Technical Documentation
2014 Production File
(Version 4.2)

March 2015
Revised December 2015

By
Quantria Strategies, LLC

Technical Documentation
2014 Production File
(Version 4.2)

Table of Contents

Overview[1]

Each year we create a production file comprised of individual tax units constructed from the three most recently released datasets from the March Current Population Survey (CPS). The year of the production file relates to the most recent CPS year. For the 2014 file, tax units are constructed from the March 2014, March 2013 and March 2012 CPS files.

Income amounts reported on the CPS are asked retrospectively and relate to the prior calendar year's income. As such, incomes on the 2014 Production File represent reported income for taxable years 2013, 2012 and 2011.

Tax units constructed from each year are combined into one master database and, at least initially, the sample weight for each return is divided by three. These preliminary weights are further adjusted to reflect control totals reported by the IRS for each state using an algorithm that minimizes the percentage change in the original sample weight subject to hitting the control totals.

This document describes each step in the data development process beginning with the processing of the original March CPS file.

Current Population Survey (CPS)

Also referred to as the Annual Social and Economic Supplement (ASEC)[2], the March Supplement to the monthly CPS is an annual survey of US households conducted in March of every year. In contains detailed information on the income, demographics, family structure and labor force participation of household members. The CPS sample represents the non-institutional population of the US at a given point in time.

Annual CPS data files have a hierarchical structure, with three distinct record types:

- <u>Household</u> records contain information on the housing unit that is sampled. This information includes such items as state of residence, ownership status (e.g., homeowner or renter), selected information on the physical characteristics of the housing units (e.g., number of rooms) and summary income information for the household.
- <u>Family</u> records contain information the families that reside in the housing unit. For CPS purposes, a family is defined a group of related individuals living together. More than one family can reside in a household.

---

[1] Our state individual income tax model has benefitted greatly from the input of many individuals, but especially Mark Robyn (Pew), Mike Allen (Maine Revenue Services), Mike Udell (District Economics), Mike Ettlinger (University of New Hampshire) and Jon Bakija (Williams College).
[2] In earlier years, the CPS was referred to as the Annual Demographic File (ADF).

- <u>Person</u> records contain information on all persons living in the housing unit. Income information is only collected on those members of the household who are 15 years of age or older.

In order to make CPS data resemble tax return data collected by the IRS, we rely on information from all three record types to create notional tax returns (i.e., "tax units) for each household in the sample. To facilitate this task, we combine information from all three record types into one, flat, extract file for processing. We use this file extract to construct tax units.

We also include information on income and demographic variables that do not appear on tax returns filed with the IRS. This can extend the types of analysis that can be supported with the Production File. For example, our tax unit file contains information on health status, health insurance coverage, pension coverage, wages received by both spouses for married couples, transfer payments (e.g., Temporary Assistance for Needy Families (TANF)) and educational attainment.

Creating Tax Units from the CPS

For most households in the CPS, constructing notional tax returns, or tax units, is a straightforward process. We make the initial assumption that <u>every</u> housing unit is comprised of at least one tax unit. Next, we assemble the income and demographic information from the household members in a manner that would emulate the filing of an individual income tax return.

For example, for a single person living alone we would construct a tax unit with a single filing status and no dependents. Income reported on the CPS is then used to construct the tax unit. The income items available on the CPS are quite comprehensive and include: wages, interest, dividends, business and self-employment income, rental income and unemployment insurance. Similarly, a household comprised of a married couple with two children would be assumed to file a joint return claiming two dependents. Demographic information relating to the ages of household members is also used and included as part of the tax unit characteristics.

In contrast to this relatively simple situation, additional complexities arise in certain cases. For example, if a child in the household is 15 years-of-age or older and reports receiving income of some type, then that child may legally be required to file a tax return.[3] Because we want our population of tax units to resemble as closely as possible the tax returns actually filed with the IRS for a particular year, we want our file of tax units to include these dependent filers.

Another common situation that complicates the construction of tax units from the CPS occurs when there is more than one family residing in a household. For these reasons, a more complicated algorithm is needed to construct tax units from the CPS extract.

---

[3] Income information is only collected on the CPS for individuals 15 years of age or older.

Briefly, the algorithm for creating tax units follows several steps:

1. Store the information for each household member into a temporary array to facilitate further processing.
2. Initialize each household by creating one tax unit. (We assume each household contains at least one tax unit.)
3. Populate the tax unit with information on the household head, whose person record is stored first in the sequence of person records.
4. Loop through the remaining person records in the household and determine if each person can be considered part of the tax unit (e.g., a spouse or child of the household head).
5. When household members are encountered who do not meet the tax filing requirements to be a member of the initial tax unit, a new tax unit is created for this person and a similar search through remaining household members is done to ascertain if they could be included as part of this tax unit.
6. Once all members of a household have been represented in at least one tax unit, all the tax units in the household are then output to a permanent file.

We rely on several modules to streamline the process of creating tax units:

CREATE: Creates a CPS Tax Unit and stores it in a temporary array.

SEARCH: The inner loop of the processing algorithm. Searches all household members except reference person & spouse to determine if they are eligible to be dependents.

ADDEPT: Adds a dependent to a Tax Unit. Arguments to this macro are the person number of the (new) dependent and the Tax Unit number to which it belongs.

IFDEPT: Determines dependency status. Relies on tax law in making this determination. For example, to qualify as a dependent, an individual must be 18 years of age or younger (or 22 years of age and a student).

FILST: Determines whether or not a CPS Tax Unit is deemed to actually file a return.

INCOME: Computes Gross Income (for determining filing requirements) for the Tax Unit.

HHSTATUS: Determines whether a single individual with dependents can or will file as a head of household return.

TOTINCX: Calculates the total income of all tax units in the household.

COMBINE: Combines another tax unit and all of its members into another tax unit, making the new members dependents.

CONVERT: Converts an existing tax unit into a dependent on another tax unit.

RELATION: Determines if any two household members are related in order to ascertain if one can be claimed as a dependent on another return.

MUSTFILE: Determines if a dependent is legally required to file a tax return.

Tax units created in this way are then tabulated along a number of dimensions, including filing status, aged status, dependency status and number of dependents and compared with IRS reported figures for that tax year.

In general, the number of tax units created from the CPS will differ from the number of tax returns filed in any tax year. There are several reasons for this:

- Respondents in the CPS generally under-report income from many sources including interest and dividends. In addition, the CPS does not ask respondents about capital gain income (or loss).
- IRS administrative data relating to returns filed, including tabulations produced by the Statistical of Income (SOI) Division will include returns filed for prior tax years.
- Residents living abroad and certain non-resident aliens may be required to file a US income tax return. These individuals will not be part of the CPS sample.
- The CPS sample only covers the non-institutional population of the US. Many institutionalized individuals (e.g., nursing homes) may also be required to file tax returns.[4]
- Because the CPS sample frame is based on housing units, it may be the case that certain individuals living in separate housing units may actually file tax returns together. This would happen, for example, if as part of a divorce decree one spouse is allowed to claim a child as a dependent even though that child does not live in the household.
- Individuals who become deceased during the year will often have a tax return filed on their behalf. Information on these individuals may or may not be included in the CPS.
- Income information on the CPS is only collected for individuals 15 years-of-age or older. To the extent children under 15 have income that would make them legally required to file an income tax return, they cannot be identified in the CPS.[5]
- In addition, many individuals and married couples file tax returns even though they aren't legally required to do so. Our approach creates tax units for these taxpayers, but we do not count them as filers.

---

[4] College students are generally included in the household of their parents for CPS reporting.
[5] In tax year1996, there were approximately 2 million tax filers under the age of 15. (Source: Independent tabulation from the SOI Complete Report File (CRF)).

