# Imputing medical out of pocket (MOOP) expenditures using SIPP and MEPS* 

Sharon I. O'Donnell ${ }^{\dagger}$<br>and<br>Rodney Beard ${ }^{\ddagger}$


#### Abstract

In this paper, we compare two parametric statistical matching methods used to generate synthetic medical out of pocket expenditure (MOOP) data. The methods differ by the mechanism used to assign synthetic values to the observations on the recipient survey. The model developed by Betson (2001) uses random assignment while the predicted mean matching method is a parametric version of cold deck imputation. Sensitivity analysis determined the extent to which the models preserved marginal and joint distributions of the host survey data and tested the overall stability of each model.

In a comparison of data generated from the Betson model using 1996-1997 Consumer Expenditure data versus host survey data current to the recipient survey, results suggest that aging data using the CPI is not a good substitute for using current data. Results found that the model based on the random assignment mechanism performed as well as the model based on cold deck imputation. Analysis indicated that both models were sensitive to differences in the host survey data and indicator variables. Future work will concentrate on improving the predicted mean matching model as the method used to generate synthetic MOOP data.


[^0]
## Introduction

In the best of all possible worlds, data required for a policy measure are extracted from a single source that represents the decisions made by the person, family or household. For the experimental poverty measure, it would require the creation of an annual national survey that collects timely information on income, taxes, program benefits, food, shelter, child care, work-related, and medical out of pocket (MOOP) expenditures. To produce a survey that collects these data in a timely manner is prohibitively expensive, both in terms of the actual production costs and the burden faced by respondents.

The response to this problem is to choose a survey that is timely with many of the needed variables and, using statistical matching techniques (Rässler 2002, D'Orazio, Di Zio, and Scanu 2006), impute missing data from a data source (host survey) to this recipient survey. To use statistical matching techniques, the host and recipient data sets must be representative samples of the same population, share a set of common variables and have no households in common. The host survey should include variable(s) of interest that are not found on the recipient survey, (e.g., variables are missing completely at random (MCAR) (D'Orazio, Di Zio, and Scanu 2006)).

Government measures generated from a single survey source are estimates of their true measures. Because the values are obtained from a single source, standard sampling techniques can generate a confidence interval that measures the interval estimate of the true parameter. When multiple sources of data are used to generate a measure, it may lead to a measure that is biased if a source does not reflect the population of the recipient survey or if the matching mechanism is not properly specified.

The experimental poverty measure is based on multiple data sources and, as such, is a synthetic point estimate. In this paper, we examine two methods to impute medical out of pocket (MOOP) expenditures to the recipient survey (Annual Social and Economic Supplement (ASEC) of the Current Population Survey (March CPS)). MOOP enters the experimental poverty measure as an expenditure that is subtracted from the total annual income of the CPS family. Our assessment of the methods is a comparison of net income from the host survey with net income based on total family income from the CPS and synthetic MOOP.

Presently, the Census uses a method developed by Betson (2001). The Betson model generates synthetic data based on MOOP expenditure data from the 1996-1997 Consumer Expenditure Survey. Synthetic MOOP data generated by the model are aged using the Consumer Price Index.

In this paper, we compare the Betson model with a predicted mean matching method. The main difference between the two parametric methods is that the Betson model uses a random assignment while the predicted mean matching method is based on a cold deck imputation method. In this paper, we will determine if random assignment matters.

Methods are expected to meet two standards. Using statistical matching techniques, we assessed each method in terms of its ability to generate a distribution of expenditures that replicate the expected distribution that would exist if the March CPS included the variables. Each method will be assessed based on its ability to preserve the joint distributions of the indicator variables and MOOP. The second standard is to test the stability of each method. The method is expected to be invariant to the choice of the survey identified as an appropriate host survey. The method also is expected to be stable over time and when there is a change in the set of common variables. A method is considered to be stable if it can generate results given these changes.

To introduce our work, we begin with a description of the host and recipient surveys,
the data of interest and the common variables used as indicator variables. This is followed by descriptions of the two methods. The results section summarizes our assessment of these methods. The paper ends with a discussion of our findings.

## Data

## Recipient and Host Surveys

The Annual Demographic Survey of the Current Population Survey (or March CPS) is an annual survey produced by the Bureau of Labor Statistics and the US Census Bureau. It is used to report estimates of geographic mobility, educational attainment and poverty status and to profile the annual US labor market and general population (U.S. Census Bureau 2001). It is the base (recipient) survey for the experimental poverty measure. Health insurance information is collected annually. Due to the survey design of the Current Population Survey, MOOP expenditures are not collected.

The present method to imput MOOP expenditure values to the March CPS is based on MOOP expenditure data from the 1996-1997 Consumer Expenditure Survey. Stakeholders raised the concern of the age of the medical expenditure data in the Betson model. The model is based on data that is slightly more than ten years old. MOOP expenditure values generated by the model are converted into current dollars using the consumer price index but researchers and policy analysts suggested that this may not be sufficient.

In this analysis, two host surveys are used as sources of data for MOOP expenditure data. The Census Bureau's Survey of Income and Program Participation (SIPP) provides "information about the income and program participation of individuals and households in the United States, and about the principal determinants of income and program participation" (U.S. Census Bureau 2006). Depending on the panel, individuals and families were interviewed every four months for two and one half to four years using a core interview and interviews on topics including program participation, health expenditures, disability, assets and liabilities, work-related expenses and child support. Health insurance information is collected every four months and MOOP expenditures annually. ${ }^{1}$ Medical Expenditure Panel Survey (MEPS) is produced by the Agency for Healthcare Research and Quality of the US Department of Health and Human Services. It contains an extensive list of health utilization, health insurance and medical expenditure data. Household members, insurance providers and employers are the sources of information. The panel is based on an overlapping panel design. To collect two years of data, individuals are interviewed five times over a two and one half year period. MOOP expenditure data are collected for specific medical/health events (hospital stays, emergency medical services, outpatient care, medical provider visits, home care, prescription drugs). Adult household members maintain a medical event calendar to facilitate recall (Agency for Healthcare Research and Quality 2009).

All three surveys are representative samples of the US population at the national level. For our analysis, the person-level data were aggregated as CPS families and individuals living in non group quarters. A CPS family is "a group of two people or more (one of whom is the householder) related by birth, marriage, or adoption and residing together; all such people (including related subfamily members) are considered as members of one family" (U.S. Census Bureau 2008a). For the purpose of this analysis, we selected recipient and host survey data for calendar year 2004.

[^1]
## Imputed variables

The analysis will examine two variables assigned to each CPS family observation, medical out of pocket expenditures (MOOP) and net income. MOOP expenditures are the annual expenditures for traditional health services and it includes the amount paid by family members for health insurance. Health services include visits to medical professionals, hospital visits and home care. The cost of prescription drugs, eyeglasses and contacts are included but the cost of over the counter medicine is not included. Reported MOOP expenditures take into account any reimbursements.

The second variable is net income $N(C P S)$ which is obtained by subtracting the synthetic MOOP expenditure from the CPS family's total annual income. Net income is an important variable in computing the experimental poverty measure based on a version known as MSI-NGA (Medical out-of-pocket expenses (MOOP) subtracted from income). MOOP enters the poverty measure as an expenditure used to calculate net income. Comments from one researcher noted that generating a large number of poverty rates using the current method produced a distribution of rates with significant variation around the mean. ${ }^{2}$ The variation may be due to the model's random assignment of MOOP to each observation. We cannot observe the percent of families in poverty using a single data source but we can observe the distribution of net income from surveys that include income and medical expenditure variables. If the Betson model generates excessive volatility, we will observe it directly in net income. Minimizing variation in net income is expected to minimize the variation in the poverty rate.

## Variables in common

Variables in common match observations from the host to the recipient survey. In the Betson model, families were grouped using four variables, a variable indicating that the reference person is 65 years old or older (dichotomous), health insurance coverage (three categories for non elderly families, two for elderly), family size (three categories for no elderly families, two for elderly) and income as a percent of poverty (dichotomous, using $150 \%$ as the cutoff).

With respect to health insurance, families are defined as "private" if any member had private health insurance during the calendar year, "public" if family members who were insured carried only public health insurance and "uninsured" if all members were uninsured during the calendar year. The differences among the surveys suggest that matching host data with the CPS via the insurance coverage variable may generate some mismatches. The CPS questions ask the family members to recall coverage over a longer reference period compared to the host surveys. Compared to SIPP and MEPS families, CPS families are more likely to forget to report coverage of short duration or coverage that only occurred in the early part of the reference period. SIPP and CPS obtain insurance coverage information from only one source (the members of the households). MEPS obtains this information initially from the household members but supplemented this information with data from medical providers and employers. Given the data quality of the two outside sources, the additional data sources used by MEPS are expected to influence respondent error. For this paper, we wish to note these issues but will not address them in this analysis.

From these variables, the survey data are partitioned into $F$ family groups. For each family group, it is assumed that the mean synthetic MOOP expenditure value $\bar{M}(C P S)_{i}$, $i=(1, \ldots, F)$ is equal to the expected mean MOOP value (or the mean values generated from a version of the CPS with MOOP expenditure variables) for each family group in the CPS.

[^2]We created an alternate set of indicator variables. It includes a measure of the overall health of the family based on self reported health status. It is a dichotomous variable indicating if all family members are in good to excellent health or if any member is in fair to poor health. We replaced the income variable in the original set of indicator variables to one based on $25^{t h}, 50^{t h}$, and $75^{t h}$ percentile income cutoffs in the CPS. If there is sufficient data in the host survey file, there may be up to four income categories for each family group.

Tables A-1 and A-2 in the appendix section list the population estimates from the CPS, SIPP and MEPS surveys by the original and revised indicator variables. Tables include population estimates and percent of total population. Table A-3 lists the estimated $25^{t h}$, $50^{t h}$, and $75^{t h}$ percentiles income cutoffs in the CPS by family group and the income breakdowns used the analysis. If the table indicates that "quartiles" were used, this means that we further partitioned the family group into four income groups in the host and recipient surveys using the quartile values. If it is a "median" breakdown, we partitioned the group into two groups using the median income value.

## Methods

## Betson Model

Betson (2001) is a version of a statistical matching model. MOOP expenditure data were extracted from the 1996 and 1997 Consumer Expenditure Survey (CEX) files. Data were retrieved from families with at least three completed interviews. From Betson (2001, page 22),

The CEX data collects the unitnet of the amount that reimbursed by insurance or any government program as well as the cost to the unit of any health care insurance including Medicare Part B premiums.

The model assumed that the heterogenous population influences the distribution of expenditures values. Betson categorized CEX families based on the age of the reference person, income as a percent of poverty, health insurance coverage and family size. Based on these characteristics, CEX families and the families in the recipient survey (March CPS) were categorized into $F$ family groups. In order to generate synthetic data, each family group required a minimum number of observations from the host surveys. There were 49 family groups identified.

The Betson model generated synthetic data in two ways. To generate zero expenditure data, the estimated percent of families ( $P_{z e r o, i}$ ) with zero expenditure in the host survey was calculated for each family group. CPS family observations were categorized by family group and assigned a random number $\left(U_{1}\right)$ from the uniform distribution. If $U_{1} \leq P_{z e r o, i}$, the CPS family observation was assigned a MOOP value of zero.

Distribution of the nonzero medical out of pocket expenses is assumed to be log-logistic with the CDF given as

$$
\operatorname{Pr}\left(m_{i} \leq M_{i}\right)=C\left(M_{i}\right)=\frac{1}{1+\exp \left(-\beta_{0, i}-\beta_{1, i} \ln \left(M_{i}\right)\right)}, \quad i=(1, \ldots, F)
$$

or

$$
\begin{equation*}
\ln \left[\frac{C\left(M_{i}\right)}{1-C\left(M_{i}\right)}\right]=\beta_{0, i}+\beta_{1, i} \ln \left(M_{i}\right) \tag{1}
\end{equation*}
$$

where $M_{i}$ is the nonzero MOOP values and $C\left(M_{i}\right)$ is the CDF values for observations in the host survey categorized as family group $i$.

To generate synthetic nonzero MOOP values, equation 1 is inverted.

$$
\begin{equation*}
m_{i}=\exp \left[\frac{\ln \left(\frac{C\left(M_{i}\right)}{1-C\left(M_{i}\right)}\right)-\beta_{0, i}}{\beta_{1, i}}\right] \tag{2}
\end{equation*}
$$

To generate synthetic values, $C(M)$ is replaced with a value generated from the uniform distribution, a standard numerical method (Lange 1999) for producing simulated values.

In estimating the model, Betson determined that cubic polynomial version of the equation generated the best fit. For his analysis, equation 1 was rewritten as

$$
\begin{equation*}
\ln \left[\frac{C\left(M_{i}\right)}{1-C\left(M_{i}\right)}\right]=\beta_{0, i}+\beta_{1, i} \ln \left(M_{i}\right)+\beta_{2, i}\left(\ln \left(M_{i}\right)\right)^{2}+\beta_{3, i}\left(\ln \left(M_{i}\right)\right)^{3}, \quad i=(1, \ldots, F) \tag{3}
\end{equation*}
$$

Inverting equation 3 to generate expenditure values produces a function that is quite complex and difficult to maintain in production code. To generate expenditures, Betson used numerical methods. For observations in the recipient survey where $U_{1}>P_{z e r o, i}$, the program assigned a second random number $\left(U_{2}\right)$ for $C\left(M_{i}\right)$ and using the estimated parameters from the host survey, numerically solved

$$
\begin{equation*}
m_{i}=h\left(U_{2} ; \beta_{0, i}, \beta_{1, i}, \beta_{2, i}, \beta_{3, i}\right) \tag{4}
\end{equation*}
$$

## Predicted mean matching

Predicted mean matching is a parametric statistical matching method. Using the host survey, the variable of interest $(M)$ is regressed on the common variables and a predicted value of the dependent variable $\left(\hat{M}_{h}\right)$ is generated for each observation in the host survey. The estimated parameters from the regression model are used to generate predicted values $\left(\hat{M}_{r}\right)$ for each observation in the recipient survey. The observations are combined. For each observations in the recipient survey, the host survey observation is chosen when the absolute difference between the predicted values from the host and recipient surveys is minimized $\left(\min \left(\operatorname{abs}\left(\hat{M}_{h}-\hat{M}_{r}\right)\right)\right)$. When a match is obtained, the MOOP value from the host observation is copied to the recipient observation. The assignment of MOOP is done with replacement. For this analysis, the predicted values of the recipient survey observations were compared with the predicted values of host survey observations within family groups.

