MODELING INCOME IN THE U.S. CONSUMER EXPENDITURE SURVEY

Geoffrey D. Paulin, U.S. Bureau of Labor Statistics, Elizabeth M. Sweet, U.S. Bureau of Census
Geoffrey D. Paulin, DCES, 2 Massachusetts Avenue NE #3985, Washington, D.C. 20212

Key Words: Missing data, Model-based imputation, Missing-at-Random

NOTE: This paper describes the results of research undertaken by the staffs of the U.S. Bureau of Labor Statistics and the U.S. Bureau of the Census. All material contained herein is solely the responsibility of the authors, and does not necessarily reflect the views of either agency. For an unabridged version (working paper), contact the authors.

I. Introduction

The Consumer Expenditure (CE) Survey, sponsored by the U.S. Bureau of Labor Statistics (BLS) and collected under contract by the U.S. Bureau of the Census, contains detailed information on family level spending, demographic characteristics, and income in a series of five quarterly interviews. Currently, consumer units, or "CUs", (see appendix) are divided into two groups: "complete" and "incomplete" income reporters, depending on the respondent's answers to income questions. Although 85 percent of CUs are classified as complete income reporters, even these families do not always provide a complete accounting of all types of income, and the classification can be arbitrary. As a result, these classifications do not completely correct for the problems caused by missing data. For example, many groups are shown on average to spend more than their reported incomes, even though only complete reporters are used to define income classes. It is hoped that imputing data to replace missing income values will improve the quality of the published CE data.

This paper describes modeling techniques currently under joint investigation by BLS and Census as part of a strategy described by Little and Rubin (1987). The work is split between the staffs in order to take advantage of the diversity of available procedures. Using different techniques, each group works on separate models to be merged at the conclusion of the project. A final imputation model including the best results from both strands of research should be attained.

Presumably, all missing income can be imputed eventually. However, wage and salary income is focused on here because it is the most frequently reported type of income; about two-thirds of completely reporting CUs report wage and salary earnings. It is also assumed to be the most accurately reported type of income, since people generally have a good idea of their own (and other members') wage or salary level. This may not be true of other types of income.

II. Preliminary Issues

Before deciding on an imputation strategy, several important issues must be decided: First, are the income data missing randomly, or is there a pattern to non-response? Second, what number of members per CU is appropriate to model? Third, should income be modeled at the member level (and then aggregated), or for the family as a whole?

Definitions of Missingness. The income data are assumed to be missing-at-random (MAR), i.e., the probability of responding to the income question is independent of the level of income, though it may be related to other characteristics. Most important in the decision is work by Crawford (internal BLS memos, 1989-90). Although earlier work by Greenlees, Reece, and Zieschang (1982) finds that income data are not MAR, their work is highly parameterized.

Family Size. In order to work with the cleanest data first, both BLS and Census agreed to start with single-member consumer units and, based on the results, build separate models for two-member and eventually multiple-member CUs. Because single-member CUs have few complications, and multiple-member CUs are highly complex, the primary focus is on two-member CUs. Single-member CUs are discussed for illustration.

Family vs. Member Level Income. Family level income is examined instead of member level income for several reasons. First, the goal is to impute family income, since expenditures are obtained and published at the family level. Second, the error in imputing family income directly is probably less than from summing across imputed member incomes, particularly if incomes are imputed for multiple members. The joint probability distribution between the variables is difficult to preserve in this case. Third, members of the family are assumed to decide how much to work based on how much non-labor income (interest, pensions, Social Security, etc.) is available to them individually, or to the family as a whole. Members also may view each other's incomes (whether from salary or not) as non-labor income. Trying to capture these interactions at the member level can be difficult even in theory; from a practical standpoint, they are often impossible to capture, since many income sources are collected for the family as a whole. But at the family level the the outcome of such interactions is observed. For these reasons addressing family level wage and salary incomes provide an important first step in the modeling procedure; member level incomes will be explored later.

III. The Models

Data. The data come from second interviews occuring between the first quarter of 1988 and the fourth quarter of 1990 for two-member CUs (husband and wife only, a single parent with one child, and other two-member CUs) in which at least one person reported wage or salary income. Although it is first hypothesized that CUs containing a husband and wife only are different than single parent and other families, a Chow test (Kennedy 1992) does not confirm a statistically significant difference between these two groups. The test fails at the 95 percent confidence level, but passes at the 90 percent confidence level. Therefore, dummy variables for single parent and other families are kept in the model.

