Imputation of Medical Out of Pocket (MOOP) Spending to CPS Records

David M. Betson
University of Notre Dame

January 2001

I. Introduction

In their final report, the NRC Panel on Poverty Measurement and Family Assistance recommended numerous changes to the method by which the US Census Bureau measures poverty.[1] The Panel sought to make recommendations that could implemented. One of the Panel's proposals was to subtract from the family's resources the amount of medical out of pocket (MOOP) spending. Given that neither the Current Population Survey (CPS) nor the preferred data set, the Survey of Income and Program Participation (SIPP), collect information on the family's medical spending, a natural question is how well can one impute this needed data from other sources to either the CPS or SIPP? [2]

The purpose of this paper is to examine the current imputation strategy, discuss its potential shortcomings, and report upon efforts to re-estimate the MOOP model on data from the Consumer Expenditure Survey (CEX).

II. Current Imputation Strategy

The current strategy to impute MOOP spending is a two step procedure. First, national control totals for MOOP spending in families headed by an individual under 65 years old (Non Elderly) and those families headed by an individual at least 65 years old (Elderly) are determined. Second, these two aggregate amounts are then allocated to individual CPS families in a manner that reflects the distribution of MOOP spending reported in the National Medical Expenditure Survey (NMES) conducted in 1987 and subsequently 'aged' to reflect spending patterns in 1992.[3]

---

[1] See Citro and Michael (1995).
[2] Data on an individual family's out of pocket medical spending has been collected only in three nationally representative surveys: the National Medical Expenditure Survey (NMES), the Consumer Expenditure Survey (CEX) and currently, the Medical Expenditure Panel Survey (MEPS).
[3] A fuller description can be found in Betson (1998) which is attached to the paper.

To formalize this procedure, let A denote the two age groups where NE signifies the non elderly families and E denotes the elderly families. In the first step, estimates of the total spending in the two age groups are determined. Let C_A denote the estimate of total national MOOP spending for the Ath age group.

The next step is to allocate C_A to the individual family records on the CPS file. This allocation is based upon the distribution of MOOP spending in a secondary data source such as the NMES or CEX. Using, this secondary data source, one can estimate a regression model that describes the distribution of MOOP spending. This regression model can be used to predict a level of MOOP each CPS record. The predicted values for MOOP in the CPS are not the expected value of MOOP spending for the CPS family based upon the estimated regression model. If the expected value of MOOP was used then the variation in MOOP in the CPS files would be smaller than the variation found in the secondary data source because of ignoring the unexplained errors in the imputation. To replicate the entire distribution of MOOP spending in the CPS, this unexplained variation needs to be included in the imputation procedure. This is accomplished by using the regression model to compute the expected value of the family's MOOP spending and then adding the 'unexplained error variance' through the use of a random number generator.

Let m_{fA} denoted the f^{th} family's <u>predicted</u> MOOP spending. The allocation of the national control totals to the individual family records is accomplished by using a proportional raking technique. In other words, the <u>imputed</u> MOOP value for the fth family record would be equal to

$$m^*_{fA} = m_{fA} \times \sum_{i \in A} \frac{C_A}{m_{iA}} = m_{fA} \times S_A \ .$$

The two scaling factors (S_A), one for each age group, are computed by predicting MOOP for each record in the file and taking their sum for each age group. Then the scaling factor is expressed as the ratio of the age's group control to the sum of predicted MOOP values.

A few remarks on the current procedure are in order at this time. The importance of trying to replicate the entire distribution of MOOP spending must be stressed. Too often, the expected value of the variable is utilized for imputation. Given that MOOP spending is to be subtracted from the family's resources, the use of the expected value of MOOP will likely overstate the true proportion of families whose actual MOOP spending would place them in poverty.[4] To avoid this

---

[4] See Betson (2000)

systematic bias in measurement, the imputation strategy must try to faithfully replicate what is known about the entire distribution of MOOP spending of similar families not just the expected amount.

Maintaining the appropriate correlation with other characteristics is equally important. Estimating the total number of individuals and families that are poor when accounting for MOOP spending, capturing the appropriate covariance between income and MOOP spending will be crucial. Accurately estimating the composition of the poverty population will depend upon how well we can reflect the covariance of demographic characteristics such as age and education with MOOP spending.

Finally, the regression approach taken in the current imputation strategy is not the only method to impute MOOP to the CPS or SIPP. Pat Doyle is investigating an alternative strategy utilizing a statistical matching technique known as 'hot decking'. Instead of predicting MOOP spending via a regression model, actual records from the secondary data source are merged onto the CPS or SIPP files. In theory, the imputation of a single variable to the primary data set (CPS or SIPP) via either method should yield approximately the same results. Any differences that occur will be the result of differences in the common variables taken into account via the matching process and the variables used in the regression models. This paper will not attempt to compare the relative merits of these two strategies but will focus upon the regression model approach.

III. A Critical Examination of the Current Imputation Strategy

A. Control Totals

The MOOP control totals were developed to reflect the <u>actual</u> amount of MOOP spending of families of a given age group in a given year. For the moment, let us assume that the primary data set to which we wish to impute MOOP is for the same year as the secondary data survey. Further, let us assume that aggregate amount of predicted MOOP is less than the control total for each age group

$$\sum_{i \in A} m_{iA} < C_A \ .$$

This is not an unanticipated result given that the predicted MOOP values should reflect not only individual MOOP spending but how the families report their actual spending to the surveys. Proportionally raking the predicted MOOP values to the controls implies that the imputation leads to estimates of actual MOOP spending.

The use of estimates of actual MOOP spending in poverty measurement is a mistake. Currently, all other sources of family resources reflect what is reported to the survey. We know that many sources of income are greatly under reported in many surveys. Even the SIPP that has better reporting of income than the CPS, has significant under reporting. If other sources of resources were similarly adjusted for under reporting then the current method would be appropriate. But since they are not adjusted, the use of estimates of actual MOOP in poverty measurement with reported amounts of other family resources will overstate the impact of the subtraction of MOOP spending on poverty counts.

The proportional raking adjustment, S_A, assumes that under reporting of MOOP spending is a constant proportion for all family units. Assuming a constant rate of under reporting for all levels of spending is dubious and most likely leads to overstating actual MOOP spending at the higher levels of spending and understating actual MOOP spending at lower levels.

The discussion to this point has assumed that we are imputing data from one survey to another survey where both surveys are for the same time period. Unfortunately, surveys that target medical expenditures are fielded very infrequently. The National Medical Expenditure Survey (NMES) was conducted only once every ten years with the most recent being in 1987. The Medical Expenditure Panel Survey (MPES) seeks to provide more frequent and hence current estimates of what individuals and families spend on health care. While MPES collects data similar to the NMES, a decision has been made that out of pocket expenditures for medical services, supplies and prescription drugs will not provided to the public with the family's cost of health care premiums. While the files will be made public separately, no identification number will be provided to match families across the two files. The aged 1992 NMES file represents the only specially targeted survey on health care that provides both out of pocket expenditures for medical care and premium payments.

Given that the regression model will be used to impute MOOP spending in years other than the year represented in the secondary data set (NMES), the question is how to reflect the changes in MOOP over time. The effect of changes in the number of individuals and families as well as the socio-economic composition of the population will be reflected in the out year primary data base and their inclusion in the regression model. However, differences in the cost of medical care and how individuals respond to the movement in the relative price of medical care will not be reflected in the predicted MOOP levels.

Even if families do not change their utilization of health care in response to changes in its price, multiplying the predicted MOOP values by the change in the price index for medical care will only crudely reflect how medical cost inflation affects individual families. As medical costs increase, insurance premiums will increase and employers may ask their employees to bear a larger share of their health care utilization (the actual cost or price of health care may rise faster to the family than in the economy). But as the price of utilization rises to the family, the family may choose to utilize less health care.[5] Without further research, it is not clear whether indexing predicted MOOP spending for changes in the cost of health care will over or understate MOOP spending.

Recommendation 1:

Imputation of MOOP spending to the CPS should not control the aggregate imputed amounts to an aggregate control total reflecting actual MOOP spending or administrative estimates of MOOP spending. The only scaling of imputed values from the regression model should be done to reflect differences in the costs of medical care between the time between the year of the primary data set and the secondary data set.

[5] Estimates of the price elasticity of health care demand range from zero to –1.00. See Phelps (1997)

Recommendation 2:

> After periods of health care inflation, the basic imputation should be re-estimated using secondary data from a time period closer to the year of the primary data. This is needed to capture any changes in utilization of health care and shifting of health care costs from employers to families.

B. Regression Model for Allocation

The prediction of MOOP spending levels for an individual family on the CPS has been described as being the result of a regression model. To examine this characterization further, let us for the time being that all families are all similar to each other except that each family has a different level of MOOP spending. Specifically, let us assume that all the individuals have private insurance coverage, are non poor (incomes in excess of 150% of their respective poverty lines, non elderly single white individuals and all have MOOP spending. Given no differences in observed characteristics in the sample, we could assume that MOOP spending in this family group is distributed log normally, in other words,

$$ln(m_f) = \alpha + \varepsilon_f$$

where α is a constant and ε_f is a random normal variable with mean zero and standard deviation σ. Using the sample of households in the secondary data of this type, we could estimate α and σ. We will denote these estimates as *a* and *s* respectively. Next we would proceed to the primary data set and impute to each single with the same characteristics a value for MOOP spending by first drawing a random number from a standard normal random number generator, e_f, for the f^{th} family in the primary data set and imputing

$$exp\left[a + s \times e_f\right].$$

While this would have been the most straightforward way to implement a regression imputation strategy, it could not be used when the NRC Panel first received data from NMES. It was provided in tabular form (the percentage of the sample with a given set of characteristics that had values of MOOP within a given interval). Lacking data on individual families, a different estimation strategy was employed. We assumed that the underlying MOOP spending was distributed as a log-logistic random variable[6] and hence

$$Prob[m \le M] = F[M] = \frac{1}{1 + exp[-(\delta + \phi\, ln(M))]}$$

or alternatively as

$$ln\left[\frac{F[M]}{1 - F[M]}\right] = \delta + \phi\, ln(M)\ .$$

Using the tabular information on families with the same characteristics, we had information on the cumulative probability of MOOP being less than M for various values of M. Based upon this data from the NMES, we could estimate δ and ϕ via OLS. These estimates will be denoted as *d* and *f* respectively.

To impute MOOP values, the first step would be to draw from an uniform random number generator. Let this draw be denoted as u_f for the f^{th} family. This draw represents where the f^{th} household in the MOOP distribution for families with identical characteristics. Given this 'place' in the MOOP distribution, we then compute the value for MOOP that corresponds to this percentile

$$exp\left[\frac{ln\left(\frac{u_f}{1-u_f}\right) - \delta}{\phi}\right]$$

This value is then used as the imputed MOOP value in the primary data set.

This description of the regression approach presents the closest link between this approach and statistical matching via a hot deck method. In a statistical match, one would collect all the observation in the secondary data source that 'close' to the characteristics of the family to which we wish to impute a value in the primary data set and randomly select one of these observation to append to the primary data set. While there is no need for statistical matching to do this, let us assume that the random selection is done in the following manner. First all of the similar observations are sorted with respect to value of MOOP. Then for each observation, the

[6] The log-logistic distribution was initially defined by Shah and Dave(1963) in a manner similar to the definition of the log normal distribution. This citation was found in Johnson and Kotz (1970).

percentage of similar observations with values less than that observation's MOOP is then computed for all observations that are similar to the one you want to impute a value. Then take a random number from an uniform random number generator and pick the observation whose cumulative probability is closest to the random number. This value of MOOP is used for the observation in the primary data base. This is identical to the procedure employed in the log-logistic regression approach where the only difference is the statistical description of the MOOP distribution is used instead of the actual MOOP from the secondary data source.

This discussion has assumed that all families have the same characteristics which clearly not the case. To allow for differences in the characteristics of the families to affect the imputation of MOOP spending, one could estimate separate sets of parameters (α,σ) or (δ,ϕ) for each family type.

This log-logistic regression approach was used for the preparation of the NRC Panel report. After the report was released, problems with the MOOP data were discovered. These problems were documented in Betson, Citro and Michael (2000). A new version of the MOOP data was provided that not only rectified the problems in the earlier data set but also provided the data from individuals observations that were used to compute the earlier tabular information provided to the Panel. Revisions to the log-logistic model are described in Betson (1998). However, when the new data was made available, a complete evaluation of modeling approach was not undertaken. However with the larger degrees of freedom provided by the individual data from the NMES, it is prudent to take a closer look at the regression strategy at this time.