Despite these limitations, the CPS over the years has proved to be an extremely useful tool for tax analysis as well as a source of supplemental information to support IRS administrative data records.

When creating tax units for each CPS year, we construct a variable, FILST, which indicates if a tax unit actually files a tax return for this year. The purpose of this variable is to better target IRS figures for the tax filing population. For those tax units legally required to file a return, we make this determination based on filing thresholds in place for each tax year. Table 1 shows the federal tax filing thresholds for tax years 2009, 2010 and 2011 (corresponding to CPS years 2010, 2011 and 2012).

Table 1 – Federal Filing Thresholds, by Tax Year

Filing Type	Filing Threshold - Tax Year		
	2009	2010	2011
Single, Under 65	$9,350	$9,500	$9,750
Single, 65 or Over	$10,750	$10,950	$11,200
Married Filing Jointly, Both Spouses Under 65	$18,700	$19,000	$19,500
Married Filing Jointly, One Spouse Over 65	$19,800	$20,150	$20,650
Married Filing Jointly, Both Spouses Over 65	$20,900	$21,300	$21,800
Married Filing Separately	$3,650	$3,700	$3,800
Head of Household, Under 65	$12,050	$12,200	$12,500
Head of Household, Over 65	$13,450	$13,650	$13,950
Qualifying Widower, Under 65	$15,050	$15,300	$15,700
Qualifying Widower, Over 65	$16,150	$16,450	$16,850
Single Dependents, Under 65, Unearned Income	$950	$950	$950
Single Dependents, Under 65 Earned Income	$5,700	$5,800	$5,950
Single Dependents, Over 65, Unearned Income	$2,350	$2,400	$2,400
Single Dependents, Over 65 Earned Income	$7,100	$7,250	$7,400
Married Dependents, Under 65, Unearned Income	$950	$950	$950
Married Dependents, Under 65 Earned Income	$5,700	$5,800	$5,950
Married Dependents, Over 65, Unearned Income	$2,050	$2,100	$2,100
Married Dependents, Over 65 Earned Income	$6,800	$6,950	$7,100

Source: Internal Revenue Service, 1040 Instructions, Filing Requirements for various years.
Note: Dependent filers have additional limitations based on whether a blind exemption is claimed.

Determining tax-filing status in this way will understate the true number of tax returns filed in any year because many taxpayers will file returns even though may not be legally required to do so. In order to approximate the number of tax filers in the population, we randomly choose tax units whose income falls below the federal filing threshold to match published IRS totals as closely as possible.[6] In determining which CPS tax units to choose, we examine the combined wage and salary income of both the primary taxpayer and spouse (if present). For taxpayers with positive wage and salary income, we select

---

[6] For an alternative method of identifying potential tax filers from the non-filer population, see Cilke, J. (1998), "A Profile of Non-Filers", Office of Tax Analysis Paper 78. Washington, DC: U.S. Department of the Treasury.

these taxpayers with a 90 percent probability. For all other taxpayers, we select them with a 60 percent probability.

Another area where the CPS differs from SOI published totals is in the number of taxpayers filing as a head of household. Historically, the SOI reports more head of household returns than are obtained from the CPS when relying on reported income and demographic information. In order to correct for this undercount, we randomly switch CPS tax units that we initially determine to be single taxpayers (with dependents) to file as head of households. We adjust the probability of selection to match IRS control totals of returns filed by filing status.

Top Coding

In order to protect the confidentiality of survey respondents in the CPS, information relating to certain income amounts are masked or "top-coded" in the CPS. Prior to 2011, a two-stage method was used to top-code selected income amounts. In the first stage, and depending on the income source, a respondent's income is flagged for top coding if it exceeds a pre-determined limit (the "censoring threshold). In the second stage, any reported amount that exceeds the censoring threshold is replaced with the <u>average</u> income for all respondents with reported income over the threshold (the "top coded amount").

For the March 2011 and March 2012 CPS, a different method was used to mask high-income amounts referred to as "rank proximity swapping". In this method, a censoring threshold is determined for each income type and each survey response that is equal to or greater than this amount is sorted from lowest to highest. In the second step, reported amounts are swapped between records that lie within certain bounds. In this way, the distribution of the top-coded amounts is preserved. Table 2 shows the censoring thresholds for each top-coded income item by CPS year.

Table 2 – Top-Coding Thresholds, by CPS Year

CPS Variable	Description	Top Coded Amount		
		2010*	2011	2012
ERN-VAL	Earnings Before Deductions	$200,000	$250,000	$250,000
WS-VAL	Wage and Salary Earnings, Other	$35,000	$47,000	$50,000
SE-VAL	Business Self-Employment	$50,000	$60,000	$60,000
FRM-VAL	Farm Self-Employment	$25,000	$30,000	$40,000
SUR-VAL1	Survivor's Income - Source 1	$50,000	$57,600	$75,000
SUR-VAL2	Survivor's Income - Source 2	$50,000	-	-
DIS-VAL1	Disability Income - Source 1	$35,000	$48,000	$44,000
DIS-VAL2	Disability Income - Source 2	$35,000	-	-
RET-VAL1	Retirement Income - Source 1	$45,000	$64,000	$67,000
RET-VAL2	Retirement Income - Source 2	$45,000	-	$67,000
INT-VAL	Interest Income	$25,000	$24,000	$22,000
DIV-VAL	Dividend Income	$15,000	$20,000	$20,000
RNT-VAL	Rents, Royalties, etc.	$40,000	$50,000	$60,000
ED-VAL	Educational Assistance	$20,000	$30,000	$25,000
CSP-VAL	Child Support Payments	$15,000	$21,000	$18,300
ALM-VAL	Alimony Income Received	$45,000	$66,000	$96,000
FIN-VAL	Financial Assistance	$30,000	$30,000	$36,000
OI-VAL	Other Income	$25,000	$30,000	$31,200

Source: Current Population Survey, Annual Social and Economic (ASEC) Supplement, various years.
*/ For 2010, these amounts refer to the censoring point, above which the average value for the reported amounts is given. In addition, the average value is calculated separately for different population sub-groups.

Prior to correcting for top coding, we make a rough adjustment to interest and dividends on the CPS to reflect the fact that these amounts are typically under-reported. Nevertheless, it is possible that some tax units might have large amounts of dividend income and small amounts of interest income. Such characteristics are not uncommon in administrative tax records and reflect a high degree of variability in income sources.

Top coding in this way means that aggregate control totals derived from the CPS are meaningful in that they represent an estimate of the population total. However, in order to make the data suitable for tax analysis, we remove the top coding for each record and create "replicate" tax units for each top coded record according to the following algorithm:

1. We search each of the income fields in the tax unit and identify those amounts that have been top coded and the tax unit is "flagged" for subsequent processing to remove the top coding.
2. We replace each top coded amount with a lognormal random variable whose mean is the top coded amount and whose variance is initially set to equal the standard error of the estimate (SEE) of the regression equation. We then multiply the SEE by a factor to more closely match published SOI totals of the income distribution at the upper end of the income scale.
3. We repeat this process N times for each top coded tax unit. In this version of the production file, we initially set N equal to 10. This means that each top coded record is represented 10 times in the database with different values for each of the top coded fields and the resulting sample weight divided by 10. (We continue to experiment with different values of N.)

We continue to implement this approach even though the top coding in CPS files subsequent to March 2010 rely on a different top-coding method. This is because, while the technique of "rank proximity swapping" retains the underlying distribution of top coded returns, it does not adequately capture the extreme values that are reported in the SOI.[7] Once the top-coding has been implemented, we re-calculate total income for each tax unit and substitute the top-coded amounts.