Predicted mean matching can be constrained or unconstrained. In constrained matching, observations from the recipient survey are removed or duplicated until the distribution of the synthetic variable is equal to the distribution in the host survey. For weighted data, the process may involve revising weights assigned to the observations (Rodgers 1984). In our case, we chose an unconstrained approach because it is important to retain the weights and observations in the CPS.

## The approach used in the Betson model

The Betson and predicted mean matching models rely strongly on the unbiasedness property of OLS estimators by substituting their estimated parameters for population parameters. For example, if the population regression model is,

$$
\begin{equation*}
Y=\beta_{0}+\beta_{1} \ln (M)+\beta_{2}(\ln (M))^{2}+\beta_{3}(\ln (M))^{3}+\epsilon \tag{5}
\end{equation*}
$$

the estimated regression equation is

$$
\begin{equation*}
Y=\hat{\beta}_{0}+\hat{\beta_{1}} \ln (M)+\hat{\beta_{2}}(\ln (M))^{2}+\hat{\beta_{3}}(\ln (M))^{3} \tag{6}
\end{equation*}
$$

where $\hat{\beta}$ is the estimate of the true parameter $\beta$.
The Betson model replaces $\hat{\beta}_{i}$ by $\beta$ thereby assuming that the estimated parameters are identical to the unknown population parameters. It also assumes that the sample error is zero. This results in

$$
\begin{equation*}
Y=\beta_{0}+\beta_{1} \ln (M)+\beta_{2}(\ln (M))^{2}+\beta_{3}(\ln (M))^{3} \tag{7}
\end{equation*}
$$

At first glance, it appears that this is simply an application of the unbiasedness property of OLS. When one inverts the equation using the Betson method, it is not done to obtain the mean value of MOOP, given the mean predicted CDF value, but to generate individual observations for CDF values from 0.0 to 1.0.

Obtaining the solution for this problem does not involve finding the roots of a polynomial. The parameter estimates $\left(\hat{\beta_{0}}, \hat{\beta_{1}}, \hat{\beta_{2}}\right.$ and $\left.\hat{\beta_{3}}\right)$ are functions of the dependent and independent variables of the regression model. To show this, we rewrite equation 7 to obtain
$Y=\hat{\beta_{0}}(Y, \ln (M))+\hat{\beta_{1}}(Y, \ln (M)) \ln (M)+\hat{\beta_{2}}(Y, \ln (M))(\ln (M))^{2}+\hat{\beta_{3}}(Y, \ln (M))(\ln (M))^{3}$.

The corresponding normal equations are:

$$
\begin{aligned}
& \sum Y_{i}=n \beta_{0}+\beta_{1} \sum \ln \left(M_{i}\right)+\beta_{2} \sum \ln \left(M_{i}\right)^{2}+\beta_{3} \sum \ln \left(M_{i}\right)^{3} \\
& \sum Y_{i} \ln \left(M_{i}\right)=\beta_{0} \sum \ln \left(M_{i}\right)+\beta_{1} \sum \ln \left(M_{i}\right)^{2}+\beta_{2} \sum \ln \left(M_{i}\right)^{3}+\beta_{3} \sum \ln \left(M_{i}\right)^{4} \\
& \sum Y_{i} \ln \left(M_{i}\right)^{2}=\beta_{0} \sum \ln \left(M_{i}\right)^{2}+\beta_{1} \sum \ln \left(M_{i}\right)^{3}+\beta_{2} \sum \ln \left(M_{i}\right)^{4}+\beta_{3} \sum \ln \left(M_{i}\right)^{5} \\
& \sum Y_{i} \ln \left(M_{i}\right)^{3}=\beta_{0} \sum \ln \left(M_{i}\right)^{3}+\beta_{1} \sum \ln \left(M_{i}\right)^{4}+\beta_{2} \sum \ln \left(M_{i}\right)^{5}+\beta_{3} \sum \ln \left(M_{i}\right)^{6} \\
& i=1, \ldots, n .
\end{aligned}
$$

The estimated equation along with the normal equations provide a system of equations that one uses to obtain the MOOP values. The estimated equation adds a restriction to the values that the estimated parameters may take on.

This system of equations in principle can be solved for $\ln (M)$ using a procedure such as Newton's method. It does not have an analytical solution (Galois' theorem). This is a large computational problem and will not necessarily have a numerical solution. Given that it is a large numerically intensive problem, one would only add it to production code if it was the only method to impute the distribution.

Although the method used by the model is incorrect, our objective was to replicate the analysis conducted by the stakeholder and determine if the observed variation of the rate was due to random assignment. This meant running the original model using parameters generated from SIPP and MEPS. The model is a cubic polynomial and has up to three real solutions. As written, the algorithm in the model only finds one of the three solutions. The solution obtained by this method may fall within the expected range of MOOP expenditures but, based on our experience, the method periodically finds one of the solutions that is well beyond the expected range. This is more likely to occur when the random number selected is large and the predicted CDF values generated from the estimated model include MOOP expenditures outside of the expected range. ${ }^{3}$ Given this, initially, the model appeared to be unstable.

As a way to restrict the range, we made two revisions to the original model. First, we restricted the random numbers (substituting for the predicted CDF values) to those that generate MOOP values between $\$ 0.01$ and the MOOP value at the 99 th percentile. We replaced the Newton algorithm with one based on the bisection method, a root finding algorithm used to find one solution within a range of values.

Again, the reader should note that these modifications do not resolve the main problem with the model but only allow us to replicate earlier analysis. Additionally, the two revisions do not resolve the problem for those cases where two or three solutions exist within the bounded range.

## Statistics

In assessing each method, we considered how the synthetic data replicated the distribution from the host file. To do this, we ran difference of means test for each partition group. If $\bar{M}($ Host ) is the mean MOOP estimate from the host survey data and $\bar{M}(C P S)$ is the synthetic mean MOOP estimate from the CPS, the distribution is preserved if

$$
\begin{equation*}
\bar{M}(C P S)-\bar{M}(H o s t)=0 . \tag{9}
\end{equation*}
$$

If the test rejects this hypothesis, it suggests that the distribution of synthetic data does not reflect the survey data distribution.

Net income is annual family income minus MOOP expenditures $(N=I-M)$. Given this, when we compare the difference in net income or,

$$
\begin{equation*}
\bar{N}(C P S)-\bar{N}(\text { Host })=? \tag{10}
\end{equation*}
$$

we can replace equation 10 with

$$
\begin{equation*}
(\bar{I}(C P S)-\bar{M}(C P S))-(\bar{I}(\text { Host })-\bar{M}(H o s t))=? \tag{11}
\end{equation*}
$$

[^3]If the distribution is preserved, $\bar{M}(C P S)=\bar{M}($ Host $)=\bar{M}$ so we can rewrite equation 11 as,

$$
\bar{I}(C P S)-\bar{I}(\text { Host })=?
$$

Although SIPP and CPS are national surveys, it may not be the case that the mean annual income estimates are equal within each family group ${ }^{4}$. To test net income, we used the following hypothesis,

$$
\begin{equation*}
(\bar{N}(C P S)-\bar{N}(\text { Host })) \lesseqgtr 0 \text { if }(\bar{I}(C P S)-\bar{I}(\text { Host })) \lesseqgtr 0 \tag{12}
\end{equation*}
$$

Tables A-4 through A-6 provide the results of the tests for differences in the estimated mean annual income for CPS and the two host surveys.

If the mean CPS income is greater than the mean host income at the $10 \%$ significance level, we expect that this relationship holds in the difference in mean net income.

The Betson model generates synthetic MOOP expenditure values using simulation methods. For this analysis, we generated 1,000 random draws/samples for each model and calculated the weighted mean MOOP and net income estimates. Each mean value presented in the tables is the mean from the distribution of means generated by the 1,000 simulations. The reported standard errors are the ones calculated from these distributions.

For the survey data, the standard errors for means were prepared using methods developed for SIPP (U.S. Census Bureau 2009, pages 13-14), MEPS (Machlin, Yu, and Zodet 2005) and CPS (U.S. Census Bureau 2008b). The methods for generating standard errors for the survey data were used to generate the standard error for synthetic net income (SIPP, MEPS). CPS replicate weights were used to generate the standard errors for the synthetic mean MOOP estimates generated by the predicted mean matching method. Internal files were used to prepare the results for SIPP and CPS, public use files were used to prepared the results for MEPS. The income values in the CPS and the income and medical expenditures values in SIPP are not top-coded. For MEPS, MOOP is not top-coded but income is top-coded. ${ }^{5}$

## Results

Table 1 is a summary of the three main variables of interest from the two host surveys and the recipient survey. Although all surveys are representative samples of the national population, the mean MOOP and net income estimates differ between the two host surveys and mean annual income estimates differ among the three surveys. Mean MOOP, net income and annual income estimates produced from MEPS are greater than the estimates from SIPP and the mean annual income estimates from the March 2005 CPS is greater than the estimates from the two host surveys.

Tables in the appendix compare the estimated mean annual income of the recipient and host surveys for the original indicator variables (Tables A-4) and the revised variables (Tables A-5 (SIPP) and A-6 (MEPS)). Differences in mean income between CPS and the host survey are also indicated.

[^4]|  | Mean MOOP |  | Mean Net Income |  | Mean Income |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Survey | Estimated | $\pm 90 \%$ | Estimated | $\pm 90 \%$ | Estimated | $\pm 90 \%$ |  |
| Survey of Income <br> and Program <br> Participation <br> (CY 2004) | $2,498.82$ | 69.60 | $38,848.68$ | 552.01 | $41,171.27$ | 557.32 |  |
| Medical |  |  |  |  |  |  |  |
| Expenditure <br> Panel Survey <br> (CY 2004) | $2,683.11$ | 67.95 | $49,129.65$ | $1,171.37$ | $51,736.67$ | $1,200.73$ |  |
| Current <br> Population |  |  |  |  |  |  |  |
| Survey <br> (March 2005) |  |  |  |  |  |  |  |

Table 1: Summary results - estimated means of medical out of pocket expenditures, net income and annual income by survey source.

Table 2 summarizes the results from the two methods. The estimated synthetic mean MOOP expenditures vary from $\$ 1,987$ (Betson model using 1996-1997 CE data) to $\$ 2,693$ (results based on predicted mean matching using SIPP and the original indicator variables). Mean estimates between the two host surveys are statistically different as well as the estimates between the two methods controlling for the host survey. The use of different indicator variables produced mean estimates that differed for the Betson method, controlling for host survey, but did not differ for the predicted mean matching method.

Given the differences, there are no consistent patterns. Using the Betson model, the mean estimates generated using SIPP are less than the ones using MEPS but, using the predicted mean matching method the synthetic mean estimates based on MEPS are less than the ones for SIPP.

| Host Survey | Method | Indicator Variables | Recipient Survey 2005 CPS |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | MOOP (M(CPS)) |  | Net income (N(CPS)) |  |
|  |  |  | Est | $\pm 90 \%$ | Est | $\pm 90 \%$ |
| Survey of Income | Betson | Original | 2,266.64 | 9.40 | 52,404.89 | 10.01 |
| and Program |  | Revised | 2,297.65 | 6.62 | 52,451.64 | 0.53 |
| Participation | Predicted mean | Original | 2,693.57 | 50.11 | 52,093.33 | 352.08 |
| (CY 2004) |  | Revised | 2,665.25 | 28.76 | 52,019.33 | 352.21 |
| Medical | Betson | Original | 2,608.77 | 6.06 | 52,071.74 | 3.15 |
| Expenditure |  | Revised | 2,666.75 | 6.68 | 52,034.04 | 6.41 |
| Panel Survey | Predicted mean | Original | 2,530.13 | 21.87 | 52,135.66 | 352.90 |
| (CY 2004) |  | Revised | 2,489.37 | 22.31 | 52,168.56 | 214.83 |
| Consumer | Betson | Original | 1,986.65 | 0.50 | 54,086.15 | 0.49 |
| Expenditure |  |  |  |  |  |  |
| Survey |  |  |  |  |  |  |
| (96-97) aged |  |  |  |  |  |  |

Table 2: Summary results - estimated means of synthetic medical out of pocket expenditures and net income by host survey, method and indicator variable.

Mean net income estimates vary from $\$ 52,019$ (SIPP, predicted mean matching and revised indicators) to $\$ 54,086$ (aged 1996-1997 CE data). Mean net income estimates differ between SIPP and MEPS for the Betson model but estimates are not statistically different for the predicted mean matching model. Synthetic mean net income estimates did not differ by host survey and type of indicator variables.

The table describes a consistent pattern with respect to net income. Using the predicted mean matching method, the synthetic mean estimates for net income do not differ by host survey or by the set of indicator variables. Variations exist with results based on the Betson model. Part of this pattern is explained by the difference in the mean MOOP estimates. Holding host survey constant, the synthetic MOOP mean estimates generated using the predicted mean matching method did not differ by the set of indicator variables $(\$ 2,694$ and $\$ 2,665$ for SIPP and $\$ 2,530$ and $\$ 2,489$ for MEPS). Compare to the Betson model, the difference in mean MOOP estimates between SIPP and MEPS using the predicted mean method is less (\$163 and $\$ 175$ for original and revised indicator variables versus - $\$ 342$ and - $\$ 369$ for the Betson model) and is a difference that is well within the standard errors for the net income estimates.