To compare the modeling strategies, we will examine one family type: a white, non-poor, non-elderly single individual with private health care insurance and MOOP spending. In the NMES sample, there are 662 observations for this family type.[7] Examining the distribution of MOOP in this subgroup, we see that it is skewed toward zero with a long upper tail. This observation suggests that the assumption of log normality may be a reasonable assumption.[8] The log normal approach would use the sample to estimate the mean (α) and standard deviation (σ) of the log of MOOP (lnmoop). For this subgroup, the estimates are -.784 and 1.401 respectively. Figure 1 plots the density of lnmoop implied by these estimates with a kernel estimate of the lnmoop distribution in the sample.

Kernel Estimate

Normal
approximation

[7] In the sample, 60 observations of this family type do not have MOOP spending reported. In the next section, we will discuss how we plan to deal these zero observations.

[8] In the remainder of the paper, I will be analyzing the log of MOOP spending where MOOP is expressed in $1,000. Further all of the results in the paper are weighted statistics.

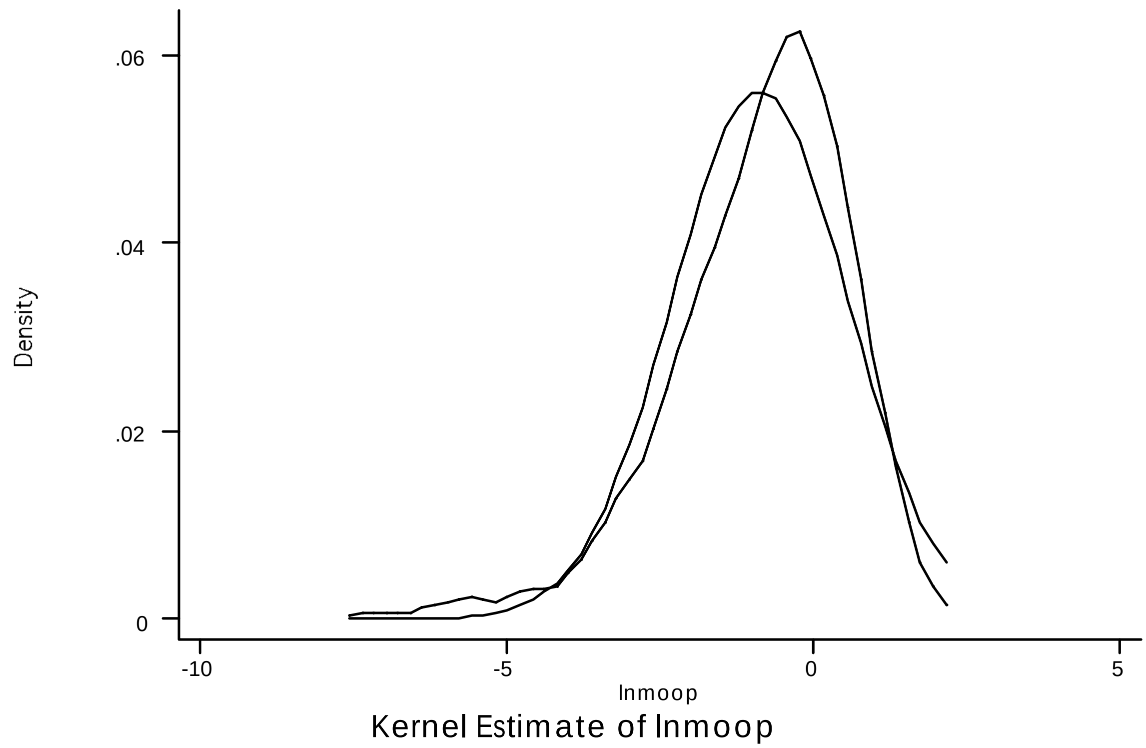


Figure 1

Figure 1 shows that the sample distribution of lnmoop is not normally distributed but is also skewed. The use of the assumption of log normality would lead to imputing too many observations with large values of MOOP spending (note the larger or fatter upper tail of the normal approximation to lnmoop compared to the kernel estimate).

The second approach was to assume that MOOP has a log-logistic distribution. To estimate this model, the log of the ratio of the cumulative probability of MOOP for that value of MOOP over one minus the cumulative probability (lnodds) was regressed against a constant and the log of MOOP (lnmoop)[9]. The results of the regression are reported below.

```
Source |       SS       df       MS                  Number of obs =     661
---------+------------------------------               F(  1,   659) =13438.41
   Model |  2027.57039     1  2027.57039               Prob > F      =  0.0000
```

[9] The value for the cumulative probability for a given observation was computed in the following manner. For each subgroup, the observations were sorted. Then for each observation, the number of weighted observations with a value of MOOP less than or equal to the current observation's value of MOOP divided by the total number of observations was recorded as the cumulative probability.

```
Residual |  99.4291262   659  .150878795             R-squared     =  0.9533
---------+------------------------------             Adj R-squared =  0.9532
   Total |  2126.99951   660  3.22272653             Root MSE      =  .38843

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.226566   .0105808    115.924   0.000        1.20579    1.247342
   _cons |   .9599494   .0176881     54.271   0.000       .9252176    .9946811
------------------------------------------------------------------------------
```

Figure 2 plots the cumulative probability function based upon the sample observations, the log normal estimates described above and the current log-logistic estimates.

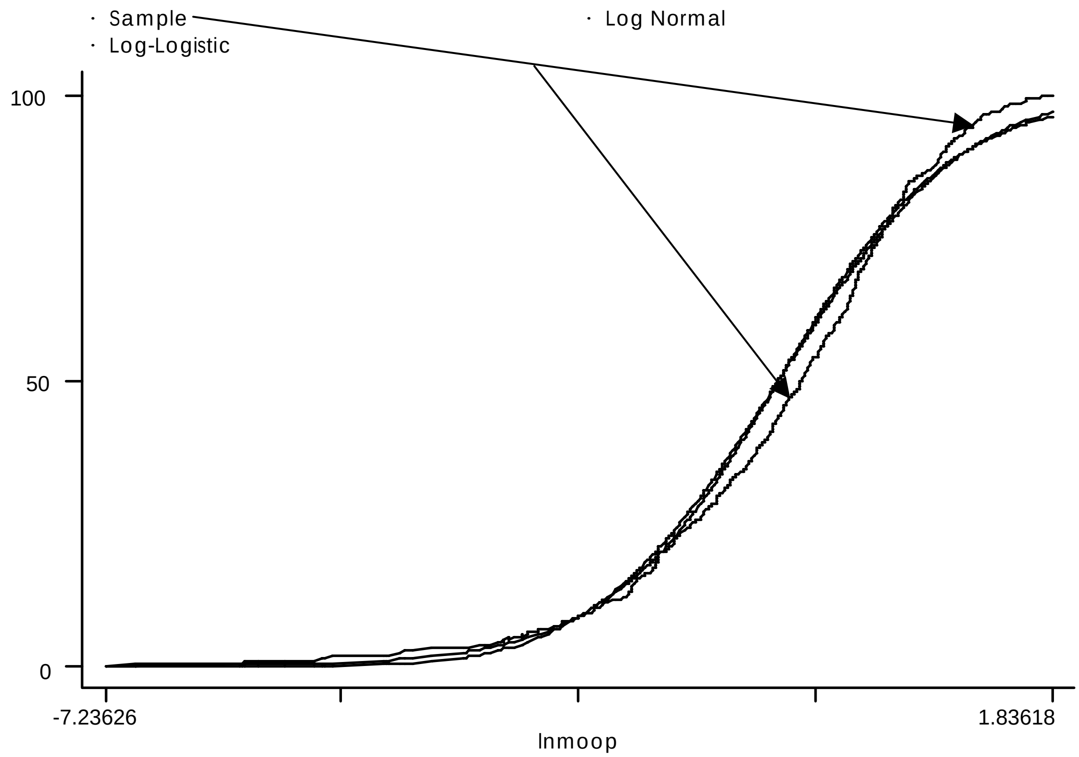


Figure 2

Figure 2 provides two important insights. First, the log normal and the log-logistic assumptions lead to almost identical cumulative probability functions. This result is not unexpected. Johnson and Kotz (1970) note that probit and logit models of discrete choice will lead to very similar results because of the similarity of cumulative probability functions of the normal and logistic distributions. Transforming the basis of the distribution to log scale should not alter this relationship. Secondly, we can conclude that our current strategy of the use of the log-logistic function will lead to too many observations with high values of MOOP spending (note that the

CDFs for the log normal and log-logistic approximations lie below the sample CDF at high values of MOOP).

What can be done to address this problem? The solution will require a better approximation. While this approach is ad hoc, I am suggesting that higher powers of the log of MOOP be included in the regression model. After some experimentation, I am proposing that a cubic approximation be employed. Specifically, the regression model will now be

$$ln\left[\frac{F[M]}{1-F[M]}\right] = \delta + \sum_{n=1}^{3}\phi_n (ln(M))^n$$

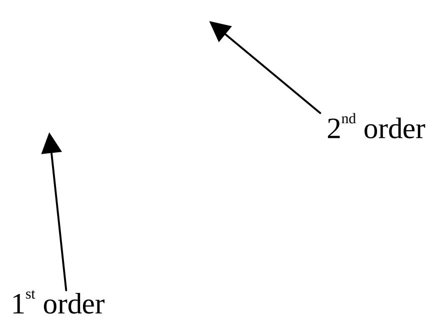
The regression results for this model are presented below

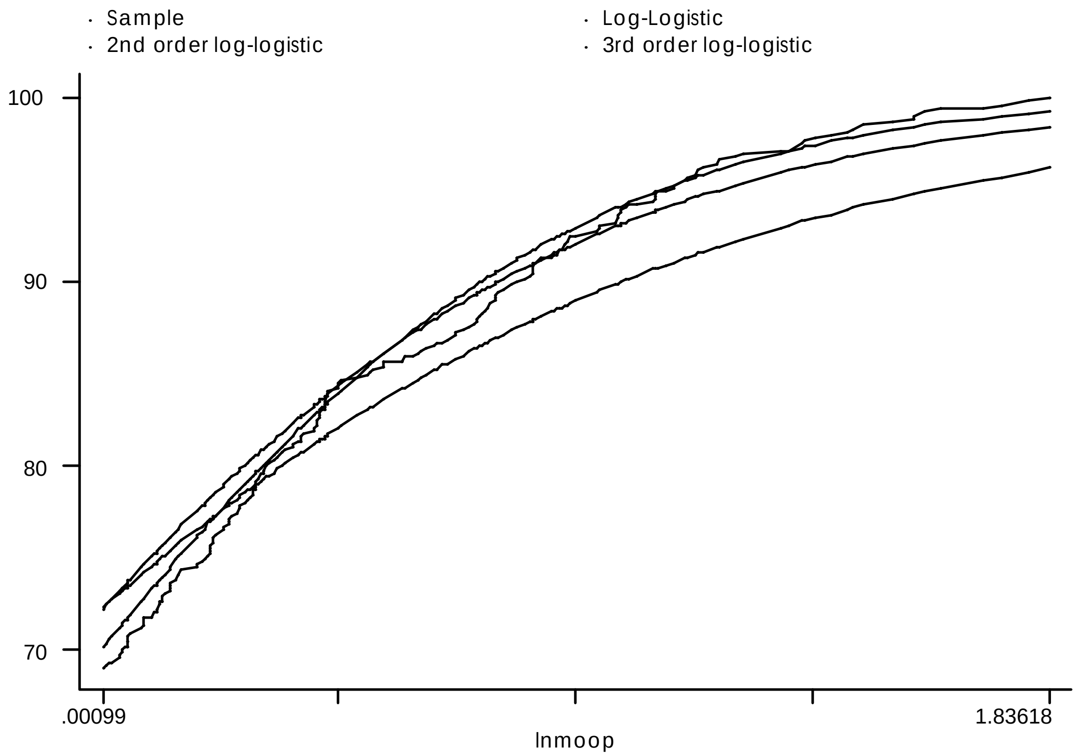
```
Source |       SS       df       MS              Number of obs =     661
---------+------------------------------           F(  3,   657) =41417.33
   Model |  2115.81186     3  705.270619           Prob > F      =  0.0000
Residual |  11.1876545   657  .017028393           R-squared     =  0.9947
---------+------------------------------           Adj R-squared =  0.9947
   Total |  2126.99951   660  3.22272653           Root MSE      =  .13049

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.640223   .0067654    242.442   0.000       1.626939    1.653508
   lnmp2 |   .2358787   .0043474     54.257   0.000       .2273422    .2444152
   lnmp3 |   .0217826   .0006417     33.947   0.000       .0205227    .0230426
   _cons |   .8545995   .0066664    128.195   0.000       .8415096    .8676895
------------------------------------------------------------------------------
```

In general, this approximation to the sample distribution will denoted as a ‘n order log-logistic’ distribution. Figure 3 plots the sample cumulative probability function with the log-logistic (1^{st} order), the 2^{nd} order and the 3^{rd} order log-logistic approximation. This figure focuses upon MOOP spending exceeding$1,000 the top one third of the MOOP distribution.