Imputation of Additional Fields

While the CPS contains a great deal of income, labor force and demographic information for US households, much tax-related information is missing. We augment our production file with additional, statistical imputations of additional fields.[8]

We rely on several external datasets to support the imputations. These include: the Statistics of Income Public Use File (PUF), the Federal Reserve Board's Survey of Consumer Finances (SCF), the Bureau of Labor Statistics Consumer Expenditure Survey (CES) and the Medical Expenditure Panel Survey (MEPS). Table 3 lists the variables used in the imputations. Imputation results are shown separately in Appendices B, C and D.

---

[7] To put this issue in perspective, the largest value of wage income reported in the CPS under the new method is approximately $1 million. In the public use version of the SOI, it is not uncommon for taxpayers to report wages in excess of $100 million.
[8] For these imputations, we construct our income measure to align with CPS concepts.

For most variables we rely on a two-step procedure. In the first step, we estimate a (weighted) logit model for the probability that a taxpayer reports a specific income or deduction amount. In the second step, we rely on a Weighted Least Squares (i.e., GLS) regression to estimate the dollar value.

In constructing the database we first calculate a probability that a particular taxpayer had a particular item from the logit equation and compare this estimated probability to a uniformly distributed random variable to determine if a value will be imputed. If the random variable is less than or equal to the predicted probability then we rely on the GLS equation to impute the dollar amounts. We add a zero mean, normally distributed random error term with a standard deviation equal to standard error of the estimate (SEE) to this predicted value to arrive at our imputed amount.

For items where microdata do not exist, we generally rely on a mean-based imputation, differentiated by state and income class obtained from SOI published tabulations.

Table 3 – Description of Variables Used in Imputations

Variable Name	Description
INCOME	Total Income (CPS Definition)
HASGAINS	Binary Variable Equal to 1 if Taxpayer Reports Capital Gains
CGAINS	Capital Gains Reported on 1040
MARS	Marital Status (1=Married Filing Joint, 0=Otherwise)
LNGAINS	Natural Logarithm of Capital Gains
LNINCOME	Natural Logarithm of Income
INTST	Interest Income
DBE	Dividend Income (Before Exclusion)
LNINTST	Natural Logarithm of Interest Income
LNDBE	Natural Logarithm of Dividend Income
DIVSHR	Share of Income Attributable to Dividend Income
INTSHR	Share of Income Attributable to Interest Income
MTXRT	Marginal Tax Rate
PENSIONS	Pension Income
LNPENSIONS	Natural Logarithm of Pension Income
FAMILYSIZE	Family Size
AGEDE	Number of Aged/Blind Exemptions

Capital gains and losses

We rely on the Public Use SOI to impute long- and short-term capital gains and losses to CPS tax units. We use a two-stage, regression-based approach where the probability of realizing a capital gain or loss is estimated in the first stage and the (conditional) amount of the gain or loss is estimated in the second stage.

Once total capital gains are imputed, we rely on tabulations from the 2004 PUF to allocate gains between short- and long-term gains by income class. Regression results are shown in Table B-1A and Table B-1B.

Self-Employed, SEP, SIMPLE and Qualified Plans

Individuals who are self-employed (i.e., sole proprietors) receive a deduction for contributions to qualified retirement plans. We rely on CPS variables to identify sole proprietors, partners or S-Corporation shareholders who are eligible to claim the deduction and rely on the PUF to estimate a two-step approach to estimate the amount of contributions. Regression results are shown in Tables B-2A and B-2B.

Self-employed health insurance

Individuals who are self-employed (i.e., sole proprietors) receive a deduction for premiums paid for health insurance. We rely on CPS variables to identify sole proprietors, partners or S-Corporation shareholders who are eligible to claim the deduction and rely on the PUF to estimate a two-step approach to estimate the amount of premiums paid. Regression results are shown in Tables B-3A and B-3B.

Self-employment taxes

Individuals are allowed a deduction for one-half of self-employed taxes (SECA) paid. We calculate this amount directly from the CPS reported values for self-employment income relying on the limitations in place for the tax year.

Individual Retirement Accounts (IRAs)

We impute contributions to individual retirement accounts based on publicly available information from the PUF. We rely on CPS variables to identify individuals eligible to make IRA contributions and estimate a two-step approach to determine the probability of making an IRA contribution and, conditional on making the contribution, the dollar value in the second step. Regression results are shown in Tables B-4A and B-4B.

Student Loan Interest Expense

We impute deductions for student loan interest expense based on publicly available information from the PUF. We estimate a two-step procedure to determine the probability of claiming a deduction for student loan interest and, conditional on claiming the deduction, the dollar value in the second step. Regression results are shown in Tables B-5A and B-5B.

Home mortgage interest expense

We rely on the Federal Reserve Board's Survey of Consumer Finances (SCF) to impute home value, outstanding mortgage amount and home equity. Mortgage interest is calculated as a fixed percent of the outstanding mortgage and calibrated to SOI totals. We identify homeowners by the CPS household variable TENURE. Regression results are shown in Tables C-1A and C-1B for home value and the ratio of home mortgage to home value.

State and local taxes

Taxpayers who itemize their deductions can deduct state and local income taxes paid on their federal tax return. We rely on our state tax calculator to obtain this amount for the state portion only. We ignore local taxes due to lack of available data. Real estate and property taxes are calculated as a state-specific percentage of home value.

Charitable Deductions and Miscellaneous Deductions

We rely on regression-based imputations from the Public Use SOI to impute charitable giving and Miscellaneous (limited and unlimited) Deductions to all tax units on the Production File. We estimate a Tobit model for those taxpayers who report a positive value for each item. Estimation results are reported in Tables B-6, B-7 and B-8.

Medical expenses

Taxpayers are allowed an itemized deduction for medical expenses to the extent they excel 7.5 percent of Adjusted Gross Income. We impute total medical expenses for all tax units using regression-based imputations based on expenses reported on the MEPS. Estimation results are shown in Table D-1.

Child credits

We calculate the child credit as the applicable amount (presently $1,000) per child times the number of qualified children in the tax unit.

Child and Dependent Care Credit

We use a two-stage procedure to impute total, qualified child and dependent care expenses. In the first stage, a logit model is estimated to calculate the probability that a taxpayer will report a qualified expense. In the second stage, we estimate an OLS regression for the amount of the expense. Results are shown in Tables B-9A and B-9B.

Earned income tax credit (EITC)

We rely on our federal tax calculator to compute the value of the EITC given the tax law parameters in effect. We use CPS variables to determine the number of qualified children for purposes of the EITC.

Other tax credits

We compute other, non-refundable tax credits as a residual amount based on publicly available tabulations from the IRS by state.

Rent Paid

Certain states allow a deduction or credit for rent paid. We estimate the amount of rent paid by each taxpayer on the file by using the TENURE variable to identify renters. Next, we rely on tabulations from the US Census Bureau, American Community Survey (ACS) on "gross rent as a percentage of income" by state to assign a rent about to each renter on the production file.

Pros and Cons in Using the CPS for Tax Policy Analysis

An underlying theme from the peer reviewers relates to our ability to match IRS/SOI income distributions for the US and for individual states. We are aware of this issue and have been for some time. While we have continued to focus on this aspect of database construction, make adjustments as necessary and continue to experiment with alternative methods, we thought it would be helpful to address this issue upfront because we consider this to be an important topic.

Relying on the CPS as our primary data source for income tax calculations presents a number of challenges, including: missing data on certain income items (e.g., capital gains), deductions (e.g., medical expenses) and the top-coding of most income fields. While the CPS will never be a substitute for actual tax return information, we believe that it can be a reliable "second best" alternative. (We elaborate on this below.)

