	25	20	21	22	18	31	32	23	30	22	20	239
Miss. Wv 4	29	30	26	17	25	37	24	19	25	36	33	272
TOTAL HHs	1265	1164	1164	1164	1164	1164	1164	1164	1164	1164	1164	N/A
Miss. Wvs 2-3	N/A	41	41	47	42	52	63	47	52	46	41	472
Miss. Wvs 1-4	N/A	90	85	88	86	103	99	88	98	98	92	927

In terms of group identification, logistic regression is performed on the 1,265 household records to uncover any variable(s) that contributes significantly to the nonresponse pattern. The number of elderly persons in the household does, indeed, show that as their number decreases, so does nonresponse, significantly. The RANTBL function of SAS is then accessed ten times to generate deviates from the probability mass function, each time randomly designating a nonresponse pattern to each of the 1,164 households. The 'Simulation Data Sets' section of the top portion of Table 1 presents the count of households, for each simulation data set, by each missing wave pattern. Note that these ten data sets do not exhibit the same number of households across each nonresponse pattern.

At this stage of data set evolution, the ten simulation data sets with 'punched-out' food stamp amounts have been constructed from the original, non-missing, 1,164 households, based on the distribution of the 1,265 households. It is critical to note that via this method of data set creation, after application of an appropriate imputation method, each imputed value has a truth or actual value to which comparison can be made and evaluated.

Differing capabilities of each imputation method require further refinement to these data sets when various accuracy measures are being calculated. Two of the imputation methods have capabilities to impute only for waves with surrounding waves of data; the other two methods are capable of performing imputations for all waves. The lower portion of Table 1 depicts the results. As an illustration, when accuracy measures are calculated from results of either of the two methods capable of imputing for waves 2 and 3, simulation data set 1 contains 41 households; when accuracy measures are calculated from results of either of the two methods capable of imputing for all waves 1 through 4, simulation data set 1 contains 90 households.

IMPUTATION METHODS

Fundamental descriptions of the four imputation methodologies follow. These summaries are not intended to be complete; rather, only those details which are pertinent to this research are presented.

Little and Su Method

The imputed values derived from this stochastic longitudinal imputation method for repeated measures data incorporate information about trend and individual levels. That is, the imputes can be based on row (unit) and column (period) fits to the variable classified by row and column fits. For this analysis, the multiplicative model is appropriate since income amounts are usually modeled on the log scale and must be positive:

imputation = (row effect) x (column effect) x (residual).

The row and column effects are proportional to their respective means. With respect to the variable being imputed, the residual is derived from a complete case, as similar as possible to the incomplete case.

The following methodology is taken from Little and Su (1989), with appropriate variable and index substitutions for this research:

(a) Column (period) effects $c_j = 16 F_j / \sum_k F_k$ are computed for each month j across the four waves, where $j=1,\ldots,16$ and F_j is the sample mean food stamp income for month j based on complete cases.

(b) Adjusted row (unit) means $F^i = n_i^{-1} \sum_j (F_{ij}/c_j)$ are computed for both complete and incomplete cases. The summation is over recorded months for case i; n_i is the number of recorded months; F_{ij} is the food stamp income for case i, month j; and c_j is the simple month correction from (a).

(c) Cases are ordered by F^i, and incomplete case i is matched to the closest complete case, say s.

(d) Missing $\hat{F}_{ij} = [F^i]\,[c_j]\,[F_{sj}/(F^s c_j)]$ where the three terms in square parentheses represent the row, column, and residual effects, the first two terms estimate the predicted mean, and the last term is the stochastic component of the imputation from the matched case.

Flexible Matching Method

The basis of this method resulted from investigation at the Bureau of the Census into imputation methodologies of the American Housing Survey (Long 1992), addressing only cross-sectional imputation. As such, its methodologies are currently used in the March Supplement of the Current Population Survey, also conducted by the Bureau of the Census.

The analysis of food stamp income in this research takes the basic flexible matching imputation method one step further: the combination of cross-sectional and longitudinal imputation. Depending upon the selected variables, potential donors can include cases undergoing wave j imputation as well as those cases with complete information in wave j. The cases themselves undergoing wave j imputation could offer information from the surrounding waves j-1 and j+1.