The table also reports the estimated means for MOOP and net income using the CE data as the host survey. The model produces the smallest estimated mean for MOOP and largest estimated mean for net income. Detailed information for this model can be found in table A-7. The difference in means between this model and the models generated with data that are current to the recipient survey suggests that aging the data may not be sufficient in generating synthetic data that replicated the expected distribution of expenditure data for the recipient survey.

## Betson model

Tables A-8 through A-15 in the appendix report the results of our analysis comparing estimated means from survey and synthetic data for each family group and for the total population. Tables 3 and 4 summarize the results from these tables.

Table 3 summarizes the difference in means tests described in equation 12. The information compares the reported differences in annual income with the differences in net income. The percent of family groups that retained the income relationship in their net income relationship are found along the diagonal of each section. Table 4 report summary findings.

The off diagonal percents in each section of Table 3 indicate the percents of groups that deviated from the difference in mean income. The sections report very few cases where this occurred suggesting that most outcomes are consistent with our hypothesis. Using the revised indicator variables, MEPS reports the greatest deviation. Only half of the groups that indicated that the CPS mean income is less than the MEPS mean income estimates reported the same relationship for net income.

Table 4 summarize the information from Table 3 and supplements it with the results from the tests for the differences in estimated mean MOOP values. Synthetic data generated from SIPP produced the largest percent of cases that retained the expected distributions ( $87 \%$ ). Using the original indicator variables, $82 \%$ of the SIPP family groups failed to reject the null that there exists no difference between the survey-based and synthetic mean MOOP estimates. For net income, $94 \%$ of the groups categorized with the original indicator variables and $87 \%$ of the groups defined using the revised set had results consistent with equation 12 .

Results for MEPS suggest that the model is sensitive to variations in host survey and indicator variables. Compared to SIPP, the percent of groups that failed to reject the difference in survey-based and synthetic MOOP estimated varied considerably by the type of indicator variables ( $82 \%$ and $87 \%$ for SIPP versus $86 \%$ and $54 \%$ for MEPS). This difference is also present for the net income results ( $93 \%$ and $87 \%$ for SIPP and $90 \%$ and $65 \%$ for MEPS).

|  |  | Differences in mean income $\bar{I}(C P S)$ and $\bar{I}$ (Host) | Differences in mean net income $\bar{N}(C P S)$ and $\bar{N}($ Host $)$ |  |  | Row Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | CPS $<$ Host | CPS $=$ Host | CPS > Host |  |
| SIPP | Original |  | CPS < Host | 20.4 | 0.0 | 0.0 | 20.4 |
|  |  | CPS $=$ Host | 4.1 | 18.4 | 2.0 | 24.5 |
|  |  | CPS > Host | 0.0 | 0.0 | 55.1 | 55.1 |
|  | Revised | Column total | 24.5 | 18.4 | 57.1 | 100.0 |
|  |  | CPS < Host | 11.4 | 0.0 | 0.0 | 11.4 |
|  |  | $\mathrm{CPS}=$ Host | 7.1 | 28.6 | 2.9 | 38.6 |
|  |  | CPS > Host | 0.0 | 1.4 | 48.6 | 50.0 |
|  | Original | Column total | 18.6 | 30.0 | 51.4 | 100.0 |
| MEPS |  | CPS < Host | 14.3 | 2.0 | 0.0 | 16.3 |
|  |  | $\mathrm{CPS}=$ Host | 2.0 | 49.0 | 0.0 | 51.0 |
|  |  | CPS > Host | 0.0 | 6.1 | 26.5 | 32.7 |
|  | Revised | Column total | 16.3 | 57.1 | 26.5 | 100.0 |
|  |  | CPS < Host | 14.8 | 0.0 | 0.0 | 14.8 |
|  |  | $\mathrm{CPS}=$ Host | 16.7 | 31.5 | 18.5 | 66.7 |
|  |  | CPS > Host | 0.0 | 0.0 | 18.5 | 18.5 |
|  |  | Column total | 31.5 | 31.5 | 37.0 | 100.0 |

Table 3: Percent of family groups by the relationship between difference in mean annual income (CPS and host) and differences in mean net income (CPS (synthetic) and host) by Host survey and indicator variables - Betson Model.

## Predicted mean matching

Tables 5 and 6 report the findings for the predicted mean matching method. Detailed tables are found in the appendix (Tables A-16 through A-23).

As with the Betson model, a large percent of the groups retains the joint distribution from the host survey. Using the original indicator variables, the MOOP results for MEPS were stronger than SIPP ( $74 \%$ (MEPS) versus $67 \%$ (SIPP)) while the opposite holds for the results using the revised indicator variables ( $84 \%$ (SIPP) versus $78 \%$ (MEPS)). The differences in the results for SIPP suggests that the predicted mean matching model was sensitive to the choice of indicator variables.

With respect to the net income tests, the results for SIPP using the predicted mean matching methods were comparable to the Betson model ( $94 \%$ and $90 \%$ for predicted mean matching versus $94 \%$ and $87 \%$ for Betson). Results for MEPS indicated that the predicted mean matching model was less sensitive to differences in indicator variables compared to the Betson model ( $88 \%$ and $91 \%$ for predicted mean matching versus $90 \%$ and $65 \%$ for Betson).

Unlike the Betson model, the predicted mean matching method did not generate any stability problems. Comparision of repeated runs assured us that the method assigned the same MOOP value to each observation in the CPS.

## Discussion

This paper explores the use of two methods used to generate synthetic data. By examining net income, we determined if assigning the values randomly generated excessive volatility. Finally, we determined if the models were stable.

| Host <br> Survey | Indicator <br> Variables | Percent of groups preserving joint distribution |  |
| :---: | :---: | :---: | :---: |
|  |  | MOOP ${ }^{1}$ | Net income ${ }^{2}$ |
| SIPP | Original | 82.0 | 93.9 |
|  | Revised | 87.3 | 88.6 |
| MEPS | Original | 86.0 | 89.8 |
|  | Revised | 54.0 | 64.8 |

${ }^{1}$ Failed to reject hypothesis
that the difference in MOOP
(CPS synthetic and host) equals zero.
${ }^{2}$ Sum of diagonal values in table 3.

Table 4: Summary statistics by host survey and indicator variables - Betson Model.

Based on the results from the Betson model with aged CE data, it suggest that aging data may not reflect the current distribution of MOOP expenditures. The results support the recommendation from the stakeholders that the method used to generate the synthetic data should be based on current host survey data.

The results generated from the two methods do not suggest that random assignments reduce a method's performance. Except in the case when the model used MEPS with the revised indicator variables, the Betson model performs well with over $80 \%$ of the grouped data indicating that the joint distributions were preserved. If a random assignment approach introduced significant variation, one would expect the synthetic MOOP and net income values generated using the Betson model to generate weaker results compared to the results generated by the predicted mean matching method. This is not the case. Any volatility that was present in the original Betson model was more likely the result of the stability problems in the model.

The use of two host surveys and sets of indicator variables allowed us to determine if the models were sensitive to the source of data and model specification. Based on the net income findings, the Betson model did not perform as well when the host survey was MEPS and the data was partitioned using the revised indicator variables. With respect to the MOOP results, the Betson model also did not perform well with MEPS and revised indicator variables. The predicted mean matching model did not perform as well with host survey data from SIPP and the original indicator variables. Net income results based on the predicted mean matching method was consistent across host surveys and set of indicator variables. The synthetic mean net income values in table 2 based on the predicted mean matching model confirm these results.

The inconsistencies in both models suggest that both may be biased due to the strong assumption that their estimated parameters reflect the true parameters. If this is the case, one can improve a model by improving its model specification. Our future work will explore this.

There are two limitations to our reported results. First, results reflect the the distribution of expenditures at the national level and subsamples of the synthetic file are not expected to reflect the distribution of expenditures at the state and local levels. The underlying assumption in statistical matching is that the synthetic distribution reflects the distribution of the host surveys. Given that the recipient and host surveys are representative samples of the national population, the use of the synthetic distribution is limited to analysis at the national level.

The second limitation is that, by imputing only one of the expenditure variables required for the poverty measure, we are assuming that medical expenditures are independent of all

|  |  | Differences in mean income $\bar{I}(C P S)$ and $\bar{I}$ (Host) | Differences in mean net income $\bar{N}(C P S)$ and $\bar{N}($ Host $)$ |  |  | Row Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | CPS < Host | $\mathrm{CPS}=$ Host | CPS > Host |  |
| SIPP | Original |  | CPS < Host | 16.3 | 4.1 | 0.0 | 20.4 |
|  |  | CPS $=$ Host | 0.0 | 22.4 | 2.0 | 24.5 |
|  |  | CPS > Host | 0.0 | 0.0 | 55.1 | 55.1 |
|  |  | Column total | 16.3 | 26.5 | 57.1 | 100.0 |
|  | Revised | CPS < Host | 8.6 | 2.9 | 0.0 | 11.4 |
|  |  | CPS $=$ Host | 2.9 | 34.3 | 1.4 | 38.6 |
|  |  | CPS > Host | 0.0 | 2.9 | 47.1 | 50.0 |
|  |  | Column total | 11.4 | 40.0 | 48.6 | 100.0 |
| MEPS | Original | CPS < Host | 8.2 | 8.2 | 0.0 | 16.3 |
|  |  | $\mathrm{CPS}=\mathrm{Host}$ | 0.0 | 51.0 | 0.0 | 51.0 |
|  |  | CPS > Host | 0.0 | 4.1 | 28.6 | 32.7 |
|  |  | Column total | 8.2 | 63.3 | 28.6 | 100.0 |
|  | Revised | CPS < Host | 11.1 | 3.7 | 0.0 | 14.8 |
|  |  | CPS $=$ Host | 0.0 | 61.1 | 5.6 | 66.7 |
|  |  | CPS $>$ Host | 0.0 | 0.0 | 18.5 | 18.5 |
|  |  | Column total | 11.1 | 64.8 | 24.1 | 100.0 |

Table 5: Percent of family groups by the relationship between difference in mean annual income (CPS and host) and differences in mean net income (CPS (synthetic) and host) by Host survey and indicator variables - Predicted mean matching.

|  |  | Percent of family groups |  |
| :--- | :--- | :---: | :---: |
| Host | Indicator | preserving joint distribution <br> MOOP | Net income $^{2}$ |
| Survey | Variables |  | 93.9 |
| SIPP | Original | 67.4 | 90.0 |
|  | Revised | 84.3 |  |
| MEPS | Original | 73.5 | 87.8 |
|  | Revised | 77.8 | 90.7 |

${ }^{1}$ Failed to reject hypothesis
that the difference in MOOP
(CPS synthetic and host) equals zero.
${ }^{2}$ Sum of diagonal values from table 5.
Table 6: Summary statistics by host survey and indicator variables - Predicted mean matching
other expenditures given the set of common variables. The purpose of the paper is to test the two imputation methods so it is reasonable to conduct the tests using one variable. In the next stage of work, we will test the Conditional Independence Assumption (CIA) that was assumed in this work.

We determined that the mechanism used to generate synthetic expenditure values in the Betson model is not correct. In replicating the analysis conducted by the stakeholder, we needed to make major revisions to the model in order to generate synthetic data that reflected the the distributions in SIPP and MEPS. The revised version of model does not resolve the overall mechanism problem. We characterized the appropriate mechanism. Given the performance of the predicted mean matching model, it seems prudent to put our resources into improving the model specification used in the predicted mean matching method.

## References

Agency for Healthcare Research and Quality (2009): "Medical Expenditure Panel Survey Home," http://www.meps.ahrq.gov/mepsweb/index.jsp (accessed: May, 2009).

Betson, D. M. (2001):"Imputation of Medical Out of Pocket (MOOP) Spending to CPS Records," (technical report)
http://www.census.gov/hhes/www/povmeas/papers/MoopReport.pdf (accessed: October, 2008).

Czajka, J. L., and G. Denmead (2008): "Income Data for Policy Analysis: a Comparative Assessment of Eight Surveys," Working paper, Mathematica Policy Research, Inc.

D’Orazio, M., M. Di Zio, and M. Scanu (2006): Statistical Matching: theory and practice. Wiley, West Sussex PO19 8SQ, England.

Lange, K. (1999): Numerical Analysis for Statisticians. Springer-Verlag, New York, NY.
Machlin, S., W. Yu, and M. Zodet (2005): "Computing Standard Errors for MEPS Estimates," Report - http://www.meps.ahrq.gov/survey_comp/standard_errors.jsp (access: March, 2009), Agency for Healthcare Research and Quality.

RÄsSLER, S. (2002): Statistical Matching: A Frequentist Theory, Practical Applications and Alternative Bayesian Approaches, Lecture Notes in Statistics. Springer-Verlag, New York, NY.

Rodgers, W. L. (1984): "An Evaluation of Statistical Matching," Journal of Business and Economic Statistics, 2, 91-102.
U.S. Census Bureau (2001): "Annual Demographic Survey (March CPS Supplement) Description," http://www.nchs.census.gov/cps/ads/adsdes.htm (accessed: May, 2009).
(2006): "Introduction to SIPP," http://www.census.gov/sipp/intro.html (accessed: May, 2009).
(2008a):"Current Population Survey (CPS) - Definitions and Explanations," http://www.census.gov/population/www/cps/cpsdef.html (accessed: April, 2009).
(2008b): "Technical Documentation, Current Population Survey, 2005 Annual Social and Economic (ASEC) Supplement," http://www.census.gov/apsd/techdoc/cps/cpsmar05.pdf (accessed: April, 2009).
_ (2009): "Source and Accuracy Statement for SIPP 2004 Wave 1 - Wave 12 (core) Public Use Files," Report -
http://www.census.gov/sipp/sourceac/s\&a04_w1tow12(sa-9).pdf (accessed: March, 2009).