While we have relied on methodologies to address the income distribution issue (e.g., Stage-I and Stage-II targeting, state-by-state calibration), we have been cautious in our approach in relying on these methods too much. The reason is that, in our experience, too much reliance on these "brute force" approaches – while guaranteeing that income distributions be matched – comes with a price. Certain unintended consequences are likely to arise by creating other, sometimes unrelated, data issues that may be difficult to identify, at least initially. Nevertheless, as we continue to experiment with alternative approaches, we are continuously updating our methodology and approach.

A second reason we have been reluctant to make extensive changes to our targeting of income distributions is that we have seen very little evidence that it would have much impact on our results. This is partly because the present tax policy simulations, the "blank slate" scenario, are relatively insensitive to the income distribution because very few states have differential tax rates at incomes above $200,000, our current threshold for targeting. We have done extensive testing to ensure that this is the case. Nevertheless, we are aware that other policy scenarios, more driven by the income distribution, will require more detailed targeting.

We chose the CPS as our primary data source for income and demographic information for several reasons. First, the CPS is a well-known and understood survey of the US population that has been collected for more than 50 years. Second, as a result of the increased need for state-level data to support the SCHIP program, sample sizes in the CPS were significantly increased in the 1990's to allow for better state-level analysis. Third, the CPS contains a wealth of demographic and labor force information that is not available on income tax returns (e.g., age, gender). Fourth, the CPS is an up-to-date and timely source of income data, with a time lag of approximately one year between the time the information is collected and released to the public. (For example, information on income received in 2014 will be available in the Fall of 2015. For administrative tax return information such as the SOI Public Use File (PUF), this lag is several years.) Fifth, the CPS sample frame allows for accurate, state-level analysis.

Such is not the case for the PUF where the sample frame is stratified across income classes. The PUF is not suitable for state-level analysis unless significant re-weighting is applied to the file.[9] Finally, the CPS contains income and demographic information on the entire US (non-institutional) population, including those households that are not legally required to file a federal income tax return. This means that certain tax proposals (e.g., lowering the filing threshold) cannot be simulated using administrative tax data unless they are augmented with information on the non-filing population. Most microsimulation tax models rely on the CPS to fill in these gaps.

American Community Survey (ACS) as an Alternative Data Source of Tax Units

In the field beginning in 2005, the ACS is an on-going, continuous survey of US households designed to replace the long form in the decennial Census. The ACS provides a great deal of detail on the income and demographics of households that is similar in most respects to the CPS.

As Mike points out, one great advantage of the ACS is the large sample size – over 1 million households in the 2011 ACS – and its state- and sub-state coverage. We have worked with the ACS in the past and seriously considered making it the source of our database in our initial planning discussions.

---

[9] The re-weighting performed by the Tax Policy Center (TPC) is a good example of how this might be accomplished.

Ultimately, we decided to rely on the CPS as our primary data file because the ACS contains more limited information on the type and sources of income. The CPS asks approximately fifty (50) questions on categories of income, benefits and transfer payments received by households while the ACS relies on eight (8) questions to capture the total amount of income received. We believe, at least at this time, that the CPS is better suited to address most tax policy issues. However, for other important state-level policy questions (e.g., poverty rates, health insurance coverage), the ACS should be considered a superior alternative to the CPS. Additionally, the CPS has the advantage in that it is the survey of record for certain official government statistics such as the unemployment rate and the number of individuals that have health insurance. As such, it's a useful reference point for policy analysis.

In addition to the CPS having more income detail than the ACS, there are at lest two other potential advantages of using the CPS. First, the CPS is more current than the ACS, which has a one-year lag in release relative to the CPS. Second, because the ACS is in the field continuously throughout the year, income reporting is with respect to the previous 12 months. This means that annual incomes are measured differently across ACS households.[10] However, be believe that these two distinctions are unlikely to have much effect on the usefulness of the ACS for tax policy analysis.[11]

A recent evaluation of the usefulness of the ACS for policy analysis sums up what is perhaps the current consensus of researchers on this topic:

> "Research is needed to understand the contributions of each of the above factors to differences between the CPS ASEC and the ACS. For users, now that the ACS is in full production with a vastly larger sample size than the CPS ASEC, it seems reasonable that they look to the ACS estimates for states and sub-state areas. However, users who want to analyze income by source and examine the correlates of income for population groups at the national level should stay with the CPS ASEC, which not only is the source of official income statistics, but also contains a wealth of variables to use in analysis."
>
> Source: P.127 "Using the American Community Survey: Benefits and Challenges", Panel on the Functionality and Usability of Data from the American Community Survey, C.F. Citro and G. Kalton, editors. National Academy Press, 2007.

---

[10] The ACS provides an inflation adjustment factor to make meaningful comparisons.

[11] Another potential advantage of the ACS relates to the sample frame and how the institutional population is treated. Because the ACS is a replacement for the decennial Census long form, those residents in nursing homes and on military bases are included. These households units are not part of the CPS sample frame.

Ultimately, our choice to rely on the CPS as our basic data source was based on its usefulness for tax policy analysis. This view may change in the future. We would add that state revenue departments that have access to confidential tax return information could greatly enhance the usefulness and accuracy of their tax policy models if these data were combined in some way (e.g., statistical matching) with the annual ACS.

ADDENDUM: Additional Imputations to Model "Blank Slate" Tax Reform

Under a scenario of a "blank slate" approach to comprehensive tax reform, all current law deductions, credits and preferences are repealed. In addition, certain types of income that are presently excluded from taxable income that are classified by the Joint Committee on Taxation (JCT) and the Treasury Department's Office of Tax Analysis (OTA) to be tax expenditures, are included in the calculation of total income.

We expand the number of imputed items to include the following amounts:

Inside Build-Up of Life Insurance Benefits

Certain types of life insurance contracts (e.g., cash value policies) are comprised of an amount that builds value over time and is paid out when the policyholder dies or terminates the contract. The annual amount of this "inside build-up" is considered a tax expenditure.

We estimate the value of this tax expenditure by relying on the two-step procedure described above and estimated from the SCF. In the first step, the probability of owning a cash-value policy is estimated. In the second step, the value of this policy is estimated. Estimation results are shown in Tables C-2a and C-2b.

Earnings on Defined Benefit (DB) Pension Plans

We use a similar procedure to estimate the earnings on DB pension arrangements. First, we identify taxpayers who are covered by an employer-provided pension from the CPS. Second, we calculate a probability that this is a DB plan from the SCF. As a further adjustment, we rely on CPS industry codes to assign certain taxpayers to DB plans (e.g., public employees). In the second stage, we estimate the amount of DB plan balances from the SCF and apply an average rate of return on these balances to arrive at an estimate of the earnings. Estimation results are shown in Tables C-3a and C-3b.

Earnings on Defined Contribution (DC) Pension Plans

Earnings on DC pension plans are estimated in a similar manner. We identify taxpayers who are covered by an employer-provided pension from the CPS. Second, if the taxpayer was not imputed to belong to a DB, we assume this taxpayer belongs to a DC plan. In the second stage, we estimate the amount of DB plan balances from the SCF and apply an average rate of return on these balances to arrive at an estimate of earnings. Estimation results are shown in Tables C-4a and C-4b.

Capital Gains on Home Sales

Taxpayers who sell their primary residence may elect to exclude a portion of capital gain on the sale from income. (Under current law, this exclusion is available for the first $500,000 for a married couple and $250,000 for single individuals.)

To estimate the value of this exclusion, we assign a probability that the taxpayer sells his or her home in a particular year. For this purpose, we rely on data collected by the National Association of Realtors (NAR) on the number of home sales by state. This state-specific probability is then used to identify homeowners who have sold their primary residence. We rely on our previous imputation of home equity and home value to represent the gain on the sale. For purposes of the tax expenditure estimate, the excluded amount is limited to $500,000/$250,000 for joint and single taxpayers, respectively.