Simplistically, it is a modified sequential hot deck procedure that matches incomplete cases to complete cases on a hierarchical basis; that is, if an incomplete case cannot be matched with a complete case on a given set of variables, then a variable is dropped and the match is tried again. The hierarchy of matching variables for continuous imputation variables is established by a multivariate forward stepwise regression procedure.

Carry-Over, with Random R, Method

Panel File Longitudinal Imputation for the SIPP is the current method, becoming operational with the 1990 Panel. It is termed the carry-over, with random r, method in this research. Basically, it is a carry-over approach, involving imputing data from the previous and subsequent waves. Simplifying Waite's memo (Census 1994), when imputing for variables with fields for each reference month, the method involves choosing one r at random (r=0,1,2,3,4) for each household. After an r is chosen, the data from the last month before the missing wave is copied into the first r months of the missing wave and data from the first month of data after the missing wave is copied into the remaining 4-r months of the missing wave. A restriction is that the missing wave must be surrounded by interviewed waves, regardless of the number of missing waves in the panel.

Carry-Over, with Population R, Method

This experimental method is a variation of the carry-over, with random r, method. Rather than choosing an r at random, this method determines r based on the distribution of the population. The distribution of across-month changes (i.e., differences greater than zero occurring between monthly food stamp incomes) is derived for the actual values of the population. The RANTBL function of SAS is then accessed to generate deviates from the probability mass function. Thus, based on the distribution of the population, the previous and subsequent non-missing interview data are copied into the appropriate months of the missing wave. As such, this method is also restricted in that the missing wave must be surrounded by interviewed waves.

EVALUATION METHODOLOGY

The first step in the evaluation is to change the level of analysis from monthly food stamp income to the average wave food stamp income. Therefore, each wave of data now consists of one food stamp income for each household, the average over the four months' actual values. Additionally, for those households that underwent imputation, the average over the four months' imputed values also exits. A second step arises from the desire to obtain average results across the data sets, for each imputation method. However, since the simulation does not result in the same number of households missing waves 2 or 3 or missing waves 1 through 4 across each data set, the data sets depicted in the lower portion of Table 1 are combined to create grand measures. For the two carry-over methods, which are capable of imputing for only waves 2 or 3 in this research, the number of households is 472; for the Little and Su and flexible matching methods, which are capable of all-wave imputations, the number of households is 927. Note that when the results from the various imputation methods are being compared to the actual data set, measures for two actual data sets are given: one that is comparable to the number of households undergoing imputation in waves 2 or 3 and the other that is comparable to the number of households undergoing imputation in any of waves 1 through 4.

As is done in Lepkowski and Kalton (1981), the primary evaluation criterion is the accuracy of the imputations; that is, the comparison of the actual and the imputed values. Several descriptive and statistical measures of the actual and the simulation data sets are provided to aid in the quality assessment of the four imputation methods. Many of them are inspired by the work of Lepkowski and Kalton, and all are computed using unweighted data. As stated by Lepkowski and Kalton, the weights in the data set are not appropriate for the calculations since simulation data sets are being used. A brief discussion of the measures follows:

(a) Since the quality of cross-wave imputations depends partly on the correlation between the actual values of the item from one wave to the next, between-wave correlations of food stamp income are derived for the actual data set of 1,164 households.

(b) The actual data set is described statistically, by wave, with and without substitution of the imputes.

(c) Considering only those households undergoing imputation, cross-wave changes are evaluated via mean and standard deviation measures.

For the following measures, it is not necessary to differentiate wave 1 imputed values from those of wave 2, and so on. It is only necessary to ascertain the quality of the imputed value as compared to the actual value, regardless of the wave the imputation occurs in. Therefore, by stacking those households which exhibit any missing wave of food stamp income, wave designation is disregarded.

(d) Considering only those households undergoing imputation, average measures of the mean and standard deviation are tabulated for the actual and imputed data sets. Other statistics include relative biases of the mean, correlations between the actual and each method's imputations, a measure of bias in the mean, and two measures which reflect the total error. The measure of bias in the mean (m_1) is defined as the mean deviation of imputed from actual values for all households with missing values on a wave. The measures which reflect the total error (m_2 and m_3) are both computed for the same cases as m_1 and are respectively defined as the mean of the absolute deviations of imputed from actual values, and the square root of the mean squared deviation of imputed from actual values.