Table A-1: Estimates of CPS, SIPP and MEPS families (in thousands) by the original indicator variables

|  | Insurance coverage | Family size | Self-reported health | CPS Families |  | SIPP Families |  | MEPS Families |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | Estimate | \% | Estimate | \% | Estimate | \% |
| Non Elderly | Private | Single | Good/excellent | 22,128 | 17.6 | 20,272 | 17.0 | 21,494 | 17.1 |
|  |  |  | Fair/poor | 1,763 | 1.4 | 1,855 | 1.6 | 2,072 | 1.6 |
|  |  | Two persons | Good/excellent | 15,975 | 12.7 | 16,602 | 13.9 | 15,976 | 12.7 |
|  |  |  | Fair/poor | 2,767 | 2.2 | 3,715 | 3.1 | 3,279 | 2.6 |
|  |  | $3+$ persons | Good/excellent | 27,805 | 22.1 | 27,524 | 23.0 | 29,255 | 23.3 |
|  |  |  | Fair/poor | 4,596 | 3.6 | 5,660 | 4.7 | 4,448 | 3.5 |
|  | Public | Single | Good/excellent | 1,812 | 1.4 | 1,553 | 1.3 | 2,693 | 2.1 |
|  |  |  | Fair/poor | 1,967 | 1.6 | 1,808 | 1.5 | 2,166 | 1.7 |
|  |  | Two persons | Good/excellent | 1,407 | 1.1 | 1,140 | 1.0 | 2,202 | 1.8 |
|  |  |  | Fair/poor | 1,089 | 0.9 | 1,057 | 0.9 | 1,165 | 0.9 |
|  |  | $3+$ persons | Good/excellent | 3,824 | 3.0 | 3,139 | 2.6 | 4,555 | 3.6 |
|  |  |  | Fair/poor | 1,860 | 1.5 | 1,679 | 1.4 | 1,774 | 1.4 |
| All | Uninsured | Single | Good/excellent | 8,498 | 6.7 | 6,543 | 5.5 | 7,317 | 5.8 |
|  |  |  | Fair/poor | 1,101 | 0.9 | 994 | 0.8 | 1,106 | 0.9 |
|  |  | Two persons | Good/excellent | 1,785 | 1.4 | 1,137 | 1.0 | 1,219 | 1.0 |
|  |  |  | Fair/poor | 456 | 0.4 | 332 | 0.3 | 416 | 0.3 |
|  |  | $3+$ persons | Good/excellent | 2,173 | 1.7 | 1,392 | 1.2 | 1,098 | 0.9 |
|  |  |  | Fair/poor | 407 | 0.3 | 359 | 0.3 | 287 | 0.2 |
| Elderly | Private | Single |  |  | 3.8 | 6,097 | 5.1 | 4,264 | 3.4 |
|  |  |  | Fair/poor | 1,856 | 1.5 | 2,546 | 2.1 | 1,293 | 1.0 |
|  |  | $2+$ persons | Good/excellent | 6,087 | 4.8 | 5,660 | 4.7 | 5,947 | 4.7 |
|  |  |  | Fair/poor | 3,663 | 2.9 | 4,310 | 3.6 | 2,967 | 2.4 |
|  | Public | Single | Good/excellent | 2,432 | 1.9 | 1,460 | 1.2 | 3,597 | 2.9 |
|  |  |  | Fair/poor | 2,221 | 1.8 | 1,343 | 1.1 | 1,518 | 1.2 |
|  |  | $2+$ persons | Good/excellent | 1,691 | 1.3 | 462 | 0.4 | 1,896 | 1.5 |
|  |  |  | Fair/poor | 1,796 | 1.4 | 801 | 0.7 | 1,778 | 1.4 |
|  |  | Total |  | 125,991 | 100.0 | 119,439 | 100.0 | 125,781 | 100.0 |

Table A-2: Estimates of CPS, SIPP and MEPS families (in thousands) by the revised indicator variables

|  | Insurance coverage | Family size | Self-reported health | Income at Percentile |  |  | Breakdown for |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | 25 th | 50th | 75th | SIPP | MEPS |
| Non Elderly | Private | Single | Good/excellent | 20,798 | 33,004 | 50,078 | Quartiles | Quartiles |
|  |  |  | Fair/poor | 14,015 | 25,050 | 39,850 | Quartiles | No breaks |
|  |  | Two persons | Good/excellent | 40,399 | 64,500 | 98,691 | Quartiles | Quartiles |
|  |  |  | Fair/poor | 27,089 | 46,000 | 73,000 | Quartiles | Median |
|  |  | $3+$ persons | Good/excellent | 48,022 | 74,081 | 109,002 | Quartiles | Quartiles |
|  |  |  | Fair/poor | 37,281 | 59,023 | 87,900 | Quartiles | Median |
|  | Public | Single | Good/excellent | 5,400 | 9,799 | 16,799 | Median | Median |
|  |  |  | Fair/poor | 6,799 | 8,899 | 13,200 | Median | Median |
|  |  | Two persons | Good/excellent | 7,000 | 15,221 | 27,687 | Median | Median |
|  |  |  | Fair/poor | 8,376 | 14,119 | 23,521 | Median | No breaks |
|  |  | $3+$ persons | Good/excellent | 10,980 | 21,000 | 36,000 | Median | Quartiles |
|  |  |  | Fair/poor | 10,100 | 19,768 | 32,364 | Median | Median |
| All | Uninsured | Single | Good/excellent | 6,000 | 15,000 | 25,010 | Quartiles | Quartiles |
|  |  |  | Fair/poor | 100 | 8,842 | 18,800 | Median | No breaks |
|  |  | Two persons | Good/excellent | 14,000 | 25,030 | 42,044 | Median | No breaks |
|  |  |  | Fair/poor | 11,180 | 21,568 | 34,800 | No breaks | No breaks |
|  |  | $3+$ persons | Good/excellent | 18,200 | 32,413 | 54,350 | Median | No breaks |
|  |  |  | Fair/poor | 15,030 | 27,400 | 45,000 | No breaks | No breaks |
| Elderly | Private | Single | Good/excellent | 13,431 | 20,400 | 34,664 | Quartiles | Median |
|  |  |  | Fair/poor | 10,997 | 14,803 | 22,869 | Quartiles | No breaks |
|  |  | $2+$ persons | Good/excellent | 30,258 | 47,341 | 76,463 | Quartiles | Quartiles |
|  |  |  | Fair/poor | 24,000 | 35,643 | 57,176 | Quartiles | Median |
|  | Public | Single | Good/excellent | 9,199 | 12,799 | 19,506 | Median | Median |
|  |  |  | Fair/poor | 7,932 | 10,555 | 14,816 | Median | No breaks |
|  |  | $2+$ persons | Good/excellent | 18,398 | 27,162 | 46,957 | No breaks | No breaks |
|  |  |  | Fair/poor | 15,199 | 21,309 | 32,040 | No breaks | Median |

Table A-3: Quartile family income cutoffs - March, 2005 CPS

|  | Insurance Coverage | Income as a \% of poverty | Family size | Age of ref. person | Mean incom Est. | $\begin{aligned} & \text { CPS } \\ & \quad \pm 90 \% \end{aligned}$ | $\begin{aligned} & \text { Mean i } \\ & \text { Est. } \end{aligned}$ | $\begin{gathered} \text { ome SIPP } \\ \pm 90 \% \end{gathered}$ |  | Mean i Est. | $\begin{aligned} & \text { Dme MEPS } \\ & \pm 90 \% \end{aligned}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Non Elderly | Private | < $150 \%$ | Single | age $\leq 25$ | 6,453.31 | 196.96 | 6,675.65 | 454.82 | $a$ $a$ | 7,859.49 | 721.65 | b $a$ |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 7,954.46 | 372.5 | 7,620.85 | 846.76 | $a$ | 7,368.97 | 2911.09 | $a$ |
|  |  |  |  | ( $35<\mathrm{age} \leq 45$ ) | 6,586.06 | 511.26 | 7,664.69 | 843.2 | $b$ | 9,715.27 | 1434.67 | $b$ |
|  |  |  |  | ( $45<\mathrm{age} \leq 55$ ) | 7,308.07 | 357.15 | 7,234.19 | 704.16 | ${ }^{a}$ | 8,289.03 | 1020.83 | $a$ |
|  |  |  |  | ( $55<$ age $<65$ ) | 8,203.36 | 335.17 | 7,863.53 | 591.48 | $a$ $a$ | 7,985.11 | 1457.23 | $\stackrel{a}{a}$ |
|  |  |  | 2-3 | age $\leq 25$ | 11,537.08 | 556.52 | 12,297.27 | 876.23 | $\stackrel{a}{a}$ | 9,024.08 | 1297.77 | ${ }_{\text {c }}$ |
|  |  |  |  | (25<age $\leq 35)$ | 13,318.90 | 481.18 | 13,515.86 | 750.15 | a | 14,235.55 | 1494.38 | $a$ |
|  |  |  |  | ( $35<$ age $\leq 45$ ) | 12,785.31 | 516.39 | 13,963.88 | 674.74 | ${ }^{\text {b }}$ | 14,595.64 | 1150.49 | ${ }^{\text {b }}$ |
|  |  |  |  | ( $45<$ age $\leq 55$ ) | 11,784.63 | 545.44 | 13,762.00 | 614.68 | $b$ | 13,742.31 | 1651.59 | $b$ |
|  |  |  |  | ( $55<$ age $<65$ ) | 10,451.42 | 508.01 | 12,684.44 | 598.77 | ${ }^{\text {b }}$ | 10,232.10 | 1774.15 | $a$ |
|  |  |  | $4+$ | age $\leq 25$ | 20,298.47 | 1379.07 | 20,679.02 | 1548.87 | $a$ | 19,533.31 | 4611.72 | $a$ |
|  |  |  |  | (25<age $\leq 35)$ | 21,909.44 | 580.98 | 22,731.76 | 670.31 | $a$ | 22,582.80 | 1336.85 | ${ }^{a}$ |
|  |  |  |  | ( $35<\mathrm{age} \leq 45$ ) | 22,170.52 | 614.36 | 23,308.77 | 634.35 | ${ }^{\text {b }}$ | 22,844.44 | 1723.9 | ${ }^{a}$ |
|  |  |  |  | (45 < age< 65) | 22,250.89 | 751.34 | 24,323.55 | 789.29 | ${ }^{\text {b }}$ | 25,142.72 | 1830.64 | ${ }^{\text {b }}$ |
|  |  | $\geq 150 \%$ | Single | age $\leq 25$ | 31,766.18 | 707 | 29,627.83 | 1319.24 | ${ }^{c}$ | 31,520.30 | 2514.30 | $a$ |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 45,960.16 | 919.04 | 43,538.44 | 1334.57 | $c$ $c$ | 44,145.70 | 2846.46 | ${ }_{a}^{a}$ |
|  |  |  |  | ( $35<\mathrm{age} \leq 45$ ) | 53,064.66 | 1538.35 | 49,518.60 | 2626.35 | ${ }^{c}$ | 51,521.29 | 3262.80 | ${ }^{a}$ |
|  |  |  |  | ( $45<\mathrm{age} \leq 55$ ) | 53,595.80 | 1572.91 | 56,021.88 | 7021.18 | $a$ $c$ | 51,369.29 | 2840.14 |  |
|  |  |  |  | ( $55<$ age $<65$ ) | 55,924.72 | 1688.17 | 45,953.37 | 2592.23 | $c$ $c$ | 51,004.26 | 4121.24 |  |
|  |  |  | 2-3 | age $\leq 25$ | 52,164.14 | 2055.4 | 38,757.91 | 2975.57 | $c$ $c$ $c$ | 45,620.42 | 3174.36 | c $a$ |
|  |  |  |  | $(25<\mathrm{age} \leq 35)$ $(35<\mathrm{age} \leq 45)$ | $73,762.33$ $82,254.31$ | 1422.17 1831.33 | $52,534.19$ $58,946.33$ | 1572.49 1746.32 | $c$ $c$ c | 73,994.63 $76,231.46$ | 3845.37 3530.09 | ${ }_{c}^{a}$ |
|  |  |  |  | ( $45<$ age $\leq 55$ ) | 91,686.41 | 1468.39 | 61,875.52 | 1757.9 | c | 80,753.06 | 3290.67 | c |
|  |  |  |  | (55 < age<65) | 91,947.81 | 1764.67 | 63,360.32 | 2842.51 | ${ }^{c}$ | 82,558.02 | 4225.77 | ${ }^{c}$ |
|  |  |  | $4+$ | age $\leq 25$ | 75,841.47 | 10265.13 | 49,252.27 | 3995.55 | c | 59,097.55 | 6592.34 | c |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 72,934.61 | 1321.38 | 61,965.82 | 4404.12 | c | 75,501.54 | 3934.99 | ${ }^{\text {b }}$ |
|  |  |  |  | ( $35<\mathrm{age} \leq 45$ ) | 101,864.15 | 2004.23 | 79,875.70 | 3711.56 | $c$ $c$ | 95,141.47 | 3771.33 | ${ }_{\text {a }}$ |
|  |  |  |  | ( $45<$ age $<65$ ) | 115,603.28 | 2177.1 | 81,985.40 | 4110.82 | c | 101,737.16 | 4877.90 | c |
|  | Public | NA | $\begin{aligned} & \text { Single } \\ & 2-3 \end{aligned}$ | age $<65$ | 12,621.08 | 408.09 | 7,803.56 | 567.23 | ${ }^{c}$ | 11,111.07 | 1176.24 | c |
|  |  |  |  | age $\leq 35$ | 19,239.10 | 1086.83 | 12,734.49 | 905.55 | ${ }^{c}$ | 15,535.62 | 1727.70 | c |
|  |  |  |  | ( $35<$ age $\leq 45$ ) | 24,571.39 | 2595.44 | 12,925.68 | 1293.49 | c | 20,357.73 | 2457.01 | c |
|  |  |  |  | (45<age $\leq 55$ ) | 26,224.80 | 1735.07 | 15,520.05 | 3536.15 | c | 32,997.50 | 6347.65 | $b$ |
|  |  |  |  | ( $55<\mathrm{age}<65$ ) | 32,032.76 | 2189.25 | 15,046.27 | 1739.2 | ${ }^{c}$ | 26,892.28 | 3724.51 | ${ }_{c}^{\text {c }}$ |
|  |  |  | $4+$ | age $\leq 45$ | 27,526.22 | 848.89 | 17,913.92 | 954.96 | ${ }^{c}$ | 25,467.82 | 1727.81 | ${ }^{\text {c }}$ |
|  |  |  |  | (45<age $<65$ ) | 38,599.43 | 2219.12 | 19,179.06 | 1994.11 | c | 39,324.86 | 6739.09 | $a$ |
|  | Uninsured | < $150 \%$ | Single | age $<65$ | 5,630.19 | 129.62 | 7,053.85 | 260.54 | $b$ | 6,015.65 | 530.68 | $a$ |
|  |  |  | 2-3 | age < 65 | 9,579.05 | 897.76 | 12,412.86 | 978.83 | $b$ | 11,387.10 | 1810.08 | ${ }^{\text {b }}$ |
|  |  |  | $4+$ | age $<65$ | 17,553.11 | 353.28 | 19,275.40 | 574.48 | ${ }^{\text {b }}$ | 17,573.13 | 1292.70 | $a$ |
|  |  | $\geq 150 \%$ | Single | age $<65$ | 31,170.40 | 1926.58 | 27,329.22 | 1749.35 | ${ }^{c}$ | 31,049.70 | 5801.54 | ${ }_{a}$ |
|  |  |  | 2-3 | age $<65$ | 50,830.21 | 640.57 | 34,644.32 | 1036.55 | ${ }^{c}$ | 52,740.78 | 2649.85 | ${ }^{a}$ |
|  |  |  | $4+$ | age < 65 | 72,342.55 | 5076.95 | 51,519.27 | 5737.6 | c | 80,850.32 | 17508.32 | $a$ |
| Elderly | All | < $150 \%$ | Single | ( $65 \leq$ age < 75) | 8,695.90 | 135.87 | 8,846.47 | 289.04 | $a$ | 9,225.23 | 474.00 | $b$ |
|  |  |  |  | age $\geq 75 \mathrm{y}$ old | 9,336.83 | 1540.72 | 9,573.02 | 2385.67 | b $a$ $a$ | 8,767.18 | 3045.93 | ${ }_{\text {c }}$ |
|  |  |  | $2+$ | ( $65 \leq$ age $<75$ ) | 12,658.88 | 389.37 | 12,908.76 | 498.74 | ${ }^{a}$ | 12,147.22 | 1412.75 | ${ }^{a}$ |
|  |  |  |  | age $\sum 75 \mathrm{y}$ old | 12,839.33 | 1866.24 | 12,426.38 | 1656.57 | ${ }^{a}$ | 12,573.58 | 3771.81 | $a$ |
|  |  | $\geq 150 \%$ | Single | $(65 \leq$ age $<75$ ) | 36,442.27 | 101.17 | 32,473.48 | 185.62 | c | 35,298.92 | 446.65 | ${ }_{a}^{a}$ |
|  |  |  |  | age $\geq 75$ y old | 28,768.71 | 964.77 | 25,448.10 | 1351.2 | ${ }^{c}$ | 29,382.46 | 1416.27 | $a$ |
|  |  |  | $2+$ | $(65 \leq$ age $<75$ ) | 66,196.62 | 351.25 | 44,636.38 | 469.23 | ${ }^{c}$ | 59,873.22 | 1201.33 | ${ }^{c}$ |
|  |  |  |  | age $\geq 75$ y old | 46,913.56 | 996.46 | 41,379.09 | 2720.26 | ${ }^{\text {c }}$ | 49,225.42 | 2867.92 | $a$ |
|  |  |  |  | Total | 54,615.28 | 341.59 | 41,171.27 | 557.32 | c | 51,736.67 | 1200.73 | ${ }^{\text {c }}$ |