Step-Up in Basis at Death

Decedents receive a step-up in the basis of appreciated property when they die and property is passed along to their heirs. We estimate the value of this step-up by first estimating the total net worth of each taxpayer from the SCF. Next, we assign an age-adjusted ratio of total net worth to cost basis obtained IRS estate tax returns. In the third step, we assign an (income adjusted) probability of death by gender obtained from the Social Security Administration's actuarial tables. Finally, we adjust this amount to match tax expenditure estimates from the JCT and OTA. Estimation results are shown in Table C-5.

Tax-Exempt Bond Interest

 CPS respondents are asked to report their total interest income and no distinction is made between interest from taxable and non-taxable sources. Because we target taxable interest in the model, we rely on SOI tabulations by state to calculate the fraction of those taxpayers that receive tax-exempt interest and rely on the ratio between taxable and non-taxable interest to determine the latter.

Employer-Provided Health Insurance Premiums

We rely on tabulations from the Medical Expenditure Panel Survey (MEPS, Employer Component) to calculate the average employer provided premium for single and family plans. We identify taxpayers who receive health insurance from their employer according to how they answered the relevant CPS questions relating to health insurance coverage. We assign a value for these employer-sponsored health premiums to both the primary taxpayer and spouse.

Employer Contributions to Pension and Retirement Plans

We rely on the Department of Labor's *National Compensation Survey* to estimate the annual contribution of employers to DB and DC pension plans. We identify workers covered by these plans from the earlier imputations on the earnings of each type of plan. Once each taxpayer is identified as being covered by a plan, we impute the amount of the employers' contribution as a fraction of wages as reported in the *National Compensation Survey*. We rely on separate tabulations for DB and DC plans and for Private and Government plans.

Education Expenses

Estimates of tax-deductible education expenses are derived from a two-step, regression-based model estimated from the PUF. Tax units on the CPS are first identified as being likely to incur educations expenses from questions in the survey (e.g., a child attending college) before an amount is assigned.

Estimates of Taxpayer Behavior

Because tax rates can affect behavior on the part of taxpayers, we rely on existing research to simulate taxpayer responses separately to changes in tax rates for capital gain realizations and charitable contributions. Both capital gains and charitable giving are discretionary decisions made by taxpayers and this research has shown that marginal tax rates (MTRs) have important, statistically significant effects on the decision to sell a capital asset or make a charitable donation.

We incorporate behavior by assuming that taxpayers respond to changes in MTRs by calculating the MTR under present law and again under some alternative proposal. The change in behavior is then calculated by applying a response parameter to the change. Different response parameters are used for capital gains and charitable giving and are culled from existing research.[12]

For charitable giving, we have presently set the response parameter to zero to reflect differing estimates in recent and on-going research. We will review this assumption in further versions of the model.

---

[12] See Auten, G., H. Sieg and C. T. Clotfelter (2002), "The Distribution of Charitable Giving, Income and Taxes: An Analysis of Panel Data", *American Economic Review*, 92(1), 371-382 or Bakija, J. and B. T. Heim (2011), "Does Charitable Giving Respond to Incentives and Income? New Estimates from Panel Data", *National Tax Journal*, 64(2, Part 2), 615-650 for a recent sample of this literature as it applies to charitable giving. For capital gains realizations, *New Evidence on the Tax Elasticity of Capital Gains: A Joint Working Paper of the Staff of Joint Committee on Taxation and the Congressional Budget Office*, JCX-56-12, June 2012, provides an up-to-date summary.

Alignment to IRS Control Totals

Each spring, the IRS publishes state-level estimates of the tax base for all fifty states and the District of Columbia. The data are tabulated by income class and include detailed information on the components of income, adjustments, exemptions, deductions, credits and federal tax liability. We utilize these tabulations to align the production file to match these control totals.

For the current 2012 production file (representing tax year 2011), we rely on state-by-state IRS tabulations for tax year 2011.[13] We specifically target, by state, the following income and deduction items:

- Wages
- Taxable Interest
- Dividends
- Business Income/Loss
- Capital Gains
- Taxable IRA distributions
- Pension Income
- Unemployment Compensation
- Total Social Security Income
- Contributions to Self-Employed Retirement Accounts (e.g., Keogh plans)
- Deduction for self-employed health insurance
- Contributions to IRAs
- Student Loan Interest
- Deductions for Domestic Production Activities

We accomplish this in two stages. In the first stage, income and deduction amounts are adjusted on a record-by-record basis according to per capita growth factors to ensure that these amounts are extrapolated, or "aged", to the year of the Production File (Stage I). Because the Production File is a composite of the three most recent CPS years, this aging is done separately for each tax unit, depending upon which CPS year the tax unit is constructed. In the second stage (Stage II), we rely on a mathematical optimization algorithm to adjust sample weights on the file to hit the prescribed control totals. A comparison of the state-by-state targeting with published IRS totals is contained in a separate document and available upon request.

---

[13] When this document was being prepared, the IRS released updated, state-level tabulations for tax year 2012. We will rely on these control totals when the 2014 production file is created this Fall.

Stage I

In the first phase of the extrapolation process we adjust the individual data elements on the base year file on a record-by-record basis to correspond to published IRS totals by state. Specifically, we multiply each data element on the file by a per capita adjustment factor that ensures that our macro targets will be hit. These adjustment factors (the "Stage I" factors) are calculated as follows. Define the following variables:

X_0:   the value of some macroeconomic aggregate in the base year (e.g., wages)
P_0:   the population total in the base year (e.g., number of tax returns)
w_{i0}:   sample weight of the ith unit in the base year
x_{i0}:   observed value of variable x reported by unit i in the base year.

With this notation, we have the following relationships that link the macro and micro variables:[14]

$$\sum w_{i0} = P_0 \tag{1}$$

$$\sum x_{i0} \bullet w_{i0} = X_0 \tag{2}$$

Now suppose we have some externally supplied forecast for total population and the aggregate level of our macro variable for some year t:

X_t: the forecasted value of our macroeconomic aggregate in year t
P_t: the forecasted population total in year t.

Define the following growth rates from the base year to year t:

$$(1+r) = \frac{X_t}{X_0} \tag{3}$$

$$(1+p) = \frac{P_t}{P_0} \tag{4}$$

$$(1+\rho) = \frac{(1+r)}{(1+p)} \qquad \text{(per capita adjustment)} \tag{5}$$

To perform the Stage I adjustment to the microdata on our base year file we first apply the population growth factor to the base year sample weights on the file, and second,

---

[14] In practice, the value of X_0 may represent a different concept that what appears in the microdata. For example, wages on the tax return may not include some types of employee compensation that is measured at the macro level.

apply the per capita adjustment factor to the reported base year income on each record. That is, we calculate

$$w_{it} = w_{i0} \bullet (1 + p) \tag{6}$$

$$x_{it} = x_{i0} \bullet (1 + \rho) \tag{7}$$

This ensures that our macro targets are hit in year t. That is, for population we calculate:

$$\sum w_{it} = \sum w_{i0} \bullet (1 + p)$$

$$= (1 + p) \sum w_{i0}$$

$$= \frac{P_t}{P_0} \sum w_{i0}$$

$$= P_t .$$

The first equality is from the definition in **(6)**; the second equality is due to the distributive property; the third equality is from the definition in **(4)**; and the final result is from **(1)**. Similarly, our aggregate macro target is achieved:

$$\sum x_{i,t} \bullet w_{i,t} = \sum x_{i,o} (1 + \rho) \bullet w_{i,o} (1 + p)$$

$$= (1 + \rho)(1 + p) \sum x_{i,o} \bullet w_{i,o}$$

$$= (1 + r) \sum x_{i,o} \bullet w_{i,o}$$

$$= (1 + r) X_0$$

$$= X_t$$

where the first equality defines the Stage I adjustment; the second is from the distributive law; the third is from **(5)**; the fourth is from **(2)**; and the final equality is from **(3)**.