The regressions are weighted to reflect the population, and their variances are multiplied by 1.44 to account for sample design effect. Of the 2,793 families initially selected, 50 have missing values for at least one independent variable and are dropped in the regression stage. There are 2,743 families included in the regression results.

Variable Selection. Since the goal of imputation is to predict income as accurately as possible, the proposed model in theory contains as many independent variables as may be plausibly related to income. However, to minimize processing costs when the imputation is implemented, both the BLS and Census attempt to find models with maximum predictive power and a minimum of variables.

BLS Procedures. The single-member full model includes numerous independent variables (e.g., age, education, sex, and race). These variables are then tested for inclusion in a reduced model in the following way: First, FSALARYX (the level of salary reported for the family as a whole) is regressed on the full model with Ordinary Least Squares (OLS). At the same time, a stepwise regression is performed on the same model. The results of each are compared. If either procedure finds that an independent variable is statistically significant, that variable is retained for further testing; otherwise, it is removed, unless there is a good reason to keep it. For example, if age is not statistically significant, but the interaction between age and education is, then age is retained. Or if most, but not all, of a group of related dummy variables are found to be statistically significant, all are retained.

The steps described above are repeated until a final reduced model emerges. Then the residuals are examined. Since they invariably are related to more than one independent variable, the absolute value of the residuals (and squared residuals) are regressed on functions of their associated predicted values, i.e., the level of wage and salary income the model predicts for the CU. Results from the residual regressions are then used to weight the final OLS model. This Weighted Least Squares (WLS) procedure helps to correct for heteroskedasticity. When more than one weight looks plausible, a series of tests (Breusch-Pagan, Goldfeld-Quandt, and Park-Glejser) are used to see which weight appears to reduce the problem the most.

The single-member CUs offers the fewest problems in initial variable selection. But for two person CUs, some variables are not so obviously chosen. For example, whose age should be chosen--that of the oldest person, the male (if there is one), or someone else? After some consideration and testing, both the age of the principal earner (i.e., the person whose contribution to family income is the highest) and the other person are included, and the procedures just described are followed.

Census Procedures. Census procedures differ from those of BLS. For example, Census uses a semilog specification instead of WLS to reduce heteroskedasticity. (The advantages and disadvantages of each specification are described in section IV, "Merging Models.") The level of reduction is tested with scatter plots and the Shapiro-Wilkes statistic.

For single-member CUs there are only a

few instances of strong interactions between variables or of variables which need collapsing. But for two-member CUs, it is not clear how member-level variables should be used or transformed into CU-level variables. Model selection and variable creation occur simultaneously since the initial variables are selected arbitrarily.

Initially, the mostly member-level variables are combined and transformed into a set of CU-level variables which are hypothesized to be related linearly to income. In the process high collinearity which sometimes results from including each member's characteristics in the model is avoided.

The ultimate goal is to find a consistent model with as few degrees of freedom and as high an R^2 value as possible. The variables should be approximately orthogonal to each other with respect to income, and make intuitive sense. Some modification to the usual forward selection process is needed because of the lack of a well-defined set of variables relevant to the problem. To achieve this, the "Transformed Main Effects" method (TME) is developed.

The TME method is an offshoot of the forward selection process. The first step of the TME method is to select the variable that produces the highest R^2 value. Here "variable" means a categorical class of variables each of which would ordinarily be a binary variable. Next, the LSMEANS (SAS procedure which adjusts means for unbalanced design) are examined. Categories are collapsed based on t-tests and plausible, intuitive interpretation to create minimum-degrees-of-freedom variable a_i; ideally, i equals one. With a_i in the model, the next strongest variable (b_j) is chosen for entry. This is the variable that produces the highest R^2 which includes an intercept, a_i, b_j, the interaction between a_i and b_j, and an error term (model A). If the interaction term is significant, then LSMEANS are examined for the model A minus a_i and b_j (model B). If the predicting power of models A and B are identical, the interaction term is treated as a categorical main effect variable, and nonsignificant categories within this main effect are collapsed again based on t-tests and intuitive interpretation. If the interaction is not significant, model C (i.e., model A minus the interaction term) is examined. The categories in variable b_j are collapsed based on their LSMEANS and intuitive interpretation, and the categories in variable a_i are reexamined for changes. This process continues until all variables are tested.