[^5]$c^{-}$- Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$host)

Table A-4: Mean annual income estimates - CPS, SIPP and MEPS by original indicator variables.

$b$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS < host)
${ }^{c}$ - Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$ host)

Table A-5: Mean annual income estimates - CPS and SIPP by revised indicator variables.


Table A-6: Mean annual income estimates - CPS and MEPS by revised indicator variables.


Table A-7: Synthetic MOOP and Net income means and $90 \%$ confidence values by family categories - Betson's indicator variables, 1996-1997 CEX host data (aged) and 2005 CPS recipient data.

|  | Insurance Coverage | Income as a \% of poverty | Family size | Age of reference person | Synthetic Est. | $\begin{array}{r} \text { MOOP CPS } \\ \pm 90 \% \end{array}$ | $\begin{aligned} & \text { MOOP } \\ & \text { Est. } \end{aligned}$ | $\begin{aligned} & \text { SIPP } \\ & \quad \pm 90 \% \end{aligned}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Non Elderly | Private | < $150 \%$ | Single | age $\leq 25$ | 766.74 | 59.41 | 626.59 | 177.49 | $a$ |
|  |  |  |  | $(25<$ age $\leq 35)$ | 990.72 | 59.64 | 912.53 | 250.91 | $a$ |
|  |  |  |  | $(35<$ age $\leq 45)$ | 1,038.56 | 47.60 | 1,078.78 | 213.06 | $a$ |
|  |  |  |  | (45<age $\leq 55)$ | 2,389.64 | 271.11 | 2,170.13 | 765.61 | $a$ |
|  |  |  |  | (55<age<65) | 1,977.29 | 80.34 | 3,093.38 | 1,484.67 | $a$ |
|  |  |  | 2-3 | age $\leq 25$ | 1,963.32 | 326.42 | 3,766.55 | 3,988.54 | $a$ |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 2,820.19 | 110.69 | 1,146.85 | 238.79 | c |
|  |  |  |  | (35<age $\leq 45$ ) | 1,635.19 | 76.79 | 1,649.19 | 378.18 | $a$ |
|  |  |  |  | $(45<$ age $\leq 55)$ | 2,580.04 | 127.87 | 2,809.08 | 1,133.69 | $a$ |
|  |  |  |  | (55<age<65) | 3,932.77 | 125.85 | 4,247.65 | 489.73 | $a$ |
|  |  |  | $4+$ | age $\leq 25$ | 954.47 | 100.78 | 1,136.21 | 364.35 | $a$ |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 1,924.31 | 74.50 | 1,916.83 | 217.12 | $a$ |
|  |  |  |  | (35<age $\leq 45$ ) | 2,677.74 | 131.31 | 2,674.17 | 338.53 | $a$ |
|  |  |  |  | (45<age<65) | 2,411.11 | 77.41 | 2,709.51 | 342.57 | $a$ |
|  |  | $\geq 150 \%$ | Single | age $\leq 25$ | 822.56 | 13.26 | 802.23 | 80.66 | $a$ |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 1,067.12 | 12.23 | 1,106.69 | 67.26 | $a$ |
|  |  |  |  | (35<age $\leq 45$ ) | 1,548.66 | 23.56 | 1,582.49 | 140.84 | $a$ |
|  |  |  |  | $(45<$ age $\leq 55)$ | 1,710.27 | 30.86 | 1,793.04 | 125.03 | $a$ |
|  |  |  |  | (55< age $<65$ ) | 1,862.60 | 31.49 | 2,046.20 | 162.91 | $b$ |
|  |  |  | 2-3 | age $\leq 25$ | 1,848.36 | 54.83 | 1,924.58 | 209.83 | $a$ |
|  |  |  |  | $(25<$ age $\leq 35)$ | 2,170.55 | 18.21 | 2,285.38 | 129.22 | $a$ |
|  |  |  |  | (35<age $\leq 45)$ | 2,750.77 | 37.70 | 2,949.89 | 170.89 | $b$ |
|  |  |  |  | $(45<$ age $\leq 55)$ | 3,266.44 | 37.42 | 3,567.99 | 242.02 | $b$ |
|  |  |  |  | $(55<$ age $<65$ ) | 3,780.47 | 44.32 | 4,105.46 | 272.58 | $b$ |
|  |  |  | $4+$ | age $\leq 25$ | 3,933.55 | 275.54 | 2,474.30 | 778.83 | c |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 3,328.43 | 38.38 | 3,445.97 | 222.68 | $a$ |
|  |  |  |  | $(35<$ age $\leq 45)$ | 3,612.82 | 15.46 | 4,473.45 | 713.75 | $b$ |
|  |  |  |  | $(45<$ age $<65$ ) | 4,389.27 | 21.85 | 4,795.31 | 347.42 | $b$ |
|  | Public | NA | $\begin{aligned} & \text { Single } \\ & 2-3 \end{aligned}$ | age $<65$ | 511.52 | 18.30 | 483.33 | 144.09 | $a$ |
|  |  |  |  | age $\leq 35$ | 427.54 | 27.32 | $361.31$ | $171.14$ | $a$ |
|  |  |  |  | $(35<\text { age } \leq 45)$ | 693.83 | 28.06 | 1,379.24 | $1,812.51$ | $a$ |
|  |  |  |  | $(45<$ age $\leq 55)$ | 1,034.66 | 77.45 | 955.28 | 375.21 | $a$ |
|  |  |  |  |  | 1,207.70 | 64.02 | 1,207.92 | $394.54$ | $a$ |
|  |  |  | $4+$ | $\text { age } \leq 45$ | 545.32 | 15.51 | 553.80 | 110.39 | $\stackrel{a}{a}$ |
|  |  |  |  | $(45<$ age $<65$ ) | 769.32 | 56.88 | 980.48 | 522.05 | $a$ |
|  | Uninsured | < $150 \%$ | Single | age $<65$ | 365.14 | 17.12 | 401.34 | 94.05 | $a$ |
|  |  |  | 2-3 | age $<65$ | 813.72 | 48.16 | 856.27 | $253.65$ | $a$ |
|  |  |  | $4+$ | age $<65$ | 1,276.96 | 75.48 | 1,285.45 | 340.42 | $a$ |
|  |  | $\geq 150 \%$ | Single | age $<65$ | 558.04 | 13.78 | $587.96$ | $151.26$ | $a$ |
|  |  |  | 2-3 | age $<65$ | 2,063.24 | 81.15 | 1,868.28 | $782.36$ | $a$ |
|  |  |  | $4+$ | age $<65$ | 1,874.02 | 61.80 | 2,081.61 | 556.34 | $a$ |
| Elderly | All | < $150 \%$ | Single | $(65 \leq$ age $<75$ ) | 1,201.92 | 37.37 | 1,223.57 | 218.28 | $a$ |
|  |  |  |  | $\text { age } \geq 75 \text { y old }$ | 1,499.79 | 28.00 | 1,481.01 | 175.90 | $a$ |
|  |  |  | $2+$ | $(65 \leq$ age $<75$ ) | 3,169.96 | 114.52 | 3,249.12 | 629.60 | $a$ |
|  |  |  |  | age $\geq 75$ y old | 3,275.79 | 86.10 | 3,740.83 | 691.65 | $a$ |
|  |  | $\geq 150 \%$ | Single | $(65 \leq$ age $<75$ ) | 1,918.11 | $32.53$ | $2,212.57$ | $520.91$ | $a$ |
|  |  |  |  | age $\geq 75$ y old | 1,992.82 | 38.68 | 2,354.19 | $392.58$ | $a$ |
|  |  |  | $2+$ | $(65 \leq \text { age }<75)$ | 3,720.12 | $44.21$ | $4,012.85$ | $368.53$ | ${ }_{a}^{a}$ |
|  |  |  |  | age $\geq 75$ y old | 3,776.14 | 57.99 | 4,357.60 | 587.76 | $a$ |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | Total | 2,266.64 | 9.40 | 2,498.82 | 69.60 | $b$ |

Table A-8: Comparison of Synthetic and Survey-based MOOP mean values by original indicator variables, Host survey SIPP, Betson model


Table A-9: Comparison of Synthetic and Survey-based MOOP mean values by original indicator variables, Host survey MEPS, Betson model

$b$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS < host)
$c$ - Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$ host)