Figure 3





While employing a quadratic term improves the fit of the cumulative probability function, adding a cubic term continues to improve the fit.[10]

Even with this improvement to the regression strategy, the probability of being in the upper tail of the MOOP distribution is still overstated by the higher order smoothing strategies. My ad hoc recommendation is to limit imputation to be less than the estimated 99th percentile of the estimated MOOP distribution. This can be easily accomplished by limiting the value of the uniform random, u_f, to a maximum value of .99. Hence once the parameters, δ, ϕ_1, ϕ_2 *and* ϕ_3 have been determined for a family type, we would impute to the f^{th} observation of the same type in the primary data set a value of M that solves the following equation

$$ln\left(\frac{min(.99,u_f)}{1-min(.99,u_f)}\right)=\delta+\sum_{n=1}^{3}\phi_n\big(ln(M_f)\big)^n$$

In summary, I would make the following recommendations.

---

[10] A 4th order approximation continues to improve the fit but increase in goodness of fit was judged to marginal. I should note that this was observation was subjective and not based upon any statistical test. The other consideration favoring the cubic approximation is that explicit solutions exist for cubic equations while they do not for 4th order equations. This will simplify the imputation procedure by not requiring numerical techniques for solving for M given a value of u_i.

Recommendation 3:

> A 3rd order log-logistic approximation to the cumulative probability used to describe the distribution of MOOP for subgroups of the population.

Recommendation 4:

> When imputing values to the primary data set, MOOP values be limited to the lower 99% of the estimated MOOP distribution.

C. Dealing with Zero MOOP

We have focused upon imputing MOOP to those observations with MOOP spending. However, not all observations in the NMES sample have MOOP spending. To impute MOOP to all of the observations in the primary data set would be wrong. While estimation problems akin to sample selection bias issues are most likely present, these issues are going to be ignored. Assignment of a non zero MOOP amount to observations will be based upon the proportion of a family type that have reported MOOP in the secondary data base. Random assignment will utilize this estimated proportion in conjunction with a draw from an uniform random number generator. If P is the proportion of the secondary data base of a given family type then a non zero MOOP level will be assigned to the f^{th} observation in the primary data base if

$$v_f < P$$

where v_f is a draw from uniform random number generator. Otherwise, a zero value for MOOP will be assigned.

D. Qualified Medicare Benefit (QMB) and MOOP

Individuals who qualify for Medicare and have incomes less than 100 percent of poverty, the Medicare program waives all cost sharing provisions and Part B premiums. For Medicare eligible individuals between 100% and 120% of poverty, Part B premiums are waved. These benefits are referred to as the Qualified Medicare Benefit (QMB). This benefit was implemented in 1990. In the current imputation procedure that uses the NMES, Part B Medicare premiums were not included in the definition of MOOP. Hence all elderly individuals are accessed a Part B premium unless they report receiving Medicaid. This procedure does not take into account the QMB portion of Medicare and leads to an overstatement of MOOP spending for this portion of the elderly population. A simple solution will be to add a Part B premium only for those elderly individual's income exceeds 120% of poverty.

QMB also waives the cost sharing provisions of Medicare eligible medical services and supplies. Given that the NMES is based upon 1987 data aged to 1992, it is doubtful that aging procedure took this provision into account. While the current imputation imputes no MOOP for elderly individuals reporting the receipt of Medicaid, not all poor elderly receive Medicaid. Hence for these individuals, the current procedure overstates their MOOP spending due to the QMB. However, to include zero MOOP for all poor elderly would also be wrong since Medicare accepts not all medical expenses. The largest single exception is prescription drugs. Since the current NMES data does not separate MOOP spending on drugs, I have chosen to continue the current practice of imputing MOOP spending to all poor elderly who do not report Medicaid.

Recommendation 5:

For those individuals over 65 years old living in a family whose income is less than 120% of poverty, no Medicare Part B premiums will be assigned.

E. Concluding Remarks

One conclusion that could be drawn is that the estimation and imputation strategy currently employed by myself and the Census Bureau produces too many observations with relatively large values for MOOP. In this section, I have proposed five recommendations aimed at improving the imputation of MOOP throughout the entire distribution. In this next section, I will discuss my re-estimation of the model on the NMES. The following section reports upon a comparison of various imputation approaches using the March 1993 CPS.

IV. Re-fitting the Model on NMES data

Before proceeding to estimate the imputation model on more recent data, I thought it would be instructive to re-fit the modified model on the economic and demographic aged NMES data.

In the previous section, the case was made for the inclusion of squared and cubed terms of the log of MOOP in the model. That is the approach that will be taken in this re-estimation.

In the former version of the model, 36 separate family types were constructed for the non elderly population. These groups were based upon the insurance coverage, the family size, poverty status, and race of the family. The elderly population was subdivided into 8 groups based upon age, family size and poverty status. For each of these 42 groups, the cumulative probability was constructed by sorting the observations and computing the percentage of the group that had MOOP spending less than the observation. The cumulative probability was then transformed into the log 'odds' that is the dependent variable of the regression analysis. Previous analysis of the data were separately performed on the non elderly and elderly samples. This analysis allowed for only the main effects of the group's other characteristics to affect the estimation of the intercept *(δ)* and slope coefficients *(ϕ)*. All interaction effects between characteristics were assumed to be zero. In retrospect, this was an unfortunate assumption. Significant interaction effects where found when the 1st order log-logistic model was recently re-estimated. This lead to separate estimates of the model for each of the 42 groups. The regression estimates for the 3rd order log-logistic model are reported in Appendixes A and B.

Since the imputation of zero MOOP values has not changed, the previous estimates of the probability of having MOOP spending will be used.

V. Comparison of MOOP Imputations on the 1993 CPS

In this section, I will report upon a Monte Carlo experiment I conducted to empirically examine the consequences of the various recommendations that I have proposed. Since the NMES data represents 1992, the choice of the March 1993 CPS was ideal since the imputation would not require any out year projections. I chose three alternative imputation implementations that were the following:

> ***Original Imputation:*** This is strategy that I have employed and forms the basis of the Census Bureau's imputations. This strategy uses a proportional rake to established national totals. For 1992, the control totals were $153 billion for the non elderly population and $55.5 billion for the elderly non Part B premium MOOP. The regression model was estimated for the non elderly and elderly populations separately as described in the previous section.[11] Finally, limitation were made on MOOP imputations. The maximum MOOP for a non elderly family was $8,200 while $18,000 for an elderly family. These limits represent the 99th percentile of the two populations and were provided by Pat Doyle.
>
> ***No Control Totals:*** This implementation was identical to the previous one except that no raking was performed to 'hit' the control totals.
>
> ***New Implementation:*** This implementation reflects recommendations 1,3, 4, and 5 made earlier. The 3rd order log-logistic model was estimated for each of the 42 different family types. Limits were placed on the maximum MOOP that was assigned. No family was assigned a MOOP that exceed the 99th percentile of the MOOP distribution for their respective family type. Elderly adults living in families whose income is less than 120% of poverty were not assigned Medicare Part B premium. And no raking was performed to achieve a control total.

For each of the three implementations, I performed 100 MOOP imputations to the entire March 1993 CPS.[12] The first variable that I examined was the mean MOOP (includes both zero and positive values) in each of the two age groups. The following table presents the Monte Carlo results for the simulations as well as the averages from the NMES (secondary file).

---

[11] See Betson (1998) for more a detailed description of the regression model and estimates. This paper is attached.

[12] Appendix C contains the FORTRAN source code for the new imputation routines.

Average MOOP in:	Non Elderly	Elderly
NMES	$1,432	$2,304
Original Imputation	$1,815	$2,600
No Control Totals	$1,735	$1,771
New Imputation	$1,398	$2,238

The use of the control totals significantly raises the average imputed MOOP from their respective averages in the original NMES file. While this difference could represent the difference between actual and reported MOOP, the differences are striking. But what is also shown is how the raking dramatically hides what a rather poor job the original regression model does in replicating the mean MOOP. Average non elderly spending is overstated while elderly spending is understated. While the previous discussion made us question the appropriateness of the model representing the upper tail of the MOOP distribution, these figures suggests it does a poor job replicating means. Given the similarity between the log-logistic and log normal models, moving toward a log normal model would not be a desirable path to follow.

The similarity of the average MOOP imputed with the New Implementation and the averages found in the NMES file are extremely comforting. They provide evidence of the gain in imputation accuracy provided by the new regression model and other recommendations.

I computed the average poverty rates for children, the elderly and for the total population for each of three implementation. For purposes of comparison, I have provided the official poverty rates for 1992. One might be concerned that the random noise in the imputation may lead to large variation in the poverty rates based upon the imputations. In the following table, I have also included the standard deviation of the estimated poverty rates.

	Poverty Rate of:		
	Children	Elderly	All Persons
Official	21.87	12.90	14.52
Old Imputation	24.90 (.11)	20.68 (.18)	17.72 (.06)
No Control Totals	24.76 (.11)	18.60 (.18)	17.32 (.06)
New Imputation	23.86 (.08)	19.87 (.20)	16.86 (.05)

The use of the control totals did lead to higher poverty rates. For children, the effect of not raking the data was minor compared to the elderly. However, the raking masked the rather poor imputation of the underlying model. When the improved model is employed, less MOOP is assigned to the non elderly and more is attributed to the elderly. This shift in the distribution between the two age groups has the expected impact on poverty rates. Children's rates fall and elderly rates rise when compared to the rates produced by the previous regression model without control totals.

The standard deviations (in parenthesis) of the poverty rates show how little possible variation in the rates can be caused by imputation procedure. In my opinion, they are quite small.

V. Imputation Model based upon CEX data

TO BE COMPLETED NEXT WEEK

V. Conclusions

References

Betson, David (1998) 'Imputation of Medical Out of Pocket (MOOP) Expenditures to CPS Analysis Files' memo, University of Notre Dame, Notre Dame, IN.

Betson, David (2000) 'Response to Bavier's Critique of the NRC Panel's Recommendations.' Appears on the Census Bureau's web site.

Betson, David, Constance Citro and Robert T. Michael (2000) 'Recent Developments for Poverty Measurement in U.S. Official Statistics' *Journal of Official Statistics*.

Citro, Constance and Robert T. Michael (1995) *Measuring Poverty : A New Approach*, National Academy Press, Washington, DC

Johnson, Norman and Samuel Kotz (1970) *Continuous Univariate Distributions – Volume 2*. John Wiley and Sons, New York, New York.

Moon, Marilyn (1996) *Medicare Now and in the Future, 2nd edition*. Urban Institute Press, Washington, DC.

Phelps, Charles, (1997) *Health Economics 2nd edition*. Addison-Wesley, Reading, MA.

Shah, B.K. and P. H. Dave (1963) "A Note on the Log-Logistic Distribution" Journal of the M.S. University of Baroda (Science Number), 12, pp. 15-20.