In theory, each step adds one degree of freedom to the model; often, though, this does not happen. For example, the strength of the interactions suggest adding several categorical variables at once. What does result is a model with fewer degrees of freedom than the full model, but with a similar R^2. Simultaneously, the effects of multicollinearity are reduced using the TME method, since the newly created variables are often by definition orthogonal to related categorical variables.

There are limitations to the TME method. For example, there is no variable indicating whether on average men receive a higher salary than women. Also, only forward selection can be used. Still, many of the variables chosen match the BLS variables, and had sensible interpretations, indicating the TME method is useful.

IV. Merging Models

Once final results from each method are obtained, the next step is to merge them into one model. One important question is whether to use the WLS or semilog specification. The main advantage of WLS is that the parameter estimates can be interpreted in the usual way. For example, if the equation turns out to be $Y_w = i_w + 5A_w$, where Y_w is weighted income, i_w is the weighted intercept, and A_w is weighted age, one can say that income increases \$5 for every year age increases. Under a semilog model, however, a similar specification, $\ln Y = i + 5A$, means that the *log* of income increases by 5 for every year age increases. Since most people do not think in log terms, the WLS method is more easily understood. (If the parameter estimate on age is small in the semilog case, it can be interpreted as the percent change in income given a unit increase in age. However, of interest here is the change in the *actual value* of the dependent variable Y, or exp(lnY), not the *percent change* in Y or even the change in lnY; also of great interest is how the actual value of Y differs from the predicted value of Y. For example, it is easier to interpret a model where the predicted Y is \$10,000 and actual Y is \$12,000 than it is to interpret one where predicted lnY equal to 9.21 and actual lnY of 9.39, since most people do not think in log terms.) But the semilog model is much easier to use than the WLS method; calculations and tests of various weighting schemes are not necessary with semilog models.

Three experiments are carried out to determine which method to use. The first step in each is to take all independent variables from the BLS and Census final models and put them into the right-hand side of one regression equation whose dependent variable changes with each experiment about to be described.

Experiment 1. Two models are run. The first uses FSALARYX as its dependent variable, as the BLS does. The second uses the natural log of FSALARYX as Census does. The BLS procedure is followed until a reduced model emerges in each case. Residuals from the FSALARYX model are then tested and an appropriate weight is found so that WLS can be run on this reduced model to get final results.

The next step is to decide which reduced model--WLS or semilog--produces the best results. To do this, the predicted values from the WLS model (Y_w's) are calculated. The predicted values from the semilog model are exponentiated to convert them to salary estimates (EXPlnY's.) These values are used to calculate error terms for each observation. In the WLS case, the error terms are:

$$\mathrm{FSALARYX}_i - Y_{wi} = e_{wi}$$

where i indicates each individual observation. For the semilog case, the error terms are:

$$\mathrm{FSALARYX}_i - \mathrm{EXPlnY}_i = e_i$$

The next step is to square the error terms and sum the squares. This grand total is then divided by the number (n) of CUs in each regression. The following results are obtained:

$$\Sigma(e_{wi}{}^2)/n = 487{,}994{,}732$$

and

$$\Sigma(e_i{}^2)/n = 479{,}931{,}419$$

Since $\Sigma(e_{wi}{}^2)/n > \Sigma(e_i{}^2)/n$, the semilog model yields better results. This may be because the semilog model never allows a negative prediction for total wage and salary income, since EXPlnY is always positive. The WLS technique, however, sets no lower bound on predicted income, and indeed some negative wage and salary incomes are predicted. Since these would be set to zero anyway if this method were used for imputation, all negative values for Y_w are converted to zero and the same procedure is followed. Although the WLS numbers improve (i.e., $\Sigma[e_{wi}{}^2]/n$ drops to 486,275,951), the semilog model still appears to be the better approach. Another possible explanation for the superiority of the semilog in this case arises from a subtlety implied by the semilog specification. If the true relationship between income and characteristics is:

$$\ln Y = X_i\beta + \varepsilon$$

then

$$E[\exp(\ln Y)\,|\,X] = E(Y\,|\,X) = E[\exp(X\beta + \varepsilon)\,|\,X] = \exp(X\beta)E[\exp(\varepsilon)]$$

The last term in the equation only holds when $E[\exp(\varepsilon)]$ equals one. Taking the antilog in this way purges some of the error, thus making the semilog perform better in this experiment, even though imputing on exp(lnY) yields biased first moments.