The procedure outlined above is easily extended to accommodate any number of macroeconomic targets coupled with their micro variables and the methodology guarantees that all the targets are reached exactly. We point out that most microsimulation models stop here and perform no other adjustments to the data. However, experience has shown that while this methodology is sufficient to reach any exogenous target, certain endogenous, or jointly determined variables, may need further calibration, especially if the aging is done for more than a few years into the feature. In

the next section, we describe a set of additional adjustments that ensure that both the exogenous and endogenous targets are reached.

Stage II

In the second stage of the extrapolation process, numerical optimization methods are used to adjust the sample weights on the file to ensure all targets are reached. We rely on a linear programming algorithm to solve for a new set of (sample) weights that minimize the absolute value of the percentage change in the weight for each record on the file subject to the constraint that each of the targets are reached.

Mathematically, we let z_i represent the percentage change in the sample weight for record i on the file relative to the base year sample weight:

$$w_{i,t} = w_{i,o}(1 + z_i).$$

Furthermore, to accommodate the absolute value metric, we decompose the z_i into a positive and negative component by defining:

$$r_i = z_i^+ \tag{8}$$

$$s_i = z_i^- \tag{9}$$

$$z_i = r_i - s_i \tag{10}$$

Here, **(8)** and **(9)** define the positive and negative parts, respectively, of the z_i and this leads to the definition of the absolute value:

$$|z_i| = r_i + s_i \tag{11}$$

Our linear programming problem then becomes one of choosing the smallest δ that solves:

$$\min_{\langle r,s \rangle} \sum |z_i| \left(= \sum r_i + s_i \right) \tag{12}$$

subject to:

$$0 \leq |z_i| \leq \delta \tag{13}$$

and

$$Ax = b \tag{14}$$

25

where $x \in (r_i, s_i)$ are the variables to be solved for, A is a coefficient matrix that will depend on the specific targets and b is a target vector.

This procedure is flexible enough to handle many types of targets. In practice, we have usually incorporated four types of constraints:

1. Amount Aggregates (AA)
2. Return Aggregates (RA)
3. Amount by Income Class (AB)
4. Returns by Income Class (RB)

An example of an amount aggregate (AA) target might be total capital gains; a return aggregate (RA) might be the total number of taxpayers reporting capital gains[15]; and targeting both these variables by income class would constitute AB and RB targets, respectively, and be equivalent to targeting average capital gains by income class.

We have recently implemented a Stage-I adjustment to various income classes for selected items to more closely align with IRS totals. These adjustments improved our overall results and reflected in our most recent estimates of the blank slate scenario.

Upon examination of certain state-specific control totals, we determined that a Stage-II adjustment was needed to align the file to IRS control totals, by state, for the following items in addition to the Stage-I targets mentioned above (the type of targeting employed is in parentheses)[16]:

- Wages by Income Class (AB)[17]
- Earned Income Tax Credit (AA)

Here, in no particular order, are some observations we have made over the years in using this Stage II procedure:

- Solving the linear programming (LP) formulation of this problem means doubling the number of variables that need to be solved for (the r_i's and s_i's). Our view is that this is a small price to pay for the assurance that the optimum is achieved under the LP framework.
- It may take several iterations of the algorithm to achieve a solution that is acceptable by iterating over δ.
- Not all variables in the model will (or should) be targeted. Those that aren't will only receive a Stage I adjustment. In fact, after the LP problem is solved, most

---

[15] Targeting both, of course, would be the same as targeting average capital gains.
[16] It's necessary to add the Stage-I targets to the Stage-II process to ensure that the control totals are maintained as the file is re-weighted.
[17] In the current version of our Stage-II adjustments, we target the income distribution according to IRS published figures however we collapse the highest income categories into one income class of $200,000 and over. We do this to minimize the distortion to the weighting for the smaller states where very few high-income returns exist. In future versions of the database, we will experiment with expanding the number of income categories in those states where there are a sufficient number of observations.

returns on the file will receive only a Stage I adjustment. That is, z_i will be zero for most records.

- Experience with this algorithm has show that it is reasonably stable over the forecasting horizon.
- More importantly, those variables that are not targeted in the Stage II extrapolation remain reasonably close to their Stage I values. Put differently, the reweighting does not materially affect the values of these variables.
- In practice, when solving the LP, we usually use the solution for year t as the starting point for obtaining a solution for year t+1. This speeds up the LP algorithm substantially.

Description of the Tax Calculator

The calculator is designed to simulate federal and state income tax liability for individuals in all 50 states and the District of Columbia. The calculator is data-driven, in that inputs from a detailed parameter file guide the calculations for each component of the tax law for each state. We chose this approach because, while more time consuming initially, subsequent updates to reflect changes in tax law can be handled through the parameter files with little if any changes to the calculator itself.

Structure of the Tax Calculator

The universal calculator, a proprietary software module developed by Quantria Strategies, LLC and Analytika to model tax law for the fifty U.S. states and the District of Columbia, utilizes a *microsimulation* approach. The microsimulation approach means that the calculator operates on a micro-level basis, or the decision-making filing unit, to determine how a particular unit would respond to various tax policy provisions. Each unit's decisions are then aggregated over a suitable population to determine the overall impact of such choices amongst various distributions. These distributions include results by state, by income class, and by filing type.

The calculator can generally be considered a *static* calculator in that it does not incorporate traditionally understood economic behavioral responses into its logic, nor does it incorporate dynamic analysis, including feedback analysis based on taxpayer decisions (with an exception noted below).

The model does allow for a limited number of taxpayer choices, although these are not typically considered "true" behavioral responses. These choices can be considered to be "tax form" choices in that the taxpayer may determine whether or not to itemize, or to elect to claim a particular credit, based on a discrete and immediate decision rule – i.e., they do not go back via a feedback loop and change their decision. These rules are based on the available information at the time of the decision.

Behavioral resposes that are not utilized in this model include elasticity assumptions, dynamic revenue feedback impacts of tax policy changes, dynamic revenue feedback impacts of economic-based changes, or the tax liability feedback impact of itemizing federal deductions of state and local taxes. In the latter case, the change in state taxes can cause the federal itemized deductions to change for some filers that would alter their state tax liability. This model does not incorporate any of these types of behavioral responses. The only exception to the generalized rule where the model does incorporate behavior is the model's determination of the benefit of deductions and exemptions available to married joint tax filers that file separate on a combined return. Otherwise, the model does not incorporate any traditonally defined economic behavioral responses.

The universal calculator interacts with a federal tax calculator on a limited basis. Depending on the policy alternatives (i.e., policy switches or parameters), the universal calculator has the ability to utilize a limited number of federal tax concepts for use. Examples of these include the starting base of income used to denote State Adjusted Gross Income (AGI), the Earned Income Tax Credit (EITC), and/or itemized deductions from Schedule A of the federal 1040.

The universal calculator is composed of over 5,000 lines of programming code and utilizes some combination of over 95,000 individual tax parameters. These 95,000 parameters are used for each of the fifty states plus the District of Columbia (hereafter to be included as a "state"), meaning that there are over nineteen hundred (1,900) parameters available for each state.

The structure of the universal calculator is such that it was designed to apply any state tax law upon another state, or to apply a uniform set of tax law to all states. The default position for each state is its current system of tax policy for the most significant provisions of tax law for each state as represented by the state's tax return (or "form"). This calculator does not capture certain provisions in law, especially esoteric or unique policies for any given state or locality. Some of these provisions might be included in the results under an "other" category. Examples types of provisions that this model was not intended to capture include Maine's quahog tax, Maryland's local taxes by county, and Hawaii's tax deduction for exceptional tree maintenance.