(e) Hypothesis testing is performed to determine if any of the methods result in observations that are significantly different from the actuals.

Additionally, the seam phenomenon is explored to determine if its relative magnitude is reduced by any of the four imputation methods. As detailed in the SIPP Quality Profile (Census 1990), the seam problem is a very perplexing manifestation of longitudinal measurement error, and most SIPP variables which are collected monthly exhibit this phenomenon. The boundary between the four-month reference periods for interviews in successive waves of a panel is designated the seam. The seam phenomenon is a tendency to over-report changes in amounts and status between adjacent months included in the reference periods for different interviews, and to under-report changes between adjacent months covered by the reference period for a single interview. When graphed, the sharp peaks show the clustering of changes at the seam.

RESULTS

The between-wave correlations of actual food stamp income are 0.83 for waves 1 and 2, 0.84 for waves 2 and 3, and 0.83 for waves 3 and 4. This level of stability may provide accurate imputed values for missing data by using longitudinal imputation strategies.

Table 2 provides the wave mean and standard deviation of food stamp income before (i.e., the actual data set) and after imputation, for each method. The differences from the actual are slight for each of the measures, regardless of the method. This may be due to the fact that the imputed values that are substituted into the actual data set account for approximately only 8% of all households.

Table 2. Mean and St. Dev., by Wave, of the 1,164 HHs Before and After Imputations ($)

Source of Values		Mean	St. Dev.
Actual	wave 1	113.88	112.13
	wave 2	114.87	108.82
	wave 3	117.78	110.73
	wave 4	124.25	115.24
Little and Su	wave 1	113.93	112.31
	wave 2	114.97	108.91
	wave 3	117.76	110.58
	wave 4	124.19	115.36
Flexible Matching	wave 1	113.84	111.89
	wave 2	114.86	108.80
	wave 3	117.81	110.82
	wave 4	124.13	115.04
Carry-over, Random r	wave 2	114.96	108.95
	wave 3	117.80	110.64
Carry-over, Population r	wave 2	114.88	108.90
	wave 3	117.87	110.66

An evaluation of cross-wave changes is provided in Table 3. Only those households that undergo imputation are included, and the measures of change for the imputed data sets are computed after imputations are made for the appropriate waves. Since only single-wave imputations are performed, all cross-wave changes are changes from actual to imputed or vice versa. These measures show a bit more contrast than has been seen up to now. For changes across waves 1 and 2, all methods exhibit a higher change than the actual data set. The flexible matching method produces a result which is closest to the actual mean for changes across waves 2 and 3. The change results across waves 3 and 4 are the opposite of those for waves 1 and 2. No real pattern can be discerned when looking at the dispersion of the cross-wave changes.

Table 3. Mean and St. Dev. of Cross-wave Changes ($)

Source of Values		n	Mean	St. Dev.
Wvs 1 and 2	Actual	416	0.83	65.58
	Little and Su	416	4.53	44.78
	Flexible Matching	416	1.41	76.57
	Actual	233	-0.33	68.14
	Carry-over, Rand. r	233	3.87	51.48
	Carry-over, Pop. r	233	-0.23	46.41
Wvs 2 and 3	Actual	472	3.50	57.05
	Little and Su	472	0.80	49.51
	Flexible Matching	472	4.55	66.37
	Carry-over, Rand. r	472	2.02	57.49
	Carry-over, Pop. r	472	5.70	63.41
Wvs 3 and 4	Actual	511	7.14	70.94
	Little and Su	511	6.20	44.21
	Flexible Matching	511	3.60	81.12
	Actual	239	5.36	62.40
	Carry-over, Rand. r	239	4.18	51.85
	Carry-over, Pop. r	239	0.91	52.76

When only the imputed households are examined and all waves are stacked, the carry-over methods provide overestimates of the mean (Table 4). The other two methods produce underestimates of the mean, and to a lesser degree. Also, the Little and Su method overestimates the actual standard deviation. The measures of dispersion of the other methods do not exhibit as striking dissimilarities from the actual.