Appendix A

3rd Order Log-Logistic Regression Results for the Non Elderly – 1992 NMES

```
-> ipl=        1
  Source |       SS       df       MS                  Number of obs =      95
---------+------------------------------               F(  3,    91) = 2405.38
   Model |  305.842493     3  101.947498               Prob > F      =  0.0000
Residual |   3.8568665    91  .042383148               R-squared     =  0.9875
---------+------------------------------               Adj R-squared =  0.9871
   Total |  309.699359    94  3.29467403               Root MSE      =  .20587

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.028337   .0302884     33.952   0.000       .9681726    1.088501
   lnmp2 |   .3279916   .0137416     23.868   0.000       .3006955    .3552877
   lnmp3 |   .1039791   .0059403     17.504   0.000       .0921795    .1157788
   _cons |   .2481396   .0278213      8.919   0.000        .192876    .3034032
------------------------------------------------------------------------------

-> ipl=        2
  Source |       SS       df       MS                  Number of obs =      18
---------+------------------------------               F(  3,    14) =  125.96
   Model |  23.5796402     3  7.85988005               Prob > F      =  0.0000
Residual |    .8735684    14  .062397743               R-squared     =  0.9643
---------+------------------------------               Adj R-squared =  0.9566
   Total |  24.4532086    17  1.43842403               Root MSE      =   .2498

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.275706   .1246871     10.231   0.000       1.008279    1.543133
   lnmp2 |   .1921735   .1078129      1.782   0.096      -.0390621    .4234092
   lnmp3 |   .0123426     .02171      0.569   0.579      -.0342207    .0589059
   _cons |   .8920519   .0885659     10.072   0.000       .7020969    1.082007
------------------------------------------------------------------------------

-> ipl=        3
  Source |       SS       df       MS                  Number of obs =     661
---------+------------------------------               F(  3,   657) =41417.33
   Model |  2115.81186     3  705.270619               Prob > F      =  0.0000
Residual |  11.1876545   657  .017028393               R-squared     =  0.9947
---------+------------------------------               Adj R-squared =  0.9947
   Total |  2126.99951   660  3.22272653               Root MSE      =  .13049

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.640223   .0067654    242.442   0.000       1.626939    1.653508
   lnmp2 |   .2358787   .0043474     54.257   0.000       .2273422    .2444152
   lnmp3 |   .0217826   .0006417     33.947   0.000       .0205227    .0230426
   _cons |   .8545995   .0066664    128.195   0.000       .8415096    .8676895
------------------------------------------------------------------------------
```

```
-> ipl=        4
  Source |       SS       df       MS              Number of obs =     135
---------+------------------------------           F(  3,   131) = 5832.04
   Model |  421.075911     3  140.358637           Prob > F      =  0.0000
Residual |  3.15275169   131  .024066807           R-squared     =  0.9926
---------+------------------------------           Adj R-squared =  0.9924
   Total |  424.228663   134  3.16588554           Root MSE      =  .15513

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   2.067369   .0237845     86.921   0.000       2.020317     2.11442
   lnmp2 |   .5307119   .0178904     29.665   0.000       .4953205    .5661032
   lnmp3 |   .0668307   .0031987     20.893   0.000       .0605029    .0731584
   _cons |   .9250693   .0178594     51.797   0.000       .8897391    .9603995
------------------------------------------------------------------------------

-> ipl=        5
  Source |       SS       df       MS              Number of obs =      97
---------+------------------------------           F(  3,    93) = 1856.86
   Model |  225.398596     3  75.1328652           Prob > F      =  0.0000
Residual |  3.76299442    93  .040462306           R-squared     =  0.9836
---------+------------------------------           Adj R-squared =  0.9830
   Total |   229.16159    96   2.3870999           Root MSE      =  .20115

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |    1.20965    .021538     56.164   0.000        1.16688     1.25242
   lnmp2 |   .0904161   .0133463      6.775   0.000       .0639129    .1169192
   lnmp3 |   .0052674   .0031591      1.667   0.099       -.001006    .0115407
   _cons |  -.2725821   .0274003     -9.948   0.000      -.3269936   -.2181706
------------------------------------------------------------------------------

-> ipl=        6
  Source |       SS       df       MS              Number of obs =      36
---------+------------------------------           F(  3,    32) =  637.87
   Model |  120.944988     3  40.3149961           Prob > F      =  0.0000
Residual |  2.02247349    32  .063202297           R-squared     =  0.9836
---------+------------------------------           Adj R-squared =  0.9820
   Total |  122.967462    35  3.51335605           Root MSE      =   .2514

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.216438   .0539722     22.538   0.000       1.106501    1.326376
   lnmp2 |   -.077237   .0439161     -1.759   0.088      -.1666911    .0122171
   lnmp3 |    .005772   .0140506      0.411   0.684      -.0228481    .0343921
   _cons |   .5502528    .063478      8.668   0.000       .4209524    .6795532
------------------------------------------------------------------------------
```

```
-> ipl=        7
  Source |       SS       df       MS                  Number of obs =    1288
---------+------------------------------               F(  3,  1284) =94094.71
   Model |  4173.02539     3  1391.00846               Prob > F      =  0.0000
Residual |  18.9814588  1284  .014783068               R-squared     =  0.9955
---------+------------------------------               Adj R-squared =  0.9955
   Total |  4192.00685  1287  3.25719258               Root MSE      =  .12159

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.530963    .003594    425.980   0.000       1.523912    1.538014
   lnmp2 |   .2010627   .0020244     99.317   0.000       .1970911    .2050343
   lnmp3 |     .01849   .0004554     40.599   0.000       .0175965    .0193835
   _cons |  -.5283127   .0042469   -124.400   0.000      -.5366443   -.5199811
------------------------------------------------------------------------------

-> ipl=        8
  Source |       SS       df       MS                  Number of obs =     246
---------+------------------------------               F(  3,   242) =12675.98
   Model |  725.871756     3  241.957252               Prob > F      =  0.0000
Residual |  4.61926039   242  .019087853               R-squared     =  0.9937
---------+------------------------------               Adj R-squared =  0.9936
   Total |  730.491017   245  2.98159599               Root MSE      =  .13816

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.710414   .0107335    159.353   0.000       1.689271    1.731557
   lnmp2 |   .2842935   .0080191     35.452   0.000       .2684974    .3000896
   lnmp3 |   .0336115   .0019594     17.154   0.000       .0297518    .0374713
   _cons |   .0543327   .0110507      4.917   0.000       .0325649    .0761005
------------------------------------------------------------------------------

-> ipl=        9
  Source |       SS       df       MS                  Number of obs =     103
---------+------------------------------               F(  3,    99) = 2035.24
   Model |  286.768458     3  95.5894861               Prob > F      =  0.0000
Residual |  4.64975632    99  .046967236               R-squared     =  0.9840
---------+------------------------------               Adj R-squared =  0.9836
   Total |  291.418215   102  2.85704132               Root MSE      =  .21672

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.521131   .0363363     41.863   0.000       1.449032     1.59323
   lnmp2 |   .1103374   .0143106      7.710   0.000       .0819421    .1387327
   lnmp3 |    .019345   .0084659      2.285   0.024       .0025468    .0361432
   _cons |  -.4183104   .0261139    -16.019   0.000       -.470126   -.3664948
------------------------------------------------------------------------------
```

```
-> ipl=       10
  Source |       SS       df       MS                  Number of obs =      36
---------+------------------------------               F(  3,    32) =  568.31
   Model |  116.655431     3  38.8851436               Prob > F      =  0.0000
Residual |   2.1895111    32  .068422222               R-squared     =  0.9816
---------+------------------------------               Adj R-squared =  0.9798
   Total |  118.844942    35  3.39556977               Root MSE      =  .26158

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.856376   .0688594     26.959   0.000       1.716114    1.996639
   lnmp2 |   .2636983   .0546097      4.829   0.000       .1524621    .3749346
   lnmp3 |   .0151784    .025188      0.603   0.551      -.0361279    .0664848
   _cons |   .4011379   .0612369      6.551   0.000       .2764025    .5258734
------------------------------------------------------------------------------

-> ipl=       11
  Source |       SS       df       MS                  Number of obs =     785
---------+------------------------------               F(  3,   781) =83980.30
   Model |   2516.2491     3  838.749702               Prob > F      =  0.0000
Residual |  7.80020426   781  .009987457               R-squared     =  0.9969
---------+------------------------------               Adj R-squared =  0.9969
   Total |  2524.04931   784  3.21945065               Root MSE      =  .09994

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.592924    .004907    324.623   0.000       1.583292    1.602557
   lnmp2 |    .233218   .0024115     96.711   0.000       .2284843    .2379518
   lnmp3 |   .0303497   .0008133     37.316   0.000       .0287532    .0319463
   _cons |    -.87723    .004545   -193.008   0.000       -.886152   -.8683081
------------------------------------------------------------------------------

-> ipl=       12
  Source |       SS       df       MS                  Number of obs =     139
---------+------------------------------               F(  3,   135) = 5072.28
   Model |  412.725524     3  137.575175               Prob > F      =  0.0000
Residual |  3.66159985   135  .027122962               R-squared     =  0.9912
---------+------------------------------               Adj R-squared =  0.9910
   Total |  416.387124   138    3.017298               Root MSE      =  .16469

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.635539   .0223557     73.160   0.000       1.591326    1.679752
   lnmp2 |   .3662254   .0134862     27.156   0.000       .3395539     .392897
   lnmp3 |   .0795462   .0058544     13.587   0.000        .067968    .0911244
   _cons |  -.6992638   .0183075    -38.196   0.000      -.7354703   -.6630572
------------------------------------------------------------------------------
```

```
-> ipl=       13
  Source |       SS       df       MS                  Number of obs =      93
---------+------------------------------               F(  3,    89) = 4039.88
   Model |  244.456556     3  81.4855186               Prob > F      =  0.0000
Residual |   1.7951569    89  .020170302               R-squared     =  0.9927
---------+------------------------------               Adj R-squared =  0.9925
   Total |  246.251713    92  2.67664905               Root MSE      =  .14202

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.192253   .0225827     52.795   0.000       1.147381    1.237124
   lnmp2 |   .1081199   .0092478     11.691   0.000       .0897447    .1264951
   lnmp3 |    .006368    .000939      6.782   0.000       .0045022    .0082337
   _cons |    1.68074   .0222479     75.546   0.000       1.636533    1.724946
------------------------------------------------------------------------------

-> ipl=       14
  Source |       SS       df       MS                  Number of obs =      41
---------+------------------------------               F(  3,    37) = 1246.93
   Model |  116.471256     3  38.8237522               Prob > F      =  0.0000
Residual |  1.15201124    37  .031135439               R-squared     =  0.9902
---------+------------------------------               Adj R-squared =  0.9894
   Total |  117.623268    40  2.94058169               Root MSE      =  .17645

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   .9352596   .0375598     24.901   0.000       .8591562    1.011363
   lnmp2 |   .0268821   .0228014      1.179   0.246       -.019318    .0730822
   lnmp3 |   .0052166   .0033453      1.559   0.127      -.0015617    .0119949
   _cons |   1.591682   .0468847     33.949   0.000       1.496684    1.686679
------------------------------------------------------------------------------

-> ipl=       15
  Source |       SS       df       MS                  Number of obs =      25
---------+------------------------------               F(  3,    21) =  189.30
   Model |  56.9611825     3  18.9870608               Prob > F      =  0.0000
Residual |  2.10632801    21  .100301334               R-squared     =  0.9643
---------+------------------------------               Adj R-squared =  0.9592
   Total |  59.0675106    24  2.46114627               Root MSE      =   .3167

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   .8886003    .079478     11.180   0.000       .7233167    1.053884
   lnmp2 |   .2116832   .0432449      4.895   0.000       .1217506    .3016158
   lnmp3 |   .0591864   .0153827      3.848   0.001       .0271963    .0911766
   _cons |   .6600614   .0977487      6.753   0.000       .4567818     .863341
------------------------------------------------------------------------------
```

```
-> ipl=       16
  Source |       SS       df       MS              Number of obs =       6
---------+------------------------------           F(  3,     2) =  230.00
   Model |  9.36582546     3  3.12194182           Prob > F      =  0.0043
Residual |  .027147655     2  .013573827           R-squared     =  0.9971
---------+------------------------------           Adj R-squared =  0.9928
   Total |  9.39297312     5  1.87859462           Root MSE      =  .11651

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   .7575798   .1381421      5.484   0.032        .1632025    1.351957
   lnmp2 |  -.1534394    .113387     -1.353   0.309       -.6413041    .3344252
   lnmp3 |  -.0359639   .0199313     -1.804   0.213       -.1217215    .0497938
   _cons |   .7252665   .0663259     10.935   0.008         .439889    1.010644
------------------------------------------------------------------------------

-> ipl=       17
  Source |       SS       df       MS              Number of obs =     123
---------+------------------------------           F(  3,   119) = 3053.57
   Model |  373.368485     3  124.456162           Prob > F      =  0.0000
Residual |  4.85014579   119  .040757528           R-squared     =  0.9872
---------+------------------------------           Adj R-squared =  0.9869
   Total |   378.21863   122  3.10015271           Root MSE      =  .20188

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.264254   .0232095     54.471   0.000        1.218297    1.310212
   lnmp2 |   .2554846    .015543     16.437   0.000         .224708    .2862612
   lnmp3 |   .0370303   .0025662     14.430   0.000        .0319491    .0421116
   _cons |   1.474751   .0299808     49.190   0.000        1.415386    1.534116
------------------------------------------------------------------------------

-> ipl=       18
  Source |       SS       df       MS              Number of obs =      95
---------+------------------------------           F(  3,    91) = 2324.83
   Model |  341.404109     3   113.80137           Prob > F      =  0.0000
Residual |  4.45448083    91  .048950339           R-squared     =  0.9871
---------+------------------------------           Adj R-squared =  0.9867
   Total |   345.85859    94   3.6793467           Root MSE      =  .22125

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.313108    .030892     42.506   0.000        1.251745    1.374471
   lnmp2 |   .1801353   .0149059     12.085   0.000        .1505266     .209744
   lnmp3 |   .0179951   .0018205      9.885   0.000         .014379    .0216113
   _cons |   2.001622   .0354134     56.522   0.000        1.931277    2.071966
------------------------------------------------------------------------------
```

```
-> ipl=       19