Experiment 2. The independent variables from the reduced WLS and semilog models calculated in Experiment 1 are merged in the same way as the final BLS and Census models are merged. But now Bera-McAleer and PE tests (Maddala, pp. 179-180) are used to test which approach might be preferred.

At first both tests are adapted for WLS by using the WLS predicted values instead of unweighted predicted values. Unfortunately, the results are ambiguous. Both the Bera-McAleer test and the PE test find that the θ_i's are significant in each case. To make sure that the adapting of the tests for WLS is not the problem, the same tests are run with OLS and semilog specifications. The results are similar.

Experiment 3. To normalize the distribution of FSALARYX, a Box-Cox transformation is tested. The optimal value for lambda (λ), found by maximum likelihood estimation, is 3/8. This value (which is confirmed by a nonlinear regression) is particularly interesting because it is almost exactly half way between 1 (i.e., WLS is appropriate) and 0 (i.e., semilog modeling is appropriate).

The BLS process of OLS and stepwise regression is conducted on the transformed values of FSALARYX, and a reduced model is found. To further confirm the transformation is appropriate, Experiment 1 is performed on the Box-Cox results. The Box-Cox results outperform the semilog specification in this test (i.e., $\Sigma[e_i^2]/n$ = 460,250,491 for the Box-Cox).

The superiority of the Box-Cox specification is finally confirmed with the Johnson-McClelland test (1992), a nonparametric specification test designed to find relationships between regressors and disturbance terms. Only under the Box-Cox specification is the null hypothesis of correct specification not rejected.

V. Results

Although a formal imputation mechanism is not yet in place, the results of the Box-Cox transformation are useful to analyze. Table 1 shows results from this model using only valid salary reporters; i.e., at least one person reports a salary amount, and no one has an invalid blank (such as a refusal to answer) for salary. This reduces the (unweighted) sample 5 percent to 2,607 CUs.

The signs for most parameter estimates make sense intuitively. But the signs for the age and education coefficients seem counterintuitive at first. This is because the interaction terms for age and squared age with education are included in the model. However, when the interactions are taken into account, the expected relationships hold in most cases.

Table 2 shows how reported incomes change if model results are substituted for invalid income reports. The variable FINCBTAX is total reported family income. UNBOXCOX is the model-predicted level of family salary. IMPUTED equals UNBOXCOX for each family for whom at least one member has an invalid blank for salary. IMPUTSAL equals IMPUTED for invalid reporters, and FSALARYX for valid reporters. IMPUTINC equals FINCBTAX minus FSALARYX, plus IMPUTSAL. SPENDGAP is the difference between FINCBTAX and approximate annual expenditures (i.e., total quarterly expenditures multiplied by four). Finally, IMPUTGAP is the difference between IMPUTINC and approximate annual expenditures. Although the model is weighted for the population, the unweighted results are shown because the unweighted means are not much different than the weighted means. Therefore, the extra time and expense needed to compute the weighted standard errors overcome the benefits of examining weighted data.

Table 2 shows that even complete income reporters have higher average incomes when salary is estimated from the model. Although the difference is small, the fact that *any* complete reporters need to have salary imputed confirms that the complete and incomplete reporter definitions do not fully correct for income reporting problems. Differences will probably be greater when estimates for other sources are also included. Even so, in each group with less than \$40,000 in income there is at least one family predicted to earn more than \$50,000 when salary *alone* is estimated from the model. And when salaries for incomplete reporters are estimated, the gap between income and expenditures drops sharply--from a \$17,222 deficit to one of \$3,872.

VI. Future Work and Conclusions

More work must be done before final imputation models can be recommended. The next step is to apply the lessons learned so far to multiple-member CUs. Other income sources must also be analyzed. The problem of underreporting (i.e., a family reports less income from a source than it actually receives) may be addressed. This is a difficult issue, since it is not clear exactly how underreporting would be detected.

The MAR assumption needs more investigation, and there are some experiments underway in that area. However, MAR assumptions have yielded new models for examination, which are being tested for predictive accuracy and ease of implementation. These results provide a valuable foundation for further research.

APPENDIX

About the Consumer Expenditure Survey (CE)

The CE Interview sample is composed of over 5,000 consumer units (defined below) per quarter. During the second and fifth interviews, which are conducted under contract by the U.S. Bureau of the Census, the respondent is asked detailed information about work experience and several sources of income for the members of the consumer unit who are at least 14 years old; other sources of income are collected for the consumer unit as a whole. Sources include:

Collected for each member: Wages and salaries; self-employment, including owned farms; Social Security and Railroad benefit checks; and supplemental security income.