Table 4. Mean and St. Dev. of Imputed HHs ($)

Source of Values	n	Mean	St.Dev.	Rel.Bias of Mean
Actual	927	135.69	116.98	N/A
Little and Su	927	135.34	120.75	-0.26%
Flexible Matching	927	134.01	116.08	-1.24%
Actual	472	135.75	113.14	N/A
Carry-over, Rand. r	472	138.42	115.74	1.97%
Carry-over, Pop. r	472	138.05	115.01	1.69%

The correlations between the actual and imputed values of each method are interesting, but not overwhelming: 0.75 for the flexible matching method, 0.83 for the Little and Su method, 0.87 for the carry-over, with random r, method, and 0.86 for the carry-over, with population r, method. The flexible matching method, more than the others, imputes a zero value when there is an actual nonzero value. These correlations can be used to determine the amount of explained variation. The flexible matching method explains 56% of the variability among the actual values; the Little and Su method explains 69%; the carry-over, with random r, method explains 76%; the carry-over, with population r, method explains 74%.

M_1 is a measure of the bias in the mean. The average deviation of imputed values from actual values is expected to be zero over many replications if the assumption of missing-at-random holds. In fact, Table 5 indicates that the mean deviation is not strikingly different from zero for any of the methods. Accuracy measures m_2 (average absolute deviation) and m_3 (square root of average squared deviation) reflect each method's total error; the two carry-over methods produce more accurate imputations than the other two.

Table 5. Accuracy Measures($) and Standard Errors(ste)

		Accuracy Measure (ste)		
Source of Values	n	m_1	m_2	m_3
Little and Su	927	0.35 (2.29)	42.66 (1.81)	69.69
Flexible Matching	927	1.67 (2.70)	49.34 (2.16)	82.25
Carry-over, Rand. r	472	-2.67 (2.69)	30.30 (2.31)	58.55
Carry-over, Pop. r	472	-2.30 (2.79)	31.06 (2.39)	60.50

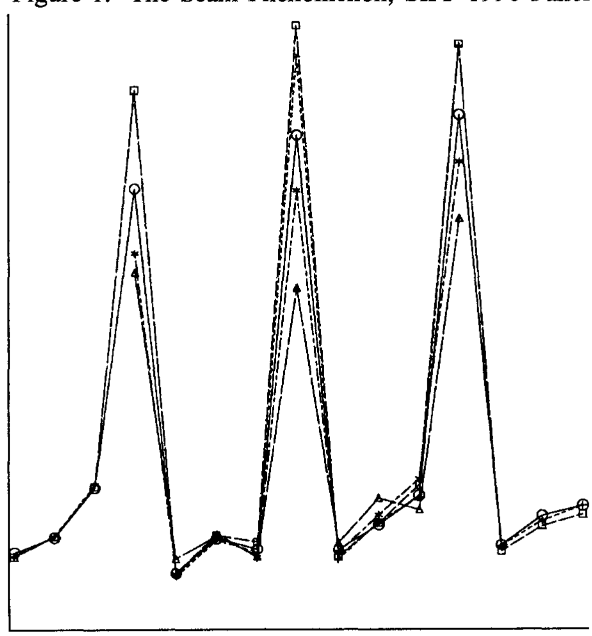
T-tests are performed on just those households undergoing imputation to test the hypothesis that the population mean of the data set before imputations (actual) is the same as the population mean of the data set after each imputation method. Table 6 provides the results. One concludes, simplistically, that the average value resulting from each imputation method is not significantly different from that of the actual.

Table 6. T Statistics

Source of Values	t	d.f.	Prob>|t|
Little and Su	0.06	1852	0.95
Flexible Matching	0.31	1852	0.76
Carry-over, Random r	-0.36	942	0.72
Carry-over, Population r	-0.31	942	0.76

The seam phenomenon is explored to determine if its magnitude is reduced by any of the four imputation methods. As shown in Figure 1, this phenomenon still exists, regardless of the method. The Little and Su method represents the greatest magnitude, followed by that of the flexible matching, the actual data, the carry-over, with population r, and the carry-over, with random r. This manifestation of longitudinal measurement error is 'worsened' by the Little and Su and flexible matching methods.