  Source |       SS       df       MS              Number of obs =      80
---------+------------------------------           F(  3,    76) = 1706.45
   Model |  191.563878     3  63.8546261           Prob > F      =  0.0000
Residual |   2.8438812    76   .03741949           R-squared     =  0.9854
---------+------------------------------           Adj R-squared =  0.9848
   Total |  194.407759    79  2.46085771           Root MSE      =  .19344

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.342251   .0239307     56.089   0.000       1.294589    1.389913
   lnmp2 |   .2363857   .0169853     13.917   0.000       .2025567    .2702148
   lnmp3 |   .0221976   .0029257      7.587   0.000       .0163706    .0280246
   _cons |   .6526474   .0317504     20.556   0.000        .589411    .7158838
------------------------------------------------------------------------------

-> ipl=       20
  Source |       SS       df       MS              Number of obs =      25
---------+------------------------------           F(  3,    21) =  396.62
   Model |  67.7309759     3   22.576992           Prob > F      =  0.0000
Residual |  1.19538687    21  .056923184           R-squared     =  0.9827
---------+------------------------------           Adj R-squared =  0.9802
   Total |  68.9263628    24  2.87193178           Root MSE      =  .23859

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   .9711176   .0566334     17.147   0.000        .853342    1.088893
   lnmp2 |   .0686191   .0358318      1.915   0.069      -.0058971    .1431354
   lnmp3 |   .0117069   .0059579      1.965   0.063      -.0006831     .024097
   _cons |   1.274808   .0717247     17.774   0.000       1.125648    1.423968
------------------------------------------------------------------------------

-> ipl=       21
  Source |       SS       df       MS              Number of obs =      83
---------+------------------------------           F(  3,    79) = 1622.14
   Model |  224.942774     3  74.9809247           Prob > F      =  0.0000
Residual |  3.65165743    79  .046223512           R-squared     =  0.9840
---------+------------------------------           Adj R-squared =  0.9834
   Total |  228.594432    82  2.78773697           Root MSE      =    .215

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.158905   .0283697     40.850   0.000       1.102436    1.215373
   lnmp2 |   .1862828   .0217528      8.564   0.000       .1429851    .2295806
   lnmp3 |    .030229   .0043844      6.895   0.000       .0215021     .038956
   _cons |    .866542   .0354968     24.412   0.000       .7958874    .9371966
------------------------------------------------------------------------------
```

```
-> ipl=       22
  Source |       SS       df       MS                  Number of obs =      58
---------+------------------------------               F(  3,    54) = 1114.55
   Model |  173.539553     3  57.8465176               Prob > F      =  0.0000
Residual |  2.80267227    54  .051901338               R-squared     =  0.9841
---------+------------------------------               Adj R-squared =  0.9832
   Total |  176.342225    57  3.09372325               Root MSE      =  .22782

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   .8611484   .0454843     18.933   0.000       .7699578     .952339
   lnmp2 |  -.0174433   .0279733     -0.624   0.536      -.0735264    .0386398
   lnmp3 |    .000104   .0039589      0.026   0.979      -.0078331    .0080411
   _cons |   1.913165   .0522935     36.585   0.000       1.808323    2.018008
------------------------------------------------------------------------------

-> ipl=       23
  Source |       SS       df       MS                  Number of obs =      61
---------+------------------------------               F(  3,    57) = 1584.69
   Model |  184.506921     3   61.502307               Prob > F      =  0.0000
Residual |  2.21218187    57  .038810208               R-squared     =  0.9882
---------+------------------------------               Adj R-squared =  0.9875
   Total |  186.719103    60  3.11198505               Root MSE      =    .197

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.267037   .0278727     45.458   0.000       1.211223    1.322851
   lnmp2 |   .1302579   .0179956      7.238   0.000       .0942223    .1662935
   lnmp3 |   .0134795   .0028976      4.652   0.000       .0076772    .0192817
   _cons |   .8463711    .035423     23.893   0.000       .7754377    .9173045
------------------------------------------------------------------------------

-> ipl=       24
  Source |       SS       df       MS                  Number of obs =      33
---------+------------------------------               F(  3,    29) =  231.95
   Model |  78.0530688     3  26.0176896               Prob > F      =  0.0000
Residual |  3.25291544    29  .112169498               R-squared     =  0.9600
---------+------------------------------               Adj R-squared =  0.9559
   Total |  81.3059842    32  2.54081201               Root MSE      =  .33492

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   .7206272   .1347306      5.349   0.000       .4450722    .9961822
   lnmp2 |  -.1105214   .0740161     -1.493   0.146      -.2619013    .0408585
   lnmp3 |  -.0121664   .0096213     -1.265   0.216      -.0318442    .0075113
   _cons |   2.101374   .1117858     18.798   0.000       1.872747    2.330002
------------------------------------------------------------------------------
```

```
-> ipl=       25
  Source |       SS       df       MS                  Number of obs =     159
---------+------------------------------               F(  3,   155) = 5001.63
   Model |  492.381375     3  164.127125               Prob > F      =  0.0000
Residual |  5.08628555   155  .032814745               R-squared     =  0.9898
---------+------------------------------               Adj R-squared =  0.9896
   Total |   497.46766   158   3.1485295               Root MSE      =  .18115

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.410947   .0205055     68.808   0.000        1.37044    1.451453
   lnmp2 |   .0891125   .0113234      7.870   0.000       .0667443    .1114807
   lnmp3 |   .0030982   .0016678      1.858   0.065      -.0001963    .0063927
   _cons |   1.995875   .0224328     88.971   0.000       1.951562    2.040189
------------------------------------------------------------------------------

-> ipl=       26
  Source |       SS       df       MS                  Number of obs =      48
---------+------------------------------               F(  3,    44) = 1057.84
   Model |  112.080781     3  37.3602602               Prob > F      =  0.0000
Residual |  1.55397074    44  .035317517               R-squared     =  0.9863
---------+------------------------------               Adj R-squared =  0.9854
   Total |  113.634751    47  2.41776067               Root MSE      =  .18793

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   .8881042   .0274997     32.295   0.000       .8326822    .9435262
   lnmp2 |  -.1773623    .014845    -11.948   0.000      -.2072804   -.1474441
   lnmp3 |  -.0239216   .0015952    -14.996   0.000      -.0271364   -.0207067
   _cons |   1.679487   .0439712     38.195   0.000       1.590869    1.768105
------------------------------------------------------------------------------

-> ipl=       27
  Source |       SS       df       MS                  Number of obs =     165
---------+------------------------------               F(  3,   161) = 2794.02
   Model |  435.680349     3  145.226783               Prob > F      =  0.0000
Residual |  8.36841432   161  .051977729               R-squared     =  0.9812
---------+------------------------------               Adj R-squared =  0.9808
   Total |  444.048763   164  2.70761441               Root MSE      =  .22799

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.476446   .0294434     50.145   0.000       1.418301    1.534591
   lnmp2 |   .1750389   .0227349      7.699   0.000       .1301418    .2199361
   lnmp3 |   .0244074   .0043731      5.581   0.000       .0157713    .0330434
   _cons |   1.649103   .0301115     54.767   0.000       1.589639    1.708568
------------------------------------------------------------------------------
```

```
-> ipl=       28
  Source |       SS       df       MS                        Number of obs =      31
---------+------------------------------                     F(  3,    27) =  607.16
   Model |  80.0117646     3  26.6705882                     Prob > F      =  0.0000
Residual |  1.18603266    27  .043927136                     R-squared     =  0.9854
---------+------------------------------                     Adj R-squared =  0.9838
   Total |  81.1977972    30  2.70659324                     Root MSE      =  .20959

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.048392   .1043977     10.042   0.000        .8341852    1.262598
   lnmp2 |  -.1262209      .0835     -1.512   0.142       -.2975487    .0451068
   lnmp3 |  -.0106782   .0156954     -0.680   0.502       -.0428824     .021526
   _cons |   2.294564   .0723208     31.728   0.000        2.146174    2.442954
------------------------------------------------------------------------------

-> ipl=       29
  Source |       SS       df       MS                        Number of obs =     117
---------+------------------------------                     F(  3,   113) = 5568.06
   Model |  394.496958     3  131.498986                     Prob > F      =  0.0000
Residual |   2.6686842   113  .023616674                     R-squared     =  0.9933
---------+------------------------------                     Adj R-squared =  0.9931
   Total |  397.165643   116  3.42384175                     Root MSE      =  .15368

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.383701   .0180565     76.632   0.000        1.347928    1.419474
   lnmp2 |   .1510551    .014352     10.525   0.000        .1226212     .179489
   lnmp3 |   .0329309   .0034669      9.499   0.000        .0260624    .0397995
   _cons |   1.289012   .0221936     58.080   0.000        1.245043    1.332982
------------------------------------------------------------------------------

-> ipl=       30
  Source |       SS       df       MS                        Number of obs =      39
---------+------------------------------                     F(  3,    35) =  613.78
   Model |  124.329781     3  41.4432603                     Prob > F      =  0.0000
Residual |     2.36323    35  .067520857                     R-squared     =  0.9813
---------+------------------------------                     Adj R-squared =  0.9797
   Total |  126.693011    38   3.3340266                     Root MSE      =  .25985

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.763089   .0954174     18.478   0.000        1.569382    1.956797
   lnmp2 |   .2158413   .0449643      4.800   0.000        .1245589    .3071237
   lnmp3 |    .015256   .0053539      2.850   0.007         .004387     .026125
   _cons |   2.111643   .0713644     29.590   0.000        1.966765     2.25652
------------------------------------------------------------------------------
```

```
-> ipl=       31
  Source |       SS       df       MS                  Number of obs =     147
---------+------------------------------               F(  3,   143) = 2957.62
   Model |  432.398629     3  144.132876               Prob > F      =  0.0000
Residual |  6.9687771    143  .048732707               R-squared     =  0.9841
---------+------------------------------               Adj R-squared =  0.9838
   Total |  439.367406   146  3.00936579               Root MSE      =  .22075

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.664321   .0250284     66.497   0.000       1.614848    1.713795
   lnmp2 |   .1682311   .0202813      8.295   0.000       .1281412     .208321
   lnmp3 |   .0173382   .0039553      4.384   0.000       .0095199    .0251566
   _cons |   .9179109    .024995     36.724   0.000       .8685034    .9673184
------------------------------------------------------------------------------

-> ipl=       32
  Source |       SS       df       MS                  Number of obs =      29
---------+------------------------------               F(  3,    25) =  702.67
   Model |  45.0813839     3  15.027128                Prob > F      =  0.0000
Residual |  .534642548    25  .021385702               R-squared     =  0.9883
---------+------------------------------               Adj R-squared =  0.9869
   Total |  45.6160265    28   1.6291438               Root MSE      =  .14624

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.085683    .037348     29.069   0.000       1.008763    1.162602
   lnmp2 |   .4554284   .0377607     12.061   0.000       .3776588     .533198
   lnmp3 |   .0980967   .0100309      9.779   0.000       .0774378    .1187557
   _cons |   .6488838   .0481696     13.471   0.000       .5496767    .7480909
------------------------------------------------------------------------------

-> ipl=       33
  Source |       SS       df       MS                  Number of obs =      94
---------+------------------------------               F(  3,    90) = 2764.60
   Model |  264.342029     3  88.1140095               Prob > F      =  0.0000
Residual |  2.86850687    90  .031872299               R-squared     =  0.9893
---------+------------------------------               Adj R-squared =  0.9889
   Total |  267.210536    93  2.87323156               Root MSE      =  .17853

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.377571   .0188401     73.119   0.000       1.340142    1.415001
   lnmp2 |   .1627608   .0157923     10.306   0.000       .1313867     .194135
   lnmp3 |   .0156125   .0034004      4.591   0.000        .008857     .022368
   _cons |   .4258514   .0268525     15.859   0.000       .3725042    .4791985
------------------------------------------------------------------------------
```

```
-> ipl=       34
  Source |       SS       df       MS                  Number of obs =      25
---------+------------------------------               F(  3,    21) =  125.37
   Model |  61.8161143     3  20.6053714               Prob > F      =  0.0000
Residual |  3.45156253    21   .16436012               R-squared     =  0.9471
---------+------------------------------               Adj R-squared =  0.9396
   Total |  65.2676768    24  2.71948654               Root MSE      =  .40541

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.303318   .1274283     10.228   0.000       1.038316     1.56832
   lnmp2 |   .2476674   .1658784      1.493   0.150      -.0972957    .5926304
   lnmp3 |   .0569785   .0448092      1.272   0.217      -.0362074    .1501645
   _cons |     1.1964   .1490861      8.025   0.000       .8863583    1.506441
------------------------------------------------------------------------------

-> ipl=       35
  Source |       SS       df       MS                  Number of obs =      95
---------+------------------------------               F(  3,    91) = 3712.32
   Model |  305.735567     3  101.911856               Prob > F      =  0.0000
Residual |  2.49816301    91  .027452341               R-squared     =  0.9919
---------+------------------------------               Adj R-squared =  0.9916
   Total |   308.23373    94  3.27908224               Root MSE      =  .16569

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.455078    .019807     73.463   0.000       1.415733    1.494422
   lnmp2 |    .400624   .0143618     27.895   0.000       .3720961    .4291519
   lnmp3 |   .0720243   .0037024     19.453   0.000       .0646699    .0793787
   _cons |   .0460845   .0224547      2.052   0.043       .0014809    .0906881
------------------------------------------------------------------------------

-> ipl=       36
  Source |       SS       df       MS                  Number of obs =      17
---------+------------------------------               F(  3,    13) =  126.96
   Model |  30.1905849     3  10.0635283               Prob > F      =  0.0000
Residual |  1.03043208    13  .079264006               R-squared     =  0.9670
---------+------------------------------               Adj R-squared =  0.9594
   Total |   31.221017    16  1.95131356               Root MSE      =  .28154

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |    1.02722   .1537007      6.683   0.000       .6951702    1.359271
   lnmp2 |  -.1764032   .0573916     -3.074   0.009      -.3003901   -.0524162
   lnmp3 |   .0762584   .0532075      1.433   0.175      -.0386894    .1912061
   _cons |   .5618133   .1106814      5.076   0.000       .3227008    .8009259
------------------------------------------------------------------------------
```