Collected for the family as a whole: Unemployment compensation; workers' compensation and veteran's benefits; public assistance and welfare; interest (savings accounts and bonds); regular income from dividends, royalties, estates, or trusts; pensions or annuities from private, military, or other government sources; net income or loss from roomers and boarders or other payments received; regular contributions for support, such as alimony and child support; money income from care for foster children, cash scholarships, and fellowships or stipends not based on working; and food stamps.

Consumer Units. Consumer units (the basic unit of comparison in the CE) are defined as a single person either living alone or sharing a household with others from whom the single person is financially independent; two or more members of a household related by blood, marriage, adoption, or other legal arrangement; or two or more persons living together who share responsibility for at least 2 out of 3 major types of expenses--food, housing, and other expenses. For convenience, "family" and "consumer unit" are used interchangeably in the text.

"Complete" and "Incomplete" Income Reporters. Families that fit one of the following criteria are classified as complete reporters:

1. All major sources of income for each member are reported as zero or valid blank, and at least one member reported a valid, non-zero value for another source of income.
2. The reference person (i.e., the first member mentioned when the respondent is asked to "Start with the name of the person or one of the persons who owns or rents the home") reports zero or valid blanks for all major sources of income, and at least one other member reported a valid, non-zero amount for at least one major source of income.
3. The reference person reported a valid, non-zero amount for at least one major source of income.

Valid blanks result when there is a good reason to leave a question unanswered. For example, if a member of the family did not work at all during the past year, then a valid blank appears for that member's salary earnings. For some sources (e.g., self-employment income) negative amounts can be valid responses.

A family whose reference person reports a major source of income is classified as a complete income reporter even if there are no valid responses for other members. But if there are no valid reports for major sources for the reference person, the family is classified as an incomplete reporter of income, even when all other members have valid responses; hence, the definition can be arbitrary.

Variable Description: Merged Model (Table 1)
*** indicates Census variable.**
Variables with "TWO" or "2" in the name are for non-principal earner.

<u>**Continuous Variables.**</u>
AGE: Age of the principal earner;
AGESQ: Squared age of the principal earner;
EDUCLEVL: Educational attainment of the principal earner, with 0 being no schooling and 18 being at least 2 years of graduate school;
HOURWEEK: Hours per week worked by the principal earner;
WEEKYEAR: Number of weeks the principal earner has worked in the last year;
TM_INTER: Length of interview in minutes;
CPI: Level of the Consumer Price Index in the month of the interview.

<u>**Interaction Terms.**</u>
AGEEDUC: AGE*EDUCLEVL;
AGESQED: AGESQ*EDUCLEVL.
AGE2ED2: AGETWO*EDUCLEV2;
AGE2SQED: AGETWOSQ*EDUCLEV2.

<u>**Dummy Variables.**</u>
RESPFEM: Respondent (respondent) is a female;
PRINERNF: Principal earner is a female;
PRINSAL:* principal earner did not receive a salary.

NUMBER OF SALARY EARNERS:
NOSAL: No member claimed to have earned the *majority* of their income from employment in a wage or salary position in the last 12 months, but some wage or salary income is reported;
TWOSAL: Both members claim to have earned the majority of their incomes from employment in a wage or salary position in the last 12 months;
Control group is family with one salary earner as described above.

FAMILY TYPE:

SINGPAR: Single parent and child;
OTHRFMLY: Other families;
Control group is CU composed of husband and wife only.

OCCUPATIONAL CLASSES:
TECHSALE: Principal earner is in technical/sales work;
PRECPROD: Principal earner is in precision/production work;
OPERATOR: Principal earner is an operative or machinist;
SERVICES: Principal earner is in service work;
Control group is managers and professionals.

REGION OF RESIDENCE:
NOREAST/MIDWEST/WEST: Indicate region in which CU is located;
Control group is located in Southern region.