Figure 1. The Seam Phenomenon, SIPP 1990 Panel



SUMMARY AND CONCLUSIONS

The first four waves of the 1990 SIPP panel provide the data to compare the accuracy of the current and three alternative longitudinal imputation methods, as applied to the item of food stamp income, for single-wave imputation. The Little and Su method is a stochastic longitudinal imputation method that incorporates information about trend and unit levels. The flexible matching method is a modified sequential hot deck procedure. Matched on a hierarchical basis, donors can include the case undergoing imputation, using data from surrounding waves, or cases with complete information for the wave of interest. The carry-over, with random r, method is the SIPP operational longitudinal imputation method. It involves imputing data from surrounding waves. This method, as the name states, fills in data from previous and subsequent waves based on a randomly chosen r. The carry-over, with population r, method differs from the previously mentioned carry-over method in that the choice of r is based on the distribution of the population. Both carry-over methods are restricted in that they can only impute for waves that are surrounded by interviewed waves. For this research, then, these two methods produce imputations for waves 2 and 3 only.

Analysis is completed on the actual data and ten simulation data sets. These simulation data sets are constructed such that they mimic the item's nonresponse pattern. Imputations for each of the four methods are then carried out. Many measures of evaluation are constructed. For example, correlations between the imputed and the actual values, measures of bias in the mean, measures of total error, descriptions of the distribution of cross-wave changes, hypothesis tests, and exploration into the seam phenomenon are provided.

Considering all the results, it is the intent of this research to specify which method is more accurate in terms of the longitudinal imputation of food stamp income, for the 1990 SIPP panel, waves 1 through 4. As discussed in future research, it is hoped that this work will be continued to arrive at the 'best' method for all waves and all longitudinal items.

The bottom line is that none of the four methods 'significantly' shines above the rest; the 'most accurate' method differs across the waves and across the various measures. However, when waves 2 and 3 are considered, either of the two carry-over methods proves to be more accurate than the Little and Su and flexible matching methods, with the random r method slightly preferred. This is good news since the carry-over, with random r, method is the current operational procedure. Considering all four waves, the flexible matching method is inferior to that of Little and Su. Although not mentioned previously, each method's computational ease should also be considered. The carry-over and Little and Su methods are definitely manageable, and the flexible matching method requires relatively difficult programming of the input files.

Further research and analyses into SIPP longitudinal imputation are definitely recommended:

(a) In terms of group identification, logistic regression indicated that the number of elderly persons in the household significantly affects nonresponse. This information was instrumental in the development of the ten simulation data sets. At the time of this research, the type of report (i.e., self or proxy) was overlooked as a possible nonresponse indicator and was not included in the logistic regression. The type of report should be included in future analyses, however, since results from the 1988 SIPP panel show that proxy-type interviews have a significantly higher nonresponse rate than self-interviews. Regardless, it is believed that its exclusion from the creation of the simulation data sets in this analysis does not alter the results.

(b) The two carry-over methods, with population r and with random r, should be modified to handle imputation of waves with only one surrounding wave. Possibly, some type of cross-sectional record matching could be utilized for these waves.

(c) Application of the imputation methods described in this research should be extended to other SIPP items, starting with other federal and state benefit programs, such as Aid to Families with Dependent Children.

ACKNOWLEDGEMENTS

The statistical guidance of Lynne Weidman and Leroy Bailey, and the programming skills of Todd Williams, all of the Bureau of the Census, are appreciated.

REFERENCES

Lepkowski, J., and Kalton, G. (1981), "Investigations of Longitudinal Imputation Strategies for the ISDP Research Panels."

Little, R. J. A., and Su, Hong-Lin (1989), "Item Nonresponse in Panel Surveys," in *Panel Surveys*, eds. D. Kasprzyk, G. Duncan, G. Kalton, M. P. Singh, New York: John Wiley & Sons, pp. 400-425.

Long, K. (1992), "Flexible Matching Imputation in the American Housing Survey," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 221-226.

U.S. Bureau of the Census (1994), "SIPP 1990+: Specifications for Panel File Longitudinal Imputation - REVISION," memorandum from Preston Jay Waite to Sherry Courtland.

U.S. Bureau of the Census (1990), *Survey of Income and Program Participation: Quality Profile*, pp.58-61.

U.S. Bureau of the Census (1991), *Survey of Income and Program Participation User's Guide*, p.2-2.

U.S. House of Representatives, Committee on Ways and Means (1992), *1992 Green Book: Overview of Entitlement Programs*, pp. 1613-1639.