Appendix B

3rd Order Log-Logistic Regression Results for the Elderly – 1992 NMES

```
-> ipl=        1
  Source |       SS       df       MS                  Number of obs =     271
---------+------------------------------               F(  3,   267) = 6312.62
   Model |  969.687286     3  323.229095               Prob > F      =  0.0000
Residual |  13.6713752   267  .051203652               R-squared     =  0.9861
---------+------------------------------               Adj R-squared =  0.9859
   Total |  983.358661   270  3.64206911               Root MSE      =  .22628

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.492615   .0138132    108.057   0.000       1.465418    1.519811
   lnmp2 |   .2132053   .0081531     26.150   0.000       .1971528    .2292577
   lnmp3 |   .0216962   .0014721     14.739   0.000       .0187979    .0245945
   _cons |   .0970413   .0172941      5.611   0.000       .0629911    .1310916
------------------------------------------------------------------------------

-> ipl=        2
  Source |       SS       df       MS                  Number of obs =     286
---------+------------------------------               F(  3,   282) = 6476.79
   Model |  896.414127     3  298.804709               Prob > F      =  0.0000
Residual |  13.0099841   282  .046134695               R-squared     =  0.9857
---------+------------------------------               Adj R-squared =  0.9855
   Total |  909.424111   285  3.19096179               Root MSE      =  .21479

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.870395    .014838    126.054   0.000       1.841187    1.899602
   lnmp2 |   .2991561   .0099501     30.066   0.000       .2795703    .3187419
   lnmp3 |   .0202544   .0018981     10.671   0.000       .0165181    .0239906
   _cons |  -.2493622   .0153731    -16.221   0.000      -.2796228   -.2191017
------------------------------------------------------------------------------

-> ipl=        3
  Source |       SS       df       MS                  Number of obs =     129
---------+------------------------------               F(  3,   125) = 8213.81
   Model |  383.006522     3  127.668841               Prob > F      =  0.0000
Residual |   1.9428987   125   .01554319               R-squared     =  0.9950
---------+------------------------------               Adj R-squared =  0.9948
   Total |  384.949421   128  3.00741735               Root MSE      =  .12467

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.052572   .0108431     97.073   0.000       1.031112    1.074032
   lnmp2 |   .2052475   .0067467     30.422   0.000        .191895    .2185999
   lnmp3 |   .0425626   .0017337     24.550   0.000       .0391314    .0459939
   _cons |   .0244914   .0158575      1.544   0.125      -.0068926    .0558753
------------------------------------------------------------------------------
```

```
-> ipl=        4
  Source |       SS       df       MS                  Number of obs =     540
---------+------------------------------               F(  3,   536) =77806.15
   Model |  1776.08577     3  592.028591               Prob > F      =  0.0000
Residual |  4.0784351    536  .007609021               R-squared     =  0.9977
---------+------------------------------               Adj R-squared =  0.9977
   Total |  1780.16421   539  3.30271653               Root MSE      =  .08723

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.456565   .0047581    306.125   0.000       1.447218    1.465912
   lnmp2 |   .0989779   .0017526     56.473   0.000        .095535    .1024208
   lnmp3 |   .0180522   .0006072     29.730   0.000       .0168593     .019245
   _cons |  -.8261255   .0046334   -178.298   0.000      -.8352274   -.8170236
------------------------------------------------------------------------------

-> ipl=        5
  Source |       SS       df       MS                  Number of obs =     329
---------+------------------------------               F(  3,   325) = 3772.72
   Model |  1164.88921     3  388.296405               Prob > F      =  0.0000
Residual |  33.4497098   325  .102922184               R-squared     =  0.9721
---------+------------------------------               Adj R-squared =  0.9718
   Total |  1198.33892   328  3.65347233               Root MSE      =  .32081

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.353372   .0170589     79.335   0.000       1.319812    1.386932
   lnmp2 |   .1354451   .0069065     19.611   0.000        .121858    .1490322
   lnmp3 |   .0077417   .0015246      5.078   0.000       .0047423     .010741
   _cons |  -.0288342   .0211457     -1.364   0.174      -.0704339    .0127655
------------------------------------------------------------------------------

-> ipl=        6
  Source |       SS       df       MS                  Number of obs =     272
---------+------------------------------               F(  3,   268) = 4425.14
   Model |  983.959069     3  327.986356               Prob > F      =  0.0000
Residual |  19.8638414   268  .074118811               R-squared     =  0.9802
---------+------------------------------               Adj R-squared =  0.9800
   Total |  1003.82291   271  3.70414358               Root MSE      =  .27225

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.716371   .0276049     62.176   0.000       1.662021    1.770722
   lnmp2 |   .1079747   .0080613     13.394   0.000       .0921032    .1238463
   lnmp3 |  -.0074709   .0050011     -1.494   0.136      -.0173173    .0023754
   _cons |   -.563099   .0209306    -26.903   0.000      -.6043084   -.5218897
------------------------------------------------------------------------------
```

```
-> ipl=        7
  Source |       SS       df       MS                  Number of obs =     107
---------+------------------------------               F(  3,   103) = 2555.51
   Model |  303.895845     3  101.298615               Prob > F      =  0.0000
Residual |  4.08284255   103  .039639248               R-squared     =  0.9867
---------+------------------------------               Adj R-squared =  0.9864
   Total |  307.978688   106  2.90545932               Root MSE      =   .1991

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   .9622642   .0214949     44.767   0.000       .9196342    1.004894
   lnmp2 |   .1523191   .0084182     18.094   0.000       .1356236    .1690145
   lnmp3 |   .0343741   .0028561     12.035   0.000       .0287097    .0400385
   _cons |  -.3060569   .0246775    -12.402   0.000      -.3549988    -.257115
------------------------------------------------------------------------------

-> ipl=        8
  Source |       SS       df       MS                  Number of obs =     239
---------+------------------------------               F(  3,   235) = 9930.15
   Model |  660.545375     3  220.181792               Prob > F      =  0.0000
Residual |  5.21066722   235  .022173052               R-squared     =  0.9922
---------+------------------------------               Adj R-squared =  0.9921
   Total |  665.756042   238  2.79729429               Root MSE      =  .14891

------------------------------------------------------------------------------
  lnodds |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  lnmoop |   1.343431   .0131444    102.206   0.000       1.317536    1.369327
   lnmp2 |   .0209974    .004437      4.732   0.000       .0122561    .0297387
   lnmp3 |   .0058502   .0019436      3.010   0.003       .0020211    .0096792
   _cons |  -.9076475   .0131758    -68.888   0.000      -.9336052   -.8816898
------------------------------------------------------------------------------
```