OWNHOME: CU owns its dwelling.
RURAL: CU is located in a rural area.
FAMIRA: At least one person has an Individual Retirement Account (IRA) or KEOGH;
FALL: CU is interviewed between October and December.
BUSINESS: At least one member has income from self-employment or own farm;
SOCSEC: At least one member has income from Social Security;
PENSION: At least one member has income from private, government, or military pensions or annuities;
INTEREST: At least one member earned interest on savings accounts or bonds;
WELFARE: At least one member received money from supplemental security income, worker's/unemployment compensation, veteran's payments, public assistance, welfare, or foodstamps;
OTHINC: At least one member had net income or loss from other sources (see appendix).

WORKSTAT*:
A=(very good jobs for both persons); CUs where both work full time/full year (ft/fy) and both employers contribute to the pensions
B=(still pretty good jobs) CUs where either both people work ft/fy and one of the persons receives the pension or CUs where one person works ft/fy and the other person works but not ft/fy but both persons' employers contribute to their pensions (i.e. good part time job)
C=employed persons, but with no retirement benefits, CUs where both work ft/fy but neither employer contributes to their pensions
D=CUs where only one person works ft/fy, and only one person's employer in the CU contributes to a pension, or CUs where neither person works ft/fy but both persons' employers contribute to pensions (i.e. so both persons are working)
E=(these CUs do not have as high paying salaries as other CUs) either there is only one person working ft/fy with no employer contribution to pensions orCUs with no persons working ft/fy but there is one employer contributing to a pension
F=(poor salaries but working) CUs where no one works ft/fy and no one's employer contributes to a pension, yet both persons in the CU currently have a job
G=(odd cases)CUs where both members work ft/fy yet at least one valid blank for whether or not the employer contributes to the pension
H=(invalid blank)all CUs where there is a G response for the employment contribution variable
I=(poor salaries and not working: Control group) CUs where no one works ft/fy, neither persons' employer contributes to a pension, and at least one person in the CU currently does not have a job.

PUBLHOUS:* CU lives in public housing.
GOVTCOST:* CU does not live in public housing, but government pays part of cost of housing.

Selected References

Crawford, Sybil. Several internal memos from the Bureau of Labor Statistics, dated 1989 through 1990.

Greenlees, J.S., Reece, W.S., and Zieschang, K.D. "Imputation of Missing Values when the Probability of Response Depends on the Variable Being Imputed," *Journal of the American Statistical Association,* 77(378), June 1982, pp. 251-261.

Johnson, David S. and McClelland, Robert, "Nonparametric Tests for the Independence of Regression and Disturbances," BLS Working Paper 231, Office of Prices and Living Conditions, June 1992.

Kennedy, Peter (1992) A Guide to Econometrics, 3rd edition, Cambridge: The MIT Press.

Little, R.J.A. and Rubin, D.B. (1987) Statistical Analysis with Missing Data, New York: John Wiley & Sons.

Maddala, G.S. (1988) Introduction to Econometrics, New York: Macmillan Publishing Company.

MERGED RESULTS

Table 1. Box-Cox Transformation Results

Dependent Variable: $(Y^{\lambda} - 1)/\lambda$
where Y = FSALARYX
and $\lambda = 3/8$

F Value: 86.204
R^2: 0.6495
Adjusted R^2: 0.6419

Independent Variables	Parameter Estimates	t-Statistics
INTERCEPT	26.328	1.543
AGE	-3.209	-6.119
AGESQ	0.032	5.127
AGETWO	-0.948	-3.156
AGETWOSQ	0.012	3.290
EDUCLEVL	-6.551	-7.787
EDUCLEV2	-1.501	-3.406
AGEEDUC	0.378	8.845
AGESQED	$-3.96*10^{-3}$	-7.820
AGE2ED2	0.099	3.713
AGE2SQED	$-1.20*10^{-3}$	-3.598
HOURWEEK	0.531	10.646
HOURWK2	0.237	4.929
WEEKYR	0.237	5.098
WEEKYR2	0.041	0.830
TM_INTER	0.025	1.690
ROOMSQ	1.075	3.163
CPI	0.519	5.289
RESPFEM	-3.096	-2.986
PRINERNF	-4.868	-3.838
PRINSAL	27.829	8.535
NOSAL	-18.255	-2.176
TWOSAL	13.264	6.063
HASJOB	4.559	2.381
HASJOB2	-0.979	-0.512
SINGPAR	-6.914	-2.043
OTHRFMLY	-3.014	-2.200
TECHSALE	-9.870	-7.198
PRECPROD	-5.490	-2.530
OPERATOR	-11.135	-6.649
SERVICES	-18.250	-9.751
TECHSAL2	-3.453	-2.207
PRECPRO2	-2.096	-0.645
OPERATO2	-7.264	-3.329
SERVICE2	-9.472	-4.160
NOREAST	4.900	3.862
OWNHOME	6.169	4.595
RURAL	-10.313	-6.919
FAMIRA	4.859	3.485
FALL	-3.895	-3.357
BUSINESS	-12.559	-7.108
SOCSEC	-9.504	-4.765
PENSION	-3.046	-1.827
INTEREST	4.903	4.453
WELFARE	-0.865	-0.564
OTHINC	-4.064	-2.480
WORKSTAT(A)	31.300	9.679
WORKSTAT(B)	26.144	9.523
WORKSTAT(C)	24.901	8.723
WORKSTAT(D)	22.395	10.784
WORKSTAT(E)	13.215	7.234
WORKSTAT(F)	4.552	1.593
WORKSTAT(G)	15.185	3.758
WORKSTAT(H)	5.872	2.008
PUBLHOUS	-17.448	-2.244
GOVTCOST	-19.366	-3.169