Table A-10: Comparison of Synthetic and Survey-based MOOP mean values by revised indicator variables, Host survey SIPP, Betson model

|  | Insurance coverage | Family size | Self-reported health | CPS income cutoffs | Synthetic MOOP Est. | $\begin{gathered} \text { CPS } \\ \pm 90 \% \end{gathered}$ | MOOP Est. | $\begin{aligned} & \text { MEPS } \\ & \quad \pm 90 \% \end{aligned}$ | $b$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Non elderly | Private | Single | Good/excellent | Q1 | 1,408.12 | 20.71 | 1,254.40 | 186.84 | c |
|  |  |  |  | Q2 | 1,468.28 | 22.78 | 1,298.49 | 138.28 | c |
|  |  |  |  | Q3 | 1,737.89 | 20.98 | 1,507.06 | 168.73 | $c$ |
|  |  |  |  | Q4 | 1,875.66 | 17.62 | 1,905.97 | 170.42 | $a$ |
|  |  |  | Fair/poor | N | 2,339.06 | 31.93 | 2,317.89 | 311.18 | $a$ |
|  |  | 2 persons | Good/excellent | Q1 | 2,423.28 | 33.40 | 2,475.81 | 278.50 | $a$ |
|  |  |  |  | Q2 | 3,243.39 | 37.13 | 3,384.71 | 507.53 | $a$ |
|  |  |  |  | Q3 | 3,593.00 | 38.71 | 3,456.60 | 341.75 | $a$ |
|  |  |  |  | Q4 | 3,839.07 | 64.28 | 4,040.85 | 493.57 | $a$ |
|  |  |  | Fair/poor | M1 | 3,474.31 | 57.05 | 3,680.08 | 409.80 | $a$ |
|  |  |  |  | M1 | 4,416.66 | 38.13 | 4,769.62 | 555.85 | $b$ |
|  |  | $3+$ persons | Good/excellent | Q1 | 2,551.26 | 19.54 | 2,493.91 | 200.45 | $a$ |
|  |  |  |  | Q2 | 3,540.40 | 28.99 | 3,676.44 | 238.94 | $a$ |
|  |  |  |  | Q3 | 3,890.45 | 18.41 | 3,990.31 | 219.09 | $a$ |
|  |  |  |  | Q4 | 4,581.45 | 32.40 | 4,782.44 | 293.35 | $b$ |
|  |  |  | Fair/poor | M1 | 3,351.40 | 51.49 | 3,516.87 | 356.58 | $a$ |
|  |  |  |  | M2 | 5,298.86 | 67.32 | $5,597.91$ | 747.67 | $a$ |
|  | Public | Single | Good/excellent | M1 | 1,695.88 | 102.98 | 728.58 | 462.36 | c |
|  |  |  |  | M2 | 2,505.18 | 80.43 | 1,291.04 | $650.09$ | c |
|  |  |  | Fair/poor | M1 | 1,592.42 | 60.35 | 765.93 | 201.73 | $c$ |
|  |  |  |  | M2 | 3,472.15 | 165.08 | 1,798.28 | 740.50 | $c$ |
|  |  | 2 persons | Good/excellent | M1 | 1,326.62 | 72.05 | 453.95 | 221.84 | c |
|  |  |  |  | M2 | 1,193.10 | 54.89 | 863.46 | 306.44 | c |
|  |  |  | Fair/poor | N | 3,667.92 | 163.29 | 2,439.88 | 1,039.35 | $c$ |
|  |  | $3+$ persons | Good/excellent | Q1 | 536.34 | 12.07 | 231.60 | 70.12 | $c$ |
|  |  |  |  | Q2 | 825.18 | 27.07 | 407.02 | 89.88 | $c$ |
|  |  |  |  | Q3 | 777.59 | 31.67 | 671.62 | 170.34 | $c$ |
|  |  |  |  | Q4 | 1,525.65 | 83.09 | 883.91 | 303.62 | $c$ |
|  |  |  | Fair/poor | M1 | 2,326.61 | 110.96 | 1,106.94 | 373.21 | c |
|  |  |  |  | M2 | 1,675.45 | 78.99 | 1,681.85 | 495.42 | $a$ |
| All | Uninsured | Single | Good/excellent | Q1 | 480.88 | 16.12 | 448.39 | 159.61 | $a$ |
|  |  |  |  | Q2 | 290.06 | 10.23 | 391.35 | 177.12 | $a$ |
|  |  |  |  | Q3 | 235.91 | 8.27 | 429.97 | 170.11 | $b$ |
|  |  |  |  | Q4 | 335.26 | 8.40 | 576.25 | 219.91 | $b$ |
|  |  |  |  | N | 999.77 | 35.70 | 1,145.68 | 287.97 | $a$ |
|  |  | 2 persons | Good/excellent | N | 1,167.22 | 34.55 | 946.05 | 306.26 | c |
|  |  |  | Fair/poor | N | 1,573.28 | 66.53 | 1,626.10 | 552.53 | $a$ |
|  |  | $3+$ persons | Good/excellent | N | 1,173.88 | 28.61 | 1,250.67 | 657.77 | $a$ |
|  |  |  | Fair/poor | N | 1,763.89 | 95.61 | 1,533.73 | 470.37 | $a$ |
| Elderly | Private | Single | Good/excellent | M1 | 3,042.22 | 43.38 | 2,940.05 | 380.52 | $a$ |
|  |  |  |  | M2 | 2,840.63 | 59.12 | 2,940.73 | 515.20 | $a$ |
|  |  |  | Fair/poor | N | 3,849.06 | 47.74 | 4,091.16 | 951.74 | ${ }^{a}$ |
|  |  | $2+$ persons | Good/excellent | Q1 | 4,050.31 | 78.78 | 4,411.66 | 956.64 | $a$ |
|  |  |  |  | Q2 | 4,183.30 | 48.11 | 4,570.68 | 564.41 | $b$ |
|  |  |  |  | Q3 | 3,952.67 | 61.71 | 5,200.13 | 759.92 | $b$ |
|  |  |  |  | Q4 | 961.32 | 101.01 | 5,705.40 | 947.35 | $b$ |
|  |  |  | Fair/poor | M1 | $5,354.52$ | $31.95$ | $5,216.43$ | $720.63$ | $a$ |
|  |  |  |  | M2 | 6,322.25 | 69.30 | 6,667.03 | 795.01 | $a$ |
|  | Public | Single | Good/excellent |  |  | 57.90 |  |  | c |
|  |  |  |  | M2 | 1,569.49 | 38.42 | 1,616.92 | 266.04 | $a$ |
|  |  |  |  | $\mathrm{N}$ | 1,456.22 | 29.31 | 1,569.99 | 229.38 | $a$ |
|  |  | $2+$ persons | Good/excellent | N | 2,456.13 | 47.16 | 2,433.16 | 297.25 | $a$ |
|  |  |  | Fair/poor | M1 | 2,454.73 | 37.08 | 2,666.70 | 456.81 | $a$ |
|  |  |  |  | M2 | 3,615.02 | 129.02 | 3,924.75 | 921.65 | $a$ |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  | Total |  | 2,666.75 | 6.68 | 2,683.11 | 67.95 | $a$ |
| ${ }^{a}$ - Differenc | between CP | and host sur | 's mean estimat | is equal to ze |  |  |  |  |  |
| $b$ - Differenc <br> ${ }^{c}$ - Differenc | between CP <br> between CP | and host sur and host sur | y's mean estimat y's mean estimat | is less to zero is greater to | (CPS < host) <br> ro. (CPS > host) |  |  |  |  |

Table A-11: Comparison of Synthetic and Survey-based MOOP mean values by revised indicator variables, Host survey MEPS, Betson model

|  | Insurance Coverage | Income as a \% of poverty | Family size | Age of reference person | Synthetic Net income Est. | $\begin{gathered} \text { CPS } \\ \pm 90 \% \end{gathered}$ | Net income Est. | $\begin{aligned} & \text { SIPP } \\ & \pm 90 \% \end{aligned}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Non Elderly | Private | < $150 \%$ | Single | age $\leq 25$ | 5,957.40 | 29.14 | 6,302.35 | 440.40 | $a$ |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 7,174.58 | 57.17 | 6,963.46 | 830.79 | $a$ |
|  |  |  |  | $(35<$ age $\leq 45)$ | 5,876.17 | 44.57 | 6,817.52 | 793.80 | $b$ |
|  |  |  |  | (45<age $\leq 55)$ | 5,994.19 | 94.18 | 6,061.88 | 687.63 | $a$ |
|  |  |  |  | (55<age<65) | 6,616.13 | 68.91 | 6,336.28 | 563.52 | $a$ |
|  |  |  | 2-3 | age $\leq 25$ | 10,476.59 | 124.89 | 11,201.54 | 841.26 | $a$ |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 10,901.91 | 80.31 | 12,427.52 | 735.50 | $b$ |
|  |  |  |  | ( $35<$ age $\leq 45$ ) | 11,358.64 | 73.64 | 12,439.22 | 669.80 | $b$ |
|  |  |  |  | $(45<$ age $\leq 55)$ | 9,800.62 | 98.27 | 11,717.25 | 645.11 | $b$ |
|  |  |  |  | (55<age<65) | 7,459.09 | 109.60 | 9,184.90 | 623.20 | $b$ |
|  |  |  | $4+$ | age $\leq 25$ | 19,374.04 | 94.23 | 19,546.59 | 1,521.14 | $a$ |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 20,044.14 | 67.33 | 20,894.82 | 642.46 | $b$ |
|  |  |  |  | $(35<$ age $\leq 45)$ | 19,670.96 | 118.76 | 20,743.35 | 652.08 | $b$ |
|  |  |  |  | $(45<$ age $<65$ ) | 19,934.07 | 70.42 | 21,694.46 | 773.19 | $b$ |
|  |  | $\geq 150 \%$ | Single | age $\leq 25$ | 30,943.62 | 13.26 | 28,825.60 | 1,322.76 | c |
|  |  |  |  | $(25<$ age $\leq 35)$ | 44,893.04 | 12.23 | 42,431.75 | 1,331.76 | c |
|  |  |  |  | ( $35<$ age $\leq 45$ ) | 51,516.00 | 23.56 | 47,949.80 | 2,615.59 | c $a$ |
|  |  |  |  | $(45<$ age $\leq 55)$ | 51,885.53 | 30.86 | 54,231.11 | 7,014.10 | a |
|  |  |  |  | $(55<$ age $<65$ ) | 54,062.12 | 31.49 | 43,919.97 | 2,582.46 | c |
|  |  |  | 2-3 | age $\leq 25$ | 50,315.78 | 54.83 | 36,833.33 | 2,947.65 |  |
|  |  |  |  | $(25<$ age $\leq 35)$ | 71,591.79 | 18.21 | 50,258.16 | 1,564.23 | c |
|  |  |  |  | ( $35<$ age $\leq 45$ ) | 79,503.55 | 37.70 | 56,021.01 | $1,729.04$ | c |
|  |  |  |  | $(45<$ age $\leq 55)$ | 88,419.97 | 37.42 | 58,404.62 | 1,753.31 | $c$ $c$ |
|  |  |  |  | $(55<$ age $<65$ ) | 88,167.92 | 44.14 | 59,380.40 | 2,829.44 | c |
|  |  |  | $4+$ | $\text { age } \leq 25$ | 71,907.92 | 275.54 | 46,777.97 | 4,019.93 | c |
|  |  |  |  | $(25<$ age $\leq 35)$ | 69,606.18 | 38.38 | 58,519.85 | 4,401.18 | c |
|  |  |  |  | $(35<$ age $\leq 45)$ | 98,251.33 | 15.46 | $75,926.12$ | 3,698.90 | $c$ |
|  |  |  |  | $(45<$ age $<65$ ) | 111,214.01 | 21.85 | 77,298.08 | 4,101.03 | $c$ |
|  | Public | NA | $\begin{aligned} & \text { Single } \\ & 2-3 \end{aligned}$ | age $<65$ | 12,201.23 | 12.77 | 7,447.89 |  | c |
|  |  |  |  | $\text { age } \leq 35$ | 18,856.53 | 23.23 | 12,423.33 | 896.68 | c |
|  |  |  |  | $(35<$ age $\leq 45)$ | 23,957.94 | 30.44 | 12,353.57 | $1,267.41$ | c |
|  |  |  |  | $(45<$ age $\leq 55)$ | 25,336.31 | 60.59 | 14,656.31 | 3,525.45 | c $c$ |
|  |  |  |  | $(55<\text { age }<65)$ | 30,910.99 | 70.88 | 13,915.72 | 1,673.70 | c |
|  |  |  | $4+$ | $\text { age } \leq 45$ | 27,015.75 | 14.29 | 17,374.21 | 926.39 | c |
|  |  |  |  | $(45<$ age $<65$ ) | 37,846.08 | 57.13 | 18,369.42 | 1,945.50 | c |
|  | Uninsured | < 150\% |  |  |  |  |  |  | $b$ |
|  |  |  | $2-3$ | age $<65$ | $9,013.66$ | 28.80 | 11,667.52 | $595.00$ | $b$ |
|  |  |  | $4+$ | age $<65$ | 16,414.36 | 64.30 | 18,033.61 | 1,045.04 | $b$ |
|  |  | $\geq 150 \%$ | Single | age $<65$ | 30,612.36 | 13.78 | 26,781.74 | 971.06 | c |
|  |  |  | 2-3 | age $<65$ | 48,905.07 | 76.27 | 33,187.45 | 1,745.50 | c |
|  |  |  | $4+$ | age $<65$ | 70,468.52 | 61.80 | 49,437.66 | 5,731.94 | $c$ |
| Elderly | All | < $150 \%$ | Single | $(65 \leq$ age $<75$ ) | 7,605.68 | 33.82 | 7,737.29 | 285.00 | $b$ |
|  |  |  |  |  | $7,949.71$ | 25.17 | 8,195.08 |  | b |
|  |  |  | $2+$ | $(65 \leq$ age $<75$ ) | 10,222.46 | 98.94 | 10,261.54 | 530.75 | $a$ |
|  |  |  |  | age $\geq 75$ y old | 10,004.46 | 73.04 | 9,347.92 | 499.48 |  |
|  |  | $\geq 150 \%$ | Single | $(65 \leq$ age $<75$ ) | 34,524.20 | 32.54 | 30,433.85 | 2,385.14 | c |
|  |  |  |  | age $\geq 75$ y old | 26,776.29 | 38.72 | 23,337.25 | 1,352.88 | c |
|  |  |  | $2+$ | $(65 \leq$ age $<75$ ) | 62,477.74 | 44.10 | 40,795.12 | 1,661.87 | c |
|  |  |  |  | age $\geq 75$ y old | 43,140.00 | 57.66 | 37,430.39 | 2,669.02 | c |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  | Total |  | 52,404.89 | 10.01 | 38,848.68 | 552.01 | c |

Table A-12: Comparison of Synthetic and Survey-based Net Income mean values by original indicator variables, Host survey SIPP, Betson model