Appendix C

Fortran Source Code for Imputation of MOOP

```
The following variables are needed prior to calling either yngoop3 or
oldoop3:

For units headed by a person less than 65 years old:

   Let:

               1 if has private insurance
      icov =   2 if Medicaid or Medicare only
               3 if uninsured

               1 if single individual
      isize =  2 if family size 2 or 3
               3 if family size is 4 or more

      npoor =  1 if census money income is less than 150% of poverty
               2 otherwise

      irace =  2 if Black
               1 otherwise

      v        a random draw from a uniform distribution
      u        a random draw from a uniform distribution

   ipl = (icov-1)*12 + (isize-1)*4 + (npoor-1)*2 + irace

   then

            oop = yngoop3(ipl,v,u) ! returns with value of MOOP in $1

For Units headed by a person 65 years old or older:

   iage=       1 if head is less than 75 years old
               2 if head is 75 years old or older

   isize =     1 if single individual
               2 if family size is 2 or more

      npoor =  1 if census money income is less than 150% of poverty
               2 otherwise

   ipl = (iage-1)*4 + (isize-1)*2 + npoor

   then

             oop = oldoop3(ipl,v,u) ! returns with value of MOOP in $1

Then add Medicare Part B premiums:

            partB =  0              if income less than 120% of poverty
                     PREM*NOLD      otherwise

Where PREM is the yearly premium and NOLD is the number of Elderly.
Source Code for yngoop3, oldoop3 and other need functions:
```

```
      function yngoop3(ipl,v,u)
      real ypzero(36),cof(4,36)

      data ypzero/
     & .065,.041,.075,.143,.061,.083,.012,.012,.031,.024,.003,.006,
     & .397,.606,.219,.628,.371,.408,.212,.279,.237,.507,.256,.345,
     & .378,.482,.248,.420,.151,.194,.103,.128,.043,.126,.036,.213/

     data (cof(n,1),n=1,4)/1.028337,.32799,.10397,.24814/
      data (cof(n,2),n=1,4)/1.2757,.19217,.01234,.89205/
      data (cof(n,3),n=1,4)/1.64022,.23587,.021783,.85459/
      data (cof(n,4),n=1,4)/2.067369,.5307119,.06683,.92506/
      data (cof(n,5),n=1,4)/1.20965,.090416,.005267,-.27258/
      data (cof(n,6),n=1,4)/1.216438,-.077237,.00577,.550253/
      data (cof(n,7),n=1,4)/1.530963,.20106,.01849,-.528312/
      data (cof(n,8),n=1,4)/1.710414,.28429,.0336115,.054332/
     data (cof(n,9),n=1,4)/1.52113,.11033,.019345,-.41831/
      data (cof(n,10),n=1,4)/1.85638,.26369,.01517,.40114/
      data (cof(n,11),n=1,4)/1.59292,.2332,.03035,-.87723/
      data (cof(n,12),n=1,4)/1.635539,.36623,.079546,-.69926/
      data (cof(n,13),n=1,4)/1.19225,.108119,.006368,1.68074/
      data (cof(n,14),n=1,4)/.93525,.02688,.0052166,1.59168/
      data (cof(n,15),n=1,4)/.8886,.21168,.059186,.66006/
      data (cof(n,16),n=1,4)/.75758,-.15343,-.03596,.725266/
     data (cof(n,17),n=1,4)/1.26425,.25548,.03703,1.47475/
      data (cof(n,18),n=1,4)/1.3131,.180135,.017995,2.001622/
      data (cof(n,19),n=1,4)/1.34225,.2363857,.0221976,.652647/
      data (cof(n,20),n=1,4)/.971117,.0686191,.011707,1.2748/
      data (cof(n,21),n=1,4)/1.1589,.186283,.030229,.86654/
      data (cof(n,22),n=1,4)/.861148,-.017444,.000104,1.913165/
      data (cof(n,23),n=1,4)/1.26704,.13025,.01348,.84627/
      data (cof(n,24),n=1,4)/.720627,-.1105214,-.012166,2.101374/
     data (cof(n,25),n=1,4)/1.410947,.089113,.003098,1.995875/
      data (cof(n,26),n=1,4)/.8881,-.17736,-.02393,1.67948/
      data (cof(n,27),n=1,4)/1.476446,.17504,.024407,1.6491/
      data (cof(n,28),n=1,4)/1.048392,-.1262209,-.010678,2.294564/
      data (cof(n,29),n=1,4)/1.383701,.151055,.03293,1.28901/
      data (cof(n,30),n=1,4)/1.76309,.21584,.01526,2.11164/
      data (cof(n,31),n=1,4)/1.664321,.1682311,.017338,.917911/
      data (cof(n,32),n=1,4)/1.08568,.45543,.098097,.648884/
     data (cof(n,33),n=1,4)/1.377571,.1627608,.0156125,.4258514/
      data (cof(n,34),n=1,4)/1.303318,.24766,.05697,1.1964/
      data (cof(n,35),n=1,4)/1.455078,.400624,.0720243,.04608/
      data (cof(n,36),n=1,4)/1.02722,-.1764032,.076258,.561813/


      yngoop3=0.0
      havemp=1.-ypzero(ipl)
      if(v.gt.havemp) return


      d=cof(4,ipl)
      f1=cof(1,ipl)
      f2=cof(2,ipl)
      f3=cof(3,ipl)
      z=amin1(.99,u)
      odds=alog(z/(1.-z))

      yngoop3=root3(odds,d,f1,f2,f3)

      return
      end
      function oldoop3(ipl,v,u)

      real opzero(8),cof(4,8)

      data (cof(n,1),n=1,4)/1.4926,.2132,.02169,.09704/
      data (cof(n,2),n=1,4)/1.8704,.2992,.2025,-.2494/
```

```
 data (cof(n,3),n=1,4)/1.05257,.20525,.04256,.02449/
 data (cof(n,4),n=1,4)/1.45657,.098978,.01805,-.82613/
 data (cof(n,5),n=1,4)/1.3534,.13545,.00774,-.02883/
 data (cof(n,6),n=1,4)/1.71637,.10798,-.00747,-.563099/
 data (cof(n,7),n=1,4)/.9226,.15232,.034374,-.30606/
 data (cof(n,8),n=1,4)/1.34343,.020997,.00585,-.90765/

data opzero/.1666,.0233,.1010,.016,.0872,.0220,.0536,.0165/

 oldoop3=0.0
 havemp=1.-opzero(ipl)
 if(v.gt.havemp) return

 d=cof(4,ipl)
 f1=cof(1,ipl)
 f2=cof(2,ipl)
 f3=cof(3,ipl)
 z=amin1(.99,u)
 odds=alog(z/(1.-z))

 oldoop3=root3(odds,d,f1,f2,f3)

 return
 end
```

```
      function root3(odds,d,f1,f2,f3)

      data tol/.001/
      con=d-odds

      y0=-con/f1
      zero0=cube(y0,con,f1,f2,f3)

      do inter=1,20

       slope=dcube(y0,f1,f2,f3)
       step=zero0/slope
       istep=1

1      y1=y0-step/float(istep)
       zero1=cube(y1,con,f1,f2,f3)
       if(abs(zero1).lt.abs(zero0)) go to 5
       istep=istep+1
       if(istep.gt.3) go to 4
       go to 1

4      y1=y0-zero0/f1
       zero1=cube(y1,con,f1,f2,f3)

5      if(abs(zero1).lt.tol) go to 10

        y0=y1
        zero0=zero1

      repeat

10    root3=1000.*exp(y1)


      return
      end


      function cube(y,con,f1,f2,f3)

      cube=con+f1*y+f2*y*y+f3*y*y*y

      return
      end


      function dcube(y,f1,f2,f3)

      dcube=f1+2.0*y*f2+3.0*f3*y*y

      return
      end
```

Imputation of Medical Out-of-Pocket (MOOP) Expenditures to CPS Analysis Files

David M. Betson
University of Notre Dame
February 1998

The purpose of this memo is to describe the methods that were employed to impute medical out-of-pocket (MOOP) expenditures to the various years of CPS data which were utilized to analyze the NRC Panel's poverty measure recommendations. The imputation procedure consisted of two parts: estimating the total amount of MOOP which would be used for a control total for the imputation, and a procedure of allocating the totals to individual records.

Constructing Control Totals for MOOP

After much searching and questioning of researchers both inside and outside of government, I concluded there did not exist a consistent series for how much the non-institutionalized population spends directly out of their own pockets for medical services and supplies. In the absence of an official series, I decided to construct one for the research project of back casting the NRC Panel's recommendation to years prior to 1992 -- the year that was reported in the Panel's report.

I began with a simple accounting relationship that states that the aggregate amount of MOOP in current dollars is equal to the real per capita MOOP (RPCMOOP) at time t times the current price of health care (PHC) times the size of the population (POP), i.e.,

$$MOOP_t \equiv POP_t \text{ x } PHC_t \text{ x } RPCMOOP_t \,.$$

Using this identity, a rather simple estimate of MOOP at t could be based upon the assumption that RPCMOOP remains constant over time? Utilizing this assumption, we could estimate MOOPt using only a single years estimate of MOOP and a historical series of population estimates and the price of health care. If B is the base year in which we have an estimate of MOOP then the specific estimate in any year t would be

$$\begin{aligned} MOOP_t &= POP_t \text{ x } PHC_t \text{ x } RPCMOOP_B \\ &= MOOP_B \text{ x } \frac{POP_t}{POP_B} \text{ x } \frac{PHC_t}{PHC_B} \;. \end{aligned}$$

.

To evaluate how well this simple estimate performs, I used a historical series published by DHHS (Table 124 in *Health, United States 1992*) which reports on the aggregate amount of MOOP (direct out-of-pocket payments for services and supplies plus the total amount of health care insurance premiums paid by households) in the total population (including the institutionalized population). Given I would be back casting data, I chose the last year of the series, 1991, as my base year. Using a historical series for the total population and medical price index, I employed the above equation to predict MOOP in each of the previous ten years in the published DHHS series. The results of this evaluation is presented in the following table.

YEAR	MOOP (in billions) DHHS	Estimate	Percentage Difference
1965	23.6	21.5	-8.8%
1967	23.8	24.6	3.5%
1970	31.6	30.6	-3.1%
1975	48.4	45.1	-6.9%
1980	76.1	75.0	-1.5%
1985	124.4	118.9	-4.4%
1987	146.3	138.8	-5.1%
1988	156.2	149.2	-4.5%
1989	168.9	162.3	-3.9%
1990	183.1	178.8	-2.4%
1991	196.5	196.5	0.0%

I feel that this comparison suggests two conclusions. While the naive model consistently underestimates the published data, it does a fairly good job of predicting previous years MOOP especially during the period which we will be imputing, 1979 to the present. Second, the consistent underestimation of the model suggests that real per capita MOOP has over time been declining not constant as assumed by the model. While it is true that over this period, the percentage of all health care directly financed out households' pockets has been significantly declining, what has not been documented has been real per capita spending.

While this exercise built some confidence in what I was going to do, I felt that some other information could also be used. The Annual Statistical Supplement to the Social Security Bulletin provides annual data on the amount of premiums paid by households to Medicare Part B. My strategy was to begin where there was some consensus, MOOP in 1992. As part of the NRC Panel's work, we received from AHCPR their estimate of MOOP for the non-institutionalized population in 1992. Their estimate was $219.4 billion dollars which included premium payments to Medicare Part B. In 1992, there was $11.0 billion of Medicare Part B payments. What I decided to do was to forecast and backcast the difference between these two numbers in 1992 ($208.4) by changes in population (using the CPS counts of total population) and prices (Medical Care Component of the CPI) employing the above naive accounting model. To arrive at the aggregate MOOP figure, I would add the published Medicare Part B payments to this estimate. The following table presents the results of the calculations for the years of the CPS to which I will be imputing MOOP values. The numbers in **bold** type face represent figures from published sources or figures which I believe there is some consensus.

Aggregate Control Totals for MOOP

YEAR	Medicare Part B	OTHER MOOP	TOTAL MOOP
1979	**2.3**	65.2	67.5
1983	**3.5**	101.2	104.7
1989	**10.5**	158.6	169.1
1992	**11.0**	**208.4**	**219.4**
1994	**16.2**	236.1	252.3

Allocating Aggregate MOOP to Elderly and Non Elderly Households

The next step is to allocate these aggregates to individual households. I first disaggregated the aggregate totals into what households (families) headed by non-elderly and headed by elderly adults would spend on MOOP. In conversations with Urban Institute and researchers in ASPE (Health), there seemed to be consensus that roughly 27% of all MOOP was made by elderly units. This percentage was based upon examination of NMES data from 1987 which was aged to 1992. However, in article by Acs and Sablehouse (Monthly Labor Review, 1995) the authors report that in 1992, 34% of MOOP expenditures were made by elderly families reported in the CEX survey (my own calculations on the CEX suggest that 33% of MOOP expenditures were made by the elderly). To be honest, I am not sure which estimate is to believe so I decided to average the two estimates and use the figure of 30.3% for split between the elderly and non-elderly populations in 1992.

The Acs and Sablehouse article show that over the period of 1980 to 1992, the share of MOOP paid by the elderly has grown. The share of MOOP of the elderly in 1992 was 9.2% higher than in 1980. To replicate the general pattern of changes in the elderly share, I assumed the following splits of the aggregate amount of MOOP :

Year :	1979	1983	1989	1992	1994
Elderly Share	27.4%	28.9%	30.3%	30.3%	30.0%

These shares were based upon the change in the relative number of families units headed by an elderly individual. In particular, I used the following adjustment process :

$$\text{Elderly Share of MOOP in Year t} = \frac{\text{\% of Families Headed by Elderly in Year t}}{\text{\% of Families Headed by Elderly in 1992}} \times 30.3$$

The following controls were derived for the Elderly and NonElderly subpopulations. Again the numbers in **bold** type face, represent figures from published sources or figures which I believe there is some consensus.

Subaggregate Control Totals for MOOP (in Billions)

YEAR	Aggregate	Medicare Part B	Other Elderly MOOP	NonElderly
1979	67.5	**2.3**	16.1	49.1
1983	104.7	**3.5**	26.7	74.5
1989	169.1	**10.5**	40.7	117.9
1992	**219.4**	**11.0**	55.4	153.0
1994	252.3	**16.2**	59.4	176.7

Allocating the Subaggregate Control Totals to Individual Records

The next step is allocate these subaggregates to individual family records. One procedure could be to compute the average family MOOP expenditure for each subgroup and then assign this average value to each record. However, given the rather skewed distribution of MOOP spending, this procedure would greatly overstate the amount of MOOP for the majority of families and hence potentially lead to an overstatement of poverty in the population.

For the NRC Panel, AHCPR produced tables from the 1987 NMES file which had been aged to 1992. These tables provided information on the cumulative distribution of MOOP for households which had MOOP expenditures as well as the percentage of households whom had MOOP expenditures. These detail tables were produced for various subgroups of the population defined by type of insurance coverage, age of the head of the family, race, income, and family size. Using the information, I was able to estimate the probability that a household with a given set of characteristics (X_h) would have MOOP. Let us denote this probability by

$$P(X_h)$$

I was also able to estimate for Elderly and NonElderly populations of the following form describing the cumulative distribution of MOOP :

$$\mathrm{Ln}(C/(1-C)) = \alpha + \beta X_h + \gamma \, \mathrm{LNMOOP} \qquad (1)$$

where

C = the percentile in the MOOP distribution

X_h = a vector of family characteristics (age, race, income, and insurance coverage)

LNMOOP = Log of Moop Spending.