Table 2. Ranges of Income for Two-Member Consumer Units, 1988-1990

Variable	All Consumer Units	Complete Income Reporters	Under $10,000	$10,000 to $19,999	$20,000 to $29,999	$30,000 to $39,999	$40,000 and Over	Incomplete Income Reporters
Sample Size	3,216	2,780	283	507	541	445	1,004	346
FSALARYX	$28,519	$31,505	$4,005	$10,868	$20,070	$29,692	$56,642	$4,532
Std. Err	529	539	236	255	329	391	1,008	1,496
Minimum	0	0	0	0	0	500	0	0
Maximum	520,000	520,000	44,000	23,157	43,200	44,000	520,000	500,000
UNBOXCOX	$28,046	$29,431	$10,297	$15,148	$22,891	$29,995	$45,312	$16,918
Std. Err	350	375	571	451	495	616	612	726
Minimum	0	0	0	27	11	1640	658	0
Maximum	123,002	123,002	65,710	64,641	73,221	80,951	123,002	71,400
IMPUTED	$16,868	$17,502	$7,575	$14,693	$27,949	$41,444	$46,614	$16,547
Std. Err	686	1,393	1,183	2,120	2,805	$3,980	6,865	757
N	470	158	67	44	28	11	8	312
Minimum	0	0	0	27	11	$22,173	18132	0
Maximum	72,459	72,459	43,457	47,593	58,339	58,150	72,459	71,400
IMPUTSAL	$30,366	$31,920	$5,473	$11,610	$20,489	$29,936	$56,670	$17,882
Std. Err	513	536	354	296	356	410	1,007	1,571
Minimum	0	0	0	27	11	500	351	0
Maximum	520,000	520,000	44,000	47,593	58,339	58,150	520,000	500,000
FINCBTAX	$34,075	$37,504	$5,866	$14,997	$24,862	$34,375	$65,986	$6,528
Std. Err	607	593	230	134	123	136	1,115	2,222
Minimum	-25,920	-25,920	-25,920	10,000	20,000	30,000	40,000	0
Maximum	750,000	524,000	9,983	19,970	29,976	39,943	524,000	750,000
IMPUTINC	$35,923	$37,919	$7,334	$15,740	$25,281	$34,619	$66,014	$19,879
Std. Err	590	591	356	205	179	171	1,121	2,249
Minimum	-25,920	-25,920	-25,920	10,000	14,028	25,558	34,185	33
Maximum	750,000	524,000	52,457	53,004	58,489	59,950	524,000	750,000
SPENDGAP	$5,554	$8,388	-$9,258	-$2,177	$1,422	$6,676	$23,210	-$17,222
Std. Err	472	432	704	399	468	542	915	2,113
Minimum	-159,234	-102,722	-102,723	-67,603	-92,011	-55,763	-99,564	-159,234
Maximum	624,897	275,173	5,410	14,610	21,351	29,667	275,173	624,897
IMPUTGAP	$7,401	$8,804	-$7,790	-$1,435	$1,841	$6,920	$23,238	-$3,872
Std. Err	452	430	706	405	478	543	920	2,089
Minimum	-159,234	-102,722	-102,722	-66,758	-92,011	-55,763	-99,564	-159,234
Maximum	624,897	293,305	23,466	31,192	34,103	37,521	293,305	2,089