Because of the sheer size and complexity of the universal calculator, it would be impossible to describe each and every detail of the calculator in a document of this size. We will, however, work to provide some detail on the workings of the calculator in this methodology.

Much of the functionality of the tax calculator can be affirmed through the detailed parameter file used by the tax calculator. For any particular detail associated with the tax calculator, a parameter is often needed to allow the calculator to compute the proper impact. However this file is quite large and may be too complex for some tax provisions. To assist in gauging general types of tax parameters, a simplified table of tax parameters was created. This table lacks the specificity of the larger parameter file, but may provide some insight into the general flavor of a state's income tax system.

Major Components of the Tax Calculator

The focus of this document is on the state tax calculator that computes the state tax revenue represented by the tax forms for all fifty states plus the District of Columbia. However, a crucial input into this universal calculator are the results from a Federal tax calculator on which much of the state tax forms are based.

Like the state calculators discussed below, the Federal calculator incorporates five (5) major components. They are:

I.   Income subject to taxation;
II. Deductions from income;
III. Exemptions from income;
IV. Tax concepts before and after credits; and,
V.   Credits, both non-refundable and refundable, against tax liability.

Each of these sections is important to determining the revenue liability due the taxpayer. The federal calculator applies equally to all of the states.

Income is determined from the basic income components. Items such as wages and salaries, interest income, dividend income, capital income, pension income, Social Security income, and others are all included in the gross income base. Adjustments to that income are provided for employment taxes paid, or contributions for medical or health related expenses, business expenses, and/or retirement contributions. This results in the concept known as Federal Adjusted Gross Income (AGI) which is the starting point for the overwhelming majority of states with an income tax system.

Deductions from federal AGI occur as as either standard deductions or, for filers having a larger amount of special deductions, itemized deductions. As with the state tax calculator, the taxpayer's choice of deduction is chosen at a distinct time. This differs from a true economic behavioral response and does not incorporate feedback mechanisms. Itemized deductions are subject to means testing and a phaseout of benefits for filers above the testing threshold.

Exemptions from AGI are provided as a deduction for tax filer(s) and certain types of dependents that rely on the tax filer(s) for their care and welfare. The exemption amount is the same for most tax filer(s) and their dependents at the Federal level, and is subject to a reduction for higher income taxpayers.

Tax concepts include items like Taxable Income, which is Federal AGI less all deductions and exemptions, and all tax liability concepts like Tax Before Credits, Tax After Credits, and the Federal Alternative Minimum Tax. The former are based on the regular progressive tax system of rates and brackets as applied against Federal Taxable Income and may include the impact of tax credits. The latter is a parallel tax system in which certain taxpayers are faced with a minimum amount of liability owed due to the disallowance of some tax preference items. For taxpayers subject to the AMT, the taxpayer owes the larger of the liability due using the regular tax system of rates and brackets as compared against the Alternative Minimum Tax system.

Credits are programs that reduce tax liability. Federal credits are typically provided to help lower-income taxpayers or special categories of taxpayers, such as parents. Federal credits come in two varieties: those that may reduce or eliminate the amount of regular tax (non-refundable credits) and those that may result in a tax rebate to the taxpayer (refundable credits).

All credits are used to reduce tax liability, but refundable credits are allowable in excess of the taxpayer's liability and can therefore be seen as a grant or budget expenditure to the taxpayer. Two refundable federal tax credits are the additional child tax credit and the Earned Income Credit (EIC). Non-refundable federal tax credits include the retirement credit, the child care and dependent tax credit, the elderly credit, and the regular child tax credit.

Income

The first major concept is for income subject to taxation. This amount begins with a gross income concept built from either a pre-existing base, such as Federal AGI or Federal taxable income, or the model calculates an income base built from individual components of income such as wages, interest, dividends, capital gains, etc. Most of these components are sourced from database concepts, but some concepts like Social Security Income, are calculated based on other database information. Other concepts, like refunds of state taxes, special other gains, and other income, were not available from the database. For a complete list of items included and excluded, please refer to the enclosed sample federal 1040 tax form.

Many of these concepts follow federal tax policy as a default assumption, but the calculator is able to adjust these definitions subject to selectable policy switches, scalar adjustments, or other available database information.

After determining the gross income concept, adjustments to income are applied. Due to limited availability of data, we were not able to include all potential federal adjustments in the calculator. Please see the enclosed sample tax form with the available data items that are available for use. All adjustments that were not available for use were imputed as a residual component ("other" income) based on aggregated tabulations from external sources (e.g., the Statistics of Income for each state).

After computing both the gross income and the adjustments, the calculator computes an Adjusted Gross Income (AGI) concept for use.

Many states, but not all, begin their tax form with the federal AGI concept. For those that utilize the federal AGI income base, we included a switch in the model to use the federal AGI as calculated in our federal calculator. However, many of the underlying data items are subject to availability, and thus will encounter the same type of missing data as described above. For some states, like Minnesota, the starting income base is federal taxable income. The calculator has a parameter to begin with federal taxable income, which is sourced from the federal calculator and includes the effects of the federal deductions and exemptions.

But some states begin their tax forms by building their own income base. New York, California, and Pennsylvania are all examples of states that build their own tax base on their state tax forms. For New York and California, this is equivalent to the federal AGI concept. The model has the ability to build these state bases from the components, but we

incorporated the state bases for California and New York as if they began with Federal AGI.

Pennsylvania is a different case, in which it builds up a base that has different definitions of income from the federal tax concept. Again, because of constraints on available data, we tried to model the Pennsylvania income base as closely as possible but it will be using the same wage concept that we use as our federal wage concept (which is based on definitions from the Current Population Survey). While this approach is imperfect, we felt that this was a reasonable approximation of the income base given the available data.

Finally, for joint units filing separate on a combined return, database values determine a suitable income split for each spouse. This will be utilized for states that allow filing separate on a combined return to reduce the tax liability due for that filing unit. That income split is based on a Current Population Survey concept.

After computing federal AGI, states adjust that figure using modifications to income in the form of additions to income and subtractions from income. After adjusting for these modifications, a State AGI concept is determined. States can have many different types of modications to income – some are based on conformity issues with the federal tax system, as evidenced by additions to income from bonus depreciation assets, others are based on providing tax relief for married couples with children, as shown by the child and dependent care subtraction from income.

Because we are working with the full set of states and we are working with a limited dataset, we were unable to capture all of the modifications to income. Our approach involved identifying the most significant areas of interest for modifications, and then, assuming that we had sufficient data to properly model the modification, including only those modifications where we had sufficient data and where the size/scope of the modification was significant.

Deductions

Deductions are split amongst two main categories – standard deductions and itemized deductions.

Standard deductions are modeled as fixed amounts relative to the filing status and available to all tax filers, whereas itemized deductions are special deductions only available to certain filers. For some states, like Maryland, there are special rules to phase out portions of the standard deductions, and these were included in our calculator. Because the states can utilize many different systems to determining deductions, the calculator has many different methods for computing standard (and itemized) deductions. For instance, the calculator requires one method for phasing out a portion of the standard deduction for Maryland, whereas Wisconsin utilizes a more complicated system of brackets and phase-outs for their standard deduction. The calculator captures a wide variety of brackets and/or continuous and discrete phase-outs to model a particular state's tax system.

Itemized deductions were computed/imputed following a similar methodology to both income items and adjustments to income. For more information on the basis of itemized deductions, please see the "Imputation of Additional Fields" section of the methodology and/or the sample federal Schedule A enclosed.