The estimated relationship (α,β,γ) was utilized in the following manner. For each household, a uniform random number was drawn from a random number generator, RN_1. If RN_1 was less than $P(X_h)$ then the household would be assigned a level of MOOP otherwise the household would be assigned a zero value. If the household was to be assigned a non zero value of MOOP, a second random was drawn, RN_2, which was to represent the percentile in the MOOP distribution to which the family was to be assigned. The level of MOOP that corresponds to this percentile was then estimated as

$$\mu \ \mathrm{x} \ \mathrm{EXP}\left[\frac{\mathrm{Ln}(RN_2/(1-RN_2)) - \alpha - \beta X_h}{\gamma} \right]$$

where μ is a proportional factor computed so that weighted sum of MOOP adds up to the control totals for the two subgroups of the population: the elderly and nonelderly.

This past summer (1997), I was able to acquire the micro (family) data from which the original AHCPR tables were constructed. Using this data I was able to reestimate the above equation (1) using the micro data instead of the aggregated tabular data. Given the larger sample size, I was able to estimate a more comprehensive model that included interaction terms of the family characteristics and level of MOOP. The description of variables used and regression results are provided below. Other than using the newly estimated regression results, the procedures to impute MOOP to the individual records remains the same as utilized for the NRC Panel report.

Description of Independent Variables

LNMOOP	Log of Medical Out of Pocket Expenses
PUBLIC	1 if Insured by Medicare or Medicaid only; 0 otherwise
UNINS	1 if Uninsured; 0 otherwise
FS23	1 if Family Size is 2 or 3; 0 otherwise
FS4M	1 if Family Size is 4 or more; 0 otherwise
FS2M	1 if Family Size is 2 or more; 0 otherwise
AGE75	1 if Head is 75 years or older; 0 otherwise
NONPOOR	1 if the ratio of the Family's Census Money Income to Poverty Line exceeds 1.50; 0 otherwise
BLACK	1 if Black; 0 otherwise
PUBLMP	= LNMOOP * PUBLIC
UNLMP	= LNMOOP * UNINS
NPLMP	= LNMOOP * NONPOOR
F23LMP	= LNMOOP * FS23
F4MLMP	= LNMOOP * FS4M
F2MLMP	= LNMOOP * FS2M
A75LMP	= LNMOOP * AGE75

Regression Model for NonElderly Population :

```
Ordinary    least squares regression.     Dep. Variable     =  LNODDS
Observations       =            5328      Weights           =  ONE
Mean of LHS        = -0.3648372D-01       Std.Dev of LHS    =  0.1757924D+01
StdDev of residuals=  0.4893229D+00       Sum of squares    =  0.1272368D+04
R-squared          =  0.9227089D+00       Adjusted R-squared=  0.9225198D+00
F[ 13,  5314]      =  0.4879926D+04       Prob value  0.3217295D-13
Log-likelihood     = -0.3744999D+04       Restr.(ß=0) Log-l = -0.1056531D+05
ANOVA  Source         Variation     Degrees ofFreedom,       Mean Square
       Regression     0.1518964D+05          13.               0.1168434D+04
       Residual       0.1272368D+04        5314.               0.2394369D+00
       Total          0.1646201D+05        5327.               0.3090297D+01

Variable  Coefficient  Std. Error   t-ratio Prob|t|>x   Mean of X  Std.Dev.of X
--------------------------------------------------------------------------------
Constant   .90283       0.2401E-01   37.595  0.00000
LNMOOP     1.2549       0.1397E-01   89.848  0.00000  -.47910        1.5211
PUBLMP    -.40385       0.1442E-01  -28.007  0.00000  -.21546        .86030
UNLMP     -.13039       0.1377E-01   -9.467  0.00000  -.19744        .74545
NPLMP     0.64415E-01   0.1184E-01    5.441  0.00000  -.19158        1.1969
F23LMP    0.92141E-01   0.1105E-01    8.338  0.00000  -.13601        .98365
F4MLMP     .14905       0.1251E-01   11.912  0.00000 -0.35087E-01    .75400
BLKLMP   -0.69204E-01   0.1100E-01   -6.290  0.00000  -.18269        .81236
PUBLIC     1.2560       0.2782E-01   45.142  0.00000   .13570        .34250
UNINS      1.0070       0.2266E-01   44.449  0.00000   .18131        .38531
FS23      -.87023       0.1916E-01  -45.427  0.00000   .43581        .49591
FS4M      -1.1897       0.2036E-01  -58.434  0.00000   .28697        .45239
NONPOOR   -.19126       0.2006E-01   -9.533  0.00000   .74474        .43605
BLACK      .38658       0.1912E-01   20.215  0.00000   .19839        .39882
```

Regression Model for Elderly Population :

```
Ordinary    least squares regression.     Dep. Variable     =  LNODDS
Observations       =            2173      Weights           =  ONE
Mean of LHS        = -0.1392984D+00       Std.Dev of LHS    =  0.1831889D+01
StdDev of residuals=  0.4271742D+00       Sum of squares    =  0.3950644D+03
R-squared          =  0.9457987D+00       Adjusted R-squared=  0.9456234D+00
F[  7,  2165]      =  0.5396953D+04       Prob value  0.3217295D-13
Log-likelihood     = -0.1231072D+04       Restr.(ß=0) Log-l = -0.4398273D+04
ANOVA  Source         Variation     Degrees ofFreedom,       Mean Square
       Regression     0.6893768D+04           7.               0.9848240D+03
       Residual       0.3950644D+03        2165.               0.1824778D+00
       Total          0.7288833D+04        2172.               0.3355816D+01

Variable  Coefficient  Std. Error   t-ratio Prob|t|>x   Mean of X  Std.Dev.of X
--------------------------------------------------------------------------------
Constant   .50786       0.1970E-01   25.777  0.00000
LNMOOP     1.2170       0.1288E-01   94.459  0.00000  0.10989E-01    1.3975
NPLMP      .44104       0.1451E-01   30.387  0.00000   .19703        .92900
A75LMP   -0.51473E-01   0.1341E-01   -3.837   .00012  0.28515E-01    .95821
F2MLMP    -.18101       0.1420E-01  -12.750  0.00000   .11100        .93611
AGE75     -.26820       0.1895E-01  -14.151  0.00000   .43580        .49598
FS2M      -.46551       0.1959E-01  -23.768  0.00000   .46710        .49903
NONPOOR   -.63639       0.2052E-01  -31.019  0.00000   .61528        .48664
```

Code Segments for Imputation of MOOP (FORTRAN)

here is the call from the main routine :

where

```
hage=reference person's age
sage=spouse's age
insstat = insurance status of reference person
premimum = yearly medicare part b premimum


 iold=0
 if(hage.gt.64) then
   out(21)=oldoop(iseed)
   iold=1
   if(sage.gt.64)iold=2
   iage=2
  else
   out(21)=yngoop(iseed)
   if(sage.gt.64) iold=1
   iage=1
 end if
 medicare=0
 if(insstat.gt.0.and.insstat.ne.4) then
  medicare=float(iold)*premimum
  if(ew.eq.0.0) medicare=0
 end if
 out(21)=out(21)+medicare
```

function yngoop(iseed)

```
real cof(7),dum(7),ycof0(7)
real ypzero(37)
data ypzero/
  & .065,.041,.075,.143,.061,.083,.012,.012,.031,.024,.003,.006,
  & .397,.606,.219,.628,.371,.408,.212,.279,.237,.507,.256,.345,
  & .378,.482,.248,.420,.151,.194,.103,.128,.043,.126,.036,.213,
  & 0./
data cof/.90283,1.256,1.007,-.87023,-1.1897,-.19126,.38658/
data ycof0/1.2549,-.40385,-.13039,.092141,.14905,.064415,-.0692/

yngoop=0.0

do k=2,7
 dum(k)=0.0
repeat

if(insstat.eq.0) then  ! insurance status of reference person
 insure=3                     ! uninsured
 dum(3)=1.0
else if(insstat.eq.2.or.insstat.eq.4) then
 insure=2                     ! public insurance only
 dum(2)=1.0
else
 insure=1                     ! private
end if


ifam=xin(13)

if(ifam.eq.1) then
 isize=1                ! family size = 1
else if(ifam.eq.2.or.ifam.eq.3) then
 isize=2                ! family size 2 or 3
 dum(4)=1.0
else
 isize=3                ! family size is four or more
 dum(5)=1.0
end if

rneeds=0.0        ! census money income to needs (poverty line) ratio
pline=xin(14)
```

```
      if(pline.ne.0.0) rneeds=cminc/pline
      if(rneeds.lt.1.5) then
       ipoor=1
      else
       ipoor=2
       dum(6)=1.0
      end if

      if(xin(29).ne.2) then                         !  race
       irace=1                                      ! non black
      else
       irace=2                                      ! black
       dum(7)=1.0
      end if
      ipl=(insure-1)*12+(isize-1)*4+(ipoor-1)*2+irace
      if(ran1(iseed).lt.ypzero(ipl)) return
      cons=cof(1)
      do k=2,7
       cons=cons+dum(k)*cof(k)
      repeat
      ycof=ycof0(1)
      do k=2,7
       ycof=ycof+dum(k)*ycof0(k)
      repeat

c ran1 is an uniform random number generator RN[0,1]

c
c If imax99 equals 1 then the distribution of MOOP is bounded at the
c pmax percentile -- this is an alternative way to bound "high values"
c of MOOP compared to the Pat Doyle's way (imax99 equal 2) which limits
c the elderly to $8200 of MOOP which she estimated to be 99th percentile
c of the elderly MOOP Distribution -- see below
c


      p=ran1(iseed)
      if(imax99.eq.1) p=p*pmax
      odds=alog(p/(1.-p))
      yngoop=1000.*exp((odds-cons)/ycof)
      if(imax99.eq.2) then
       yngoop=amin1(amax1(1.,yngoop),8200.)
      end if
      yngoop=yngoop*yfac     ! yfac multiplicative factor to hit aggregrate
      return
      end
```

```
    function oldoop(iseed)
    real dum(4),cof(4),ycof0(4)
    real opzero(9)
      data opzero/.1666,.0233,.1010,.016,.0872,.0220,.0536,.0165,0./
    data cof/.50786,-.2682,-.46551,-.63639/
    data ycof0/1.2170,-.05147,-.18101,.44104/

    oldoop=0.0

    do k=2,4
     dum(k)=0.0
    repeat


    if(hage.lt.75) then     ! age of reference person
     iage=1
    else
     iage=2
     dum(2)=1.0
    end if

    if(xin(13).eq.1) then   ! family size
     isize=1
    else
     isize=2
     dum(3)=1.0
    end if

    rneeds=0.0              ! census money income to needs (Poverty line) ratio
    pline=xin(14)
    if(pline.ne.0) rneeds=cminc/pline
    if(rneeds.lt.1.5) then
     ipoor=1
    else
     ipoor=2
     dum(4)=1.0
    end if

    ipl=(iage-1)*4+(isize-1)*2+ipoor

    if(ran1(iseed).lt.opzero(ipl)) return  ! pzero is the probability of not
having MOOP

    ycof=ycof0(1)
    do j=2,4
     ycof=ycof+dum(j)*ycof0(j)
    repeat

    cons=cof(1)
    do k=2,4
     cons=cons+dum(k)*cof(k)
    repeat

    p=ran1(iseed)    ! p is the random percentile in the MOOP distribution

c ran1 is an uniform random number generator RN[0,1]


c
c If imax99 equals 1 then the distribution of MOOP is bounded at the
c pmax percentile -- this is an alternative way to bound "high values"
c of MOOP compared to the Pat Doyle's way (imax99 equal 2) which limits
c the elderly to $18000 of MOOP which she estimated to be 99th percentile
c of the elderly MOOP Distribution -- see below
c


    if(imax99.eq.1) p=p*pmax
```

```
odds=alog(p/(1.-p))


oldoop=1000.*exp((odds-cons)/ycof)

if(imax99.eq.2) then
 oldoop=amin1(amax1(1.,oldoop),18000.)
end if

oldoop=oldoop*ofac  ! ofac = a multiplicative factor to hit aggregate total

return
end
```

Appendix A

NMES Regression Results for the Non Elderly