Table A-13: Comparison of Synthetic and Survey-based Net Income mean values by original indicator variables, Host survey MEPS, Betson model

|  | Insurance coverage | Family size | Self-reported health | CPS income cutoffs | Synthetic N Est. | $\begin{aligned} & \text { e CPS } \\ & \pm 90 \% \end{aligned}$ | Net income Est. | $\begin{aligned} & \hline \text { SIPP } \\ & \pm 90 \% \\ & \hline \end{aligned}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Non elderly | Private | Single | Good/excellent | Q1 | 10,402.38 | 13.03 | 11,039.45 | 306.44 | $b$ |
|  |  |  |  | Q2 | 25,924.77 | 14.87 | 25,747.96 | 192.00 | $a$ |
|  |  |  |  | Q3 | 39,623.10 | 16.46 | 39,557.05 | 260.16 | $a$ |
|  |  |  |  | Q4 | 86,970.64 | 22.64 | 83,829.27 | 6,238.75 | ${ }_{a}$ |
|  |  |  | Fair/poor | Q1 | 5,492.38 | 88.05 | 5,171.12 | 667.33 | $a$ |
|  |  |  |  | Q2 | 17,141.40 | 89.34 | 16,839.90 | 660.64 | $a$ |
|  |  |  |  | Q3 | 28,837.27 | 119.95 | 28,874.62 | 962.27 | $a$ |
|  |  |  |  | Q4 | 63,163.72 | 114.19 | 57,585.35 | 4,530.11 | c |
|  |  | 2 persons | Good/excellent | Q1 | 23,214.38 | 26.38 | 24,186.24 | 387.00 | ${ }^{\text {b }}$ |
|  |  |  |  | Q2 | 49,607.25 | 29.51 | 48,437.73 | 369.86 | c |
|  |  |  |  | Q3 | 76,899.42 | 28.55 | 75,352.35 | 662.79 | ${ }^{c}$ |
|  |  |  |  | Q4 | 160,991.42 | 41.08 | 155,457.56 | 11,824.89 | $a$ |
|  |  |  | Fair/poor | Q1 | 13,838.13 | 82.37 | 14,063.12 | 652.59 | $b$ |
|  |  |  |  | Q2 | 31,630.41 | 144.92 | 31,700.52 | 705.11 | $a$ |
|  |  |  |  | Q3 | 54,240.61 | 92.31 | 52,607.81 | 1,054.04 | ${ }^{c}$ |
|  |  |  |  | Q4 | 111,881.89 | 157.73 | 101,871.54 | 7,322.94 | ${ }^{c}$ |
|  |  | $3+$ persons | Good/excellent | Q1 | 28,995.09 | 24.59 | 29,046.04 | 305.58 | $b$ |
|  |  |  |  | Q2 | 57,410.94 | 22.91 | 55,824.48 | 324.22 | ${ }^{c}$ |
|  |  |  |  | Q3 | 86,284.86 | 30.57 | 84,194.99 | 593.86 | ${ }^{\text {c }}$ |
|  |  |  |  | Q4 | 179,287.88 | 54.58 | 187,632.19 |  | $b$ |
|  |  |  | Fair/poor | Q1 | 21,185.34 | 112.13 | $20,693.86$ | 581.56 | a |
|  |  |  |  | Q2 | 43,907.13 | 83.27 | 41,765.55 | 613.29 | ${ }^{\text {c }}$ |
|  |  |  |  | Q3 | 67,434.11 | 132.92 | 65,655.72 | 1,056.17 | ${ }^{\text {c }}$ |
|  |  |  |  | Q4 | 138,897.46 | 103.26 | 129,461.43 | 10,236.77 | $a$ |
|  | Public | Single | Good/excellent | M1 | 4,491.19 | 13.43 | 3,481.90 | 347.75 | c |
|  |  |  |  | M2 | 23,030.15 | 35.13 | 17,853.99 | 2,451.03 | c |
|  |  |  | Fair/poor | M1 | 5,261.70 | 19.41 | 3,064.00 | 308.87 | c |
|  |  |  |  | M2 | 15,909.69 | 72.38 | 13,674.32 | 1,297.16 |  |
|  |  | 2 persons | Good/excellent | M1 | 6,666.88 | 35.86 | 6,962.58 | 561.97 | b |
|  |  |  |  | M2 | 37,696.83 | 121.42 | 26,446.56 | 5,422.31 |  |
|  |  |  | Fair/poor | M1 | 7,567.64 | 41.57 | 5,638.12 | 532.52 | c |
|  |  |  |  | M2 | 29,992.91 | 111.94 | 21,297.81 | 2,662.85 |  |
|  |  | $3+$ persons | Good/excellent | M1 | 10,164.05 | 12.08 | 10,653.96 | 450.72 | ${ }^{\text {b }}$ |
|  |  |  |  | M2 | 45,979.37 | 23.21 | 31,839.35 | 1,431.03 | $c$ |
|  |  |  | Fair/poor | M1 | 9,605.03 | 23.70 | 8,406.55 | 577.97 | ${ }_{c}^{\text {c }}$ |
|  |  |  |  | M2 | 40,476.47 | 112.72 | 30,529.81 | 2,827.91 | c |
| All | Uninsured | Single | Good/excellent | Q1 | 1,328.92 | 7.40 | 1,741.63 | 190.64 | b $c$ |
|  |  |  |  | Q2 | 10,857.07 | 16.68 | 10,568.30 | 199.97 | ${ }^{c}$ |
|  |  |  |  | Q3 | 19,702.03 | 13.06 | 19,011.88 | 224.82 | ${ }^{\text {c }}$ |
|  |  |  |  | Q4 | 44,228.01 | 16.77 | 40,545.20 | 2,132.01 | c |
|  |  |  | Fair/poor | M1 | 2,155.10 | 22.94 | 2,805.49 | 434.11 | $b$ |
|  |  |  |  | M2 | 20,783.75 | 113.14 | 18,655.41 | 1,932.99 | c |
|  |  | 2 persons | Good/excellent | M1 | 12,233.51 | 37.36 | 13,216.52 | 853.58 | ${ }^{\text {b }}$ |
|  |  |  |  | M2 | 56,756.13 | 119.25 | 39,159.99 | 3,378.11 | c |
|  |  |  | Fair/poor Good/excellent | N | 27,683.11 | 68.19 | 19,326.29 | 2,494.76 | ${ }^{\text {c }}$ |
|  |  | $3+$ persons |  | M1 | 16,410.20 | 44.48 | 18,300.16 | 840.09 | ${ }^{\text {b }}$ |
|  |  |  |  | M2 | 70,459.84 | 47.66 | 47,597.71 | 3,676.42 | ${ }^{c}$ |
|  |  |  | Fair/poor | N | 34,644.94 | 94.51 | 24,940.19 | 4,947.82 | c |
| Elderly | Private | Single | Good/excellent |  |  | 45.13 | 8,371.50 |  | $b$ |
|  |  |  |  | $\begin{aligned} & \text { Q2 } \\ & \text { Q3 } \end{aligned}$ | $14,788.10$ | 34.41 | $14,586.12$ | $242.92$ | $a$ |
|  |  |  |  | Q3 | 24,207.71 | 75.01 | $24,136.62$ | $350.08$ | $a$ |
|  |  |  |  | Q4 | $63,325.30$ | 62.00 | $59,263.65$ | $6,464.48$ | $a$ |
|  |  |  | Fair/poor | Q1 | 5,888.30 | 63.08 | 6,489.00 | 398.01 | $b$ |
|  |  |  |  | Q2 | 10,765.27 | 81.11 | 10,678.61 | 301.03 | $a$ |
|  |  |  |  | Q3 | 15,950.16 | 183.86 | 15,799.30 | 444.07 | ${ }_{a}$ |
|  |  |  |  | Q4 | 35,796.99 | 151.06 | 34,623.49 | 7,113.11 | a |
|  |  | $2+$ persons | Good/excellent | Q1 | 17,506.25 | 57.77 | 15,632.98 | -469.95 | c a |
|  |  |  |  | Q2 | 34,975.80 | 72.16 | 34,552.42 | $581.52$ | $a$ |
|  |  |  |  | Q3 | 57,317.91 | 70.90 | 54,551.49 | 965.31 | ${ }^{\text {c }}$ |
|  |  |  |  | Q4 | 144,895.10 | 82.23 | 109,955.40 | 7,906.92 | ${ }^{c}$ |
|  |  |  | Fair/poor | Q1 | 13,492.31 | 89.55 | 11,743.70 | 431.14 | ${ }_{\text {c }}$ |
|  |  |  |  | Q2 | 26,222.01 | 80.19 | 26,161.15 | 560.31 | a |
|  |  |  |  | Q3 | 41,218.85 | 125.21 | 39,779.85 | 812.22 | ${ }^{c}$ |
|  |  |  |  | Q4 | 96,419.46 | 94.15 | 92,024.31 | 14,543.82 | ${ }^{a}$ |
|  | Public | Single | Good/excellent | M1 |  | 18.22 | 7,757.13 | 367.32 | ${ }^{a}$ |
|  |  |  |  | M2 | 26,802.60 | 27.81 | $23,508.44$ | 3,590.12 | ${ }^{\text {a }}$ |
|  |  |  | Fair/poor | M1 | 7,047.80 | 30.49 | $5,979.29$ | $346.75$ | c a |
|  |  |  |  | M2 | 16,999.29 | 63.15 | $16,051.22$ | $2,045.20$ | $a$ $c$ |
|  |  | $2+$ persons | Good/excellent Fair/poor | $\stackrel{N}{N}$ | $37,773.87$ $26,353.15$ | $23.23$ | $18,412.04$ | $2,685.21$ | c $c$ |
|  |  |  |  | N | 26,353.15 | 76.56 | 12,784.31 | 1,269.59 | ${ }^{\text {c }}$ |
|  |  |  |  | Total | 52,451.64 | 0.53 | 38,848.68 | 552.01 | ${ }^{c}$ |

$b$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS < host)
${ }^{c}$ - Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$ host)

Table A-14: Comparison of Synthetic and Survey-based Net Income mean values by revised indicator variables, Host survey SIPP, Betson model

|  | Insurance coverage | Family size | Self-reported health | CPS income cutoffs | Synthetic N Est. | $\begin{aligned} & \text { e CPS } \\ & \pm 90 \% \end{aligned}$ | Net income Est. | MEPS $\pm 90 \%$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Non elderly | Private | Single | Good/excellent | Q1 | 10,107.33 | 15.67 | 10,813.00 | 524.06 | $b$ |
|  |  |  |  | Q2 | 25,628.53 | 22.78 | 25,952.00 | 333.42 | $b$ |
|  |  |  |  | Q3 | 39,151.83 | 20.98 | 39,443.00 | 473.05 | $b$ |
|  |  |  |  | Q4 | 86,763.63 | 17.62 | 76,364.00 | 2,760.53 | $c$ |
|  |  |  | Fair/poor | N | 28,710.71 | 30.44 | 27,044.00 | 2,789.90 | $a$ |
|  |  | 2 persons | Good/excellent | Q1 | 23,081.59 | 36.49 | 24,014.00 | 941.34 | $b$ |
|  |  |  |  | Q2 | 48,877.97 | 37.13 | 48,545.00 | 806.69 | $a$ |
|  |  |  |  | Q3 | 76,149.60 | 38.71 | 76,589.00 | 872.85 | $a$ |
|  |  |  |  | Q4 | 160,862.34 | 64.28 | 138,811.00 | 4,610.25 | $c$ |
|  |  |  | Fair/poor | M1 | 22,933.01 | 63.01 | 25,090.00 | 1,563.37 | $b$ |
|  |  |  |  | M1 | 82,967.76 | 38.13 | 82,257.00 | 7,293.65 | $a$ |
|  |  | $3+$ persons | Good/excellent | Q1 | 29,018.72 | 21.49 | 30,202.00 | 712.34 | $b$ |
|  |  |  |  | Q2 | 57,159.80 | 28.99 | 56,779.00 | 532.21 | c |
|  |  |  |  | Q3 | 86,039.26 | 18.41 | 85,800.00 | 755.11 | $a$ |
|  |  |  |  | Q4 | 179,252.00 | 32.40 | 158,249.00 | 4,574.58 | $c$ |
|  |  |  | Fair/poor | M1 | 32,682.12 | 53.27 | 32,176.00 | $1,603.10$ | $a$ |
|  |  |  |  | M2 | 103,486.75 | 67.32 | 88,878.00 | 4,607.98 | c |
|  | Public | Single | Good/excellent | M1 | 3,839.91 | 36.11 | 4,107.81 | 529.75 | $a$ |
|  |  |  |  | M2 | 21,568.09 | 66.25 | 21,421.00 | 2,829.77 | $a$ |
|  |  |  | Fair/poor | M1 | 4,471.58 | 28.79 | 4,845.62 | 471.20 | $b$ |
|  |  |  |  | M2 | 14,476.35 | 92.81 | 14,104.00 | 1,660.72 | a |
|  |  | 2 persons | Good/excellent | M1 | 5,993.29 | 56.51 | 6,872.24 | 630.53 | $b$ |
|  |  |  |  | M2 | 37,533.92 | 54.89 | 34,222.00 | 3,422.02 | c |
|  |  |  | Fair/poor | N | 17,454.86 | 80.01 | 22,327.00 | 4,910.43 | $b$ |
|  |  | $3+$ persons | Good/excellent | Q1 | 4,605.98 | 14.43 | 5,346.05 | 416.91 | $b$ |
|  |  |  |  | Q2 | 15,065.22 | 27.07 | 15,742.00 | 312.02 | $b$ |
|  |  |  |  | Q3 | 27,051.54 | 31.67 | 27,321.00 | 560.52 | $a$ |
|  |  |  |  | Q4 | 64,129.03 | 83.09 | 57,996.00 | 4,886.26 | $c$ |
|  |  |  | Fair/poor | M1 | 8,385.45 | 71.90 | 9,276.39 | 601.19 | $b$ |
|  |  |  |  | M2 | 40,507.05 | 78.99 | 36,362.00 | 2,863.30 | $c$ |
| All | Uninsured | Single | Good/excellent | Q1 | 1,258.17 | 9.60 | 2,457.64 | 377.61 | $b$ |
|  |  |  |  | Q2 | 10,847.99 | 10.23 | 10,529.00 | 455.08 | c |
|  |  |  |  | Q3 | 19,820.61 | 8.27 | 19,274.00 | 443.41 | c |
|  |  |  |  | Q4 | 44,600.96 | 8.40 | 43,602.00 | 2,864.92 | $a$ |
|  |  |  |  | N | 11,286.13 | 23.99 | 11,885.00 | 3,157.53 | $a$ |
|  |  | 2 persons | Good/excellent | $\mathrm{N}$ | 34,302.35 | 30.51 | 41,475.00 | $6,127.03$ | $b$ |
|  |  |  | Fair/poor | $\mathrm{N}$ | $27,760.21$ | $76.26$ | $25,060.00$ | $7,714.18$ | $a$ |
|  |  | $3+$ persons | Good/excellent | $\mathrm{N}$ | 43,633.87 | $24.88$ | $55,859.00$ | $11,446.03$ | $b$ |
|  |  |  | Fair/poor | N | 34,370.86 | 78.06 | 38,146.00 | 5,505.59 | $b$ |
| Elderly | Private | Single | Good/excellent | M1 |  |  |  | 611.90 | c |
|  |  |  |  | M2 | $42,824.10$ | 58.95 | 38,887.00 | 3,011.39 | c |
|  |  |  |  | N | 15,887.39 | 35.01 | 16,881.00 | 1,839.10 | $a$ |
|  |  | $2+$ persons | Good/excellent | Q1 | 17,079.02 | 76.20 | 15,667.00 | 1,311.18 | ${ }^{\text {c }}$ |
|  |  |  |  | Q2 | 34,185.25 | 48.11 | 33,594.00 | 954.64 | $c$ |
|  |  |  |  | Q3 | 56,872.65 | 61.71 | 54,690.00 | 1,545.45 | $c$ |
|  |  |  |  | Q4 | 148,915.29 | 101.01 | 113,805.00 | 7,084.21 | c |
|  |  |  | Fair/poor | M1 | 18,155.15 | 29.05 | 17,760.00 | 1,305.01 | $a$ |
|  |  |  |  | M2 | 66,768.15 | 69.30 | 63,107.00 | 4,966.51 | $c$ |
|  | Public | Single | Good/excellent | M1 | 6,989.70 | 45.54 | 6,860.24 | 632.18 | $a$ |
|  |  |  |  | M2 | 26,023.70 | 38.42 | 30,112.00 | 2,755.59 | $b$ |
|  |  |  | Fair/poor | N | 11,431.06 | 27.75 | 12,352.00 | 1,526.57 | a |
|  |  | $2+$ persons | Good/excellent | N | 36,430.67 | 46.51 | 37,076.00 | 4,276.53 | $a$ |
|  |  |  | Fair/poor | M1 | $11,709.13$ | $29.48$ | $10,647.00$ | $1,018.57$ | ${ }^{\text {c }}$ |
|  |  |  |  | M2 | 38,152.14 | 129.02 | 34,521.00 | 3,164.41 | $c$ |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  | Total |  | 52,034.04 | 6.41 | 49,129.65 | 1,171.37 | $c$ |