Itemized deductions may be subject to a phase out at the federal level and/or for some states. The calculator takes into account these interactions subject to prevailing tax law as presented on the relevant tax return. For many states, the itemized deductions are sourced from the federal tax return and reduced by the deduction for state and local taxes paid by the host state. Since we do not have data on taxes paid by locality, some assumptions were made as to the overall amount of taxes paid reported on federal Schedule A, and on how much of these aggregate amounts were sourced from the host state. In the current version of the calculator, we simulated the taxes paid for each state by running the calculator at current levels under current law. The tax computed from that step was used as an approximation of the taxes paid by a given state for the prior year.

Once the deduction amounts are determined, the calculator takes both the standard deduction and the potential itemized deduction amounts into account. The larger of the two types of deductions (standard or itemized) is used by the calculator in determining the eventual deduction from income used in the model. The calculator does allow for switching from standard to itemized deduction subject to available data. This is labeled in this methodology as a taxpayer "choice" but is not a traditionally defined economic "behavioral" assumption.

The calculator also computes the "maximum" benefit attributable to joint tax filers that file separate on a combined tax return should the state allow such a filing. The model computes taxes for joint filers under two sets of assumptions: first, that they joint couple file a single return, and second (if allowed under the tax law), to compute the tax as if the joint couple filed separately on a combined return. The model will then determine the minimum amount of taxes paid by that joint couple, and use the lower amount of tax liability. Because this approach does incorporate feedback that occurs within a loop framework, this is a form of the traditionally defined economic behavioral response that incorporates dynamic feedback and is the sole exception (inclusive for both deductions and exemptions) to the behavioral and/or dynamic feedback rule as previously noted. This "behavioral-type" response differs from a taxpayer choice to elect deduction type and is done to determine the most efficient usage of deductions available to the filing unit(s). However, there are often many idiosyncratic rules for each state about which deductions are allowed for which half of the joint filing unit, and therefore our assumption and calculation of the benefit will not be fully representative of what the filing unit does in "real" life. Therefore, our model is likely to _overstate_ the benefit relative to married joint filers filing separate on a combined return, and therefore _understate_ their overall tax liability, as we assume taxpayers will make the optimal use of deductions.

Exemptions

Exemptions may be treated as a reduction in income amounts or as a credit against tax liability. We will cover both types of exemptions in this section.

There are two basic forms of exemptions claimed against income in the universal tax calculator. The first are personal exemptions that are subject to the individual status of the filing unit. There are basic exemptions for filing status and exemptions for elderly, blind, or disabled status. Personal exemptions apply to a person, but apply also to the secondary (or spouse) tax filing unit for some filing statuses.

The second category of exemptions applies to the non-personal exemptions, mainly for children and/or other dependents reliant on the filing unit. Subject to available data, the calculator computes amounts for various categories of children, parents, or other persons whom are economically dependent on the filing unit for support.

The calculator incorporates phase-outs for the value of the exemption whether stated as a deduction and/or a credit, and contains phase-out provisions by discrete income classes, continuous phase-outs, and hard limits (which vary according to state law) that might be due to the maximum amount of exemptions allowed or by the maximum income allowed to claim an exemption.

For married joint units filing separate on a combined return, the value of exemptions claimed are computed by dividing the number of exemptions between spouses. Each spouse begins with no exemptions, and the calculator adds exemptions to the spouse having the most taxable income amount. The value of these exemptions reduces each spouse's taxable income and optimizes the use of exemptions for the mutual benefit of both spouses.

As noted above, some exemptions are taken as non-refundable credits against tax liability. California uses exemptions as credits instead of as a reduction in AGI. The calculator has the ability to model both types of exemptions and allows for one or both situations to be modeled using the parameters.

When exemptions are claimed against tax liability, they generally follow the same procedures as when taken as a reduction from income. There are personal exemption credits available to the tax filer(s) as well as exemption credits for children/parents/other dependents. We did identify states that phase out the value of the exemption credits (e.g., Oregon in 2012, although they adjusted their law in 2013). Therefore, the calculator includes a phase out of exemption credits based on AGI.

Exemption credits are treated as non-refundable tax credits in the calculator, and are the first non-refundable tax credits taken in the stacking order.

Tax

The tax concept is the basis for all of the work performed by the calculator. Tax is simply the amount of liability, not the cash concept, owed to or by the taxpayer based upon their taxable income (a.k.a. regular tax before credits), and subject to reduction from various tax credits (a.k.a. tax after credits). This calculator primarily deals with state income tax liability on a calendar year basis. It does not currently measure/output any payroll tax concepts, local tax liability, or other forms of liability/benefits. Because some income tax credits are refundable (e.g., the EIC), the tax liability can be negative in value and this represents a benefit accrued to the taxpayer/filing unit.

Tax is based on taxable income. Taxable income is simply the gross income concept less the following concepts:

I.   Adjustments to gross income (equal to State AGI);
II. Deductions from State AGI; and,
III. Exemptions from State AGI.

As noted earlier, some states, such as Minnesota, begin their tax system using federal taxable income. In the current version of the output, Minnesota would have no federal AGI, no State AGI, no deductions, and no exemptions shown in the output. Their output would "begin" with Taxable Income.

Most state tax systems use a progressive tax rate structure. Some have rate structures that vary by filing status, while others utilize a single flat tax rate. Some states use as many as twelve tax rates and brackets for determining their regular tax before credits. The calculator is capable of modeling all of these types of tax systems.

Regular tax liability is computed for the filing unit. For married joint units filing separate on a combined return, individual regular tax amounts are computed per spouse and added together to the form the final amount of regular tax due. These tax amounts are computed based on each spouse's individual taxable income concept, which begins with income splitting based on database reported amounts, and reduced by the impact of deductions and exemptions. As noted earlier, the deductions and exemptions computed for each spouse are maximized/optimized following simple algorithms. The algorithms attempt to maximize the benefits for each spouse by adjusting the value of deductions and exemptions claimed for each taxpayer.

The key assumptions in these algorithms are that deductions and exemptions can be moved between spouses, and that the taxpayers will take advantage of the tax brackets to move deductions and exemptions such that each spouse uses as much income as possible in the lower tax brackets. By using up the lower tax brackets, less income is therefore taxed at higher brackets, and in theory the most optimal use of deductions and exemptions to lower taxable income is achieved.

Once regular tax is computed, credits are applied against the regular tax amount (see following section for details). After applying all non-refundable and then the refundable tax credits, the final state tax liability concept is determined. This estimate is the total amount of income tax revenue due to the State. This estimate forms the numerator for the effective tax rate (ETR) calculation in the output tables, with a broad-based income measure serving as the current denominator.

Credits

There are two types of tax credits used in the universal calculator – non-refundable tax credits and refundable tax credits. The difference is that refundable tax credits can reduce tax liability beyond zero (or negative liability), and this represents a benefit accrued to the filing unit. Non-refundable tax credits can only reduce liability to zero (or no tax liability is owed by the filing unit).

There are many credits in use for any given state. Our preliminary tax research included hundreds of non-refundable tax credits in use by the states, and this did not include obviously targeted tax credits for idiosyncratic areas of State tax law. For our purposes, we were interested in capturing the largest tax credits in a given State, and therefore we dramatically reduced the list of usable credits by 90% (or more). Where we once had hundreds of credits, the universal calculator only explicitly models on average about 4-5 non-refundable tax credits per state. This includes a "catch-all" category for "Other State Tax Credits". These "Other State Tax Credits" are assumed to be independent of the federal tax calculator and unique to the states providing them. However, the model also has the option of including the impacts on the states for credits based on a percentage of a Federal tax credit such as the elderly credit or the retirement credit. None