$a$ - Difference between CPS and host survey's mean estimates is equal to zero.
$b$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS < host)
${ }^{c}$ - Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$ host)

Table A-15: Comparison of Synthetic and Survey-based Net Income mean values by revised indicator variables, Host survey MEPS, Betson model

|  | Insurance <br> Coverage | Income as a \% of poverty | Family size | Age of reference person | Synthetic Est. | $\begin{array}{r} \text { MOOP CPS } \\ \pm 90 \% \end{array}$ | $\begin{gathered} \text { MOOP } \\ \text { Est. } \end{gathered}$ | $\begin{aligned} & \text { SIPP } \\ & \quad \pm 90 \% \end{aligned}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Non Elderly | Private | < $150 \%$ | Single | age $\leq 25$ | 268.66 | 47.68 | 626.59 | 177.49 | $b$ |
|  |  |  |  | (25< age $\leq 35$ ) | 1,524.00 | 247.84 | 912.53 | 250.91 | c |
|  |  |  |  | ( $35<\mathrm{age} \leq 45$ ) | 1,001.57 | 205.84 | 1,078.78 | 213.06 | $a$ |
|  |  |  |  | (45<age $\leq 55$ ) | 1,715.62 | 327.40 | 2,170.13 | 765.61 | $a$ |
|  |  |  |  | (55<age<65) | 2,136.17 | 592.16 | 3,093.38 | 1,484.67 |  |
|  |  |  | 2-3 | age $\leq 25$ | 1,400.88 | 460.96 | 3,766.55 | 3,988.54 | $a$ |
|  |  |  |  | (25< age $\leq 35$ ) | -860.21 | 166.24 | 1,146.85 | 238.79 | $a$ |
|  |  |  |  | (35<age $\leq 45$ ) | 1,581.48 | 323.89 | 1,649.19 | 378.18 | $a$ |
|  |  |  |  | (45<age $\leq 55$ ) | 1,911.81 | 313.18 | 2,809.08 | 1,133.69 | $a$ |
|  |  |  |  | ( $55<$ age $<65$ ) | 3,843.53 | 459.88 | 4,247.65 | 489.73 | ${ }^{a}$ |
|  |  |  | $4+$ | age $\leq 25$ | 1,534.79 | 371.91 | 1,136.21 | 364.35 | $a$ |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 1,895.95 | 261.93 | 1,916.83 | 217.12 | $a$ |
|  |  |  |  | (35<age $\leq 45$ ) | 2,414.99 | 317.64 | 2,674.17 | 338.53 |  |
|  |  |  |  | (45<age<65) | 2,824.18 | 297.41 | 2,709.51 | 342.57 | ${ }_{a}^{a}$ |
|  |  | $\geq 150 \%$ | Single | age $\leq 25$ | 899.71 | 63.66 | 802.23 | 80.66 | $a$ |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 1,022.12 | 45.66 | 1,106.69 | 67.26 | ${ }^{\text {b }}$ |
|  |  |  |  | ( $35<$ age $\leq 45$ ) | 1,953.12 | 134.64 | 1,582.49 | 140.84 |  |
|  |  |  |  | (45<age $\leq 55$ ) | 2,222.17 | 121.40 | 1,793.04 | 125.03 |  |
|  |  |  |  | ( $55<$ age $<65$ ) | 1,983.25 | 97.87 | 2,046.20 | 162.91 |  |
|  |  |  | 2-3 | age $\leq 25$ | 2,166.45 | 176.70 | 1,924.58 | 209.83 |  |
|  |  |  |  | $(25<$ age $\leq 35)$ | 2,784.23 | 97.79 | 2,285.38 | 129.22 | $c$ $c$ |
|  |  |  |  | $(35<\mathrm{age} \leq 45)$ $(45<\mathrm{age} \leq 55)$ | $3,225.76$ $3,877.93$ | 119.91 129.08 | $2,949.89$ $3,567.99$ | 170.89 242.02 | $c$ $c$ |
|  |  |  |  | (55<age<65) | 5,096.58 | 339.18 | 4,105.46 | 272.58 |  |
|  |  |  | $4+$ | age $\leq 25$ | 2,320.02 | 400.96 | 2,474.30 | 778.83 |  |
|  |  |  |  | ( $25<$ age $\leq 35$ ) | 3,844.36 | 165.49 | 3,445.97 | 222.68 |  |
|  |  |  |  | ( $35<$ age $\leq 45$ ) | 4,895.21 | 337.39 | 4,473.45 | 713.75 | ${ }^{a}$ |
|  |  |  |  | (45<age<65) | 5,597.51 | 206.72 | 4,795.31 | 347.42 | c |
|  | Public | NA | $\begin{aligned} & \hline \text { Single } \\ & 2-3 \end{aligned}$ | age < 65 | 677.01 | 107.58 | 483.33 | 144.09 | c |
|  |  |  |  | age $\leq 35$ | 430.47 | 56.53 | 361.31 | 171.14 | $a$ |
|  |  |  |  | (35< age $\leq 45$ ) | 6,265.69 | 2,654.71 | 1,379.24 | 1,812.51 | ${ }^{c}$ |
|  |  |  |  | (45<age $\leq 55$ ) | 1,317.31 | 221.26 | 955.28 | 375.21 | ${ }_{a}^{a}$ |
|  |  |  |  | ( $55<$ age $<65$ ) | 1,678.59 | 308.55 | 1,207.92 | 394.54 | a |
|  |  |  | $4+$ | age $\leq 45$ | 709.53 | 74.71 | 553.80 | 110.39 | c |
|  |  |  |  | (45<age $<65$ ) | 2,877.78 | 511.68 | 980.48 | 522.05 | c |
|  | Uninsured | < 150\% | Single | age < 65 | 333.26 | 43.16 | 401.34 | 94.05 | $a$ |
|  |  |  | 2-3 | age $<65$ | 664.79 | 130.91 | 856.27 | 253.65 | $a$ |
|  |  |  | $4+$ | age $<65$ | 1,059.44 | 202.43 | 1,285.45 | 340.42 | ${ }_{a}$ |
|  |  | $\geq 150 \%$ | Single | age $<65$ | 572.27 | 50.10 | 587.96 | 151.26 | ${ }_{a}$ |
|  |  |  | 2-3 | age < 65 | 2,175.59 | 815.49 | 1,868.28 | 782.36 | ${ }_{a}^{a}$ |
|  |  |  | $4+$ | age < 65 | 1,564.55 | 183.97 | 2,081.61 | 556.34 | $a$ |
| Elderly | All | < $150 \%$ | Single | $(65 \leq$ age < 75) | 1,123.76 | 146.14 | 1,223.57 | 218.28 | $a$ |
|  |  |  |  | age $\geq 75$ y old | 1,638.91 | 207.80 | 1,481.01 | 175.90 | $a$ |
|  |  |  | $2+$ | ( $65 \leq$ age $<75$ ) | 2,696.94 | 557.65 | 3,249.12 | 629.60 | $a$ |
|  |  |  |  | age $\geq 75$ y old | 2,784.51 | 422.10 | 3,740.83 | 691.65 | $b$ |
|  |  | $\geq 150 \%$ | Single | $(65 \leqq$ age $<75$ ) | 2,034.22 | 389.96 | 2,212.57 | 520.91 | ${ }_{a}^{a}$ |
|  |  |  |  | age $\geq 75 \mathrm{y}$ old | 2,296.70 | 291.68 | 2,354.19 | 392.58 | ${ }_{a}^{a}$ |
|  |  |  | $2+$ | $(65 \leq$ age $<75$ ) | 3,976.16 | 156.32 | 4,012.85 | 368.53 | ${ }_{a}^{a}$ |
|  |  |  |  | age $\geq 75$ y old | 4,083.00 | 271.13 | 4,357.60 | 587.76 | $a$ |
|  |  |  | Total |  | 2,693.57 | 50.11 | 2,498.82 | 69.60 | c |

$b$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS < host)
$c$ - Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$ host)

Table A-16: Comparison of Synthetic and Survey-based MOOP mean values by original indicator variables, Host survey SIPP, Predictive Mean Matching


Table A-17: Comparison of Synthetic and Survey-based MOOP mean values by original indicator variables, Host survey MEPS, Predictive Mean Matching

$b$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS < host)
$c$ - Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$ host)

Table A-18: Comparison of Synthetic and Survey-based MOOP mean values by the revised indicator variables, Host survey SIPP, Predictive Mean Matching


Table A-19: Comparison of Synthetic and Survey-based MOOP mean values by the revised indicator variables, Host survey MEPS, Predictive Mean Matching

${ }^{a}$ - Difference between CPS and host survey's mean estimates is equal to zero.
${ }^{b}$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS $<$ host)
${ }^{c}$ - Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$ host)

Table A-20: Comparison of Synthetic and Survey-based Net income mean values by original indicator variables, Host survey SIPP, Predictive Mean Matching

$b$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS < host)
${ }^{c}$ - Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$ host)

Table A-21: Comparison of Synthetic and Survey-based Net income mean values by original indicator variables, Host survey MEPS, Predictive Mean Matching

${ }^{b}$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS < host)
${ }^{c}$ - Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$ host)

Table A-22: Comparison of Synthetic and Survey-based Net income mean values by revised indicator variables, Host survey SIPP, Predictive Mean Matching

$b$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS < host)
$c$ - Difference between CPS and host survey's mean estimates is greater to zero. (CPS $>$ host)

Table A-23: Comparison of Synthetic and Survey-based Net income mean values by revised indicator variables, Host survey MEPS, Predictive Mean Matching


[^0]:    *This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone more limited review than official publications. Any views expressed on statistical and methodical issues are those of the authors and not necessarily those of the U.S. Census Bureau. The authors wish to thank Robin Anderson, Rebecca Blank, Judy Hubbard Eargle, David Hornick, David Johnson, Tracy Mattingly, Arloc Sherman, Kathy Short, and Mahdi Sundukchi for their helpful comments, suggestions and access to needed supplemental files. The authors blame each other for all remaining mistakes.
    ${ }^{\dagger}$ (corresponding author) Dynamics of Economic Well Being Management Staff, Housing and Household Economic Statistics Division, US Census Bureau, Washington, DC 20236 (1.301.763.5132) (Sharon.I.O'Donnell@census.gov)
    ${ }^{\ddagger}$ Groupe Sup de Co, La Rochelle

[^1]:    ${ }^{1}$ Beginning 2013, the SIPP survey (known as Re-engineered SIPP or RESIPP) will collect health insurance data annually through an event history calendar.

[^2]:    ${ }^{2}$ We wish to acknowledge Arloc Sherman for raising this point.

[^3]:    ${ }^{3}$ When the predicted CDF values deviate from the actual CDF values, the predicted distribution has a longer right tail compared to the actual distribution.

[^4]:    ${ }^{4}$ For a thorough review of the differences in the income distributions of eight major U.S. national surveys, see Czajka and Denmead (2008).
    ${ }^{5}$ Personal communication, Jessica Banthin

[^5]:    ${ }^{b}$ - Difference between CPS and host survey's mean estimates is less to zero. (CPS $<$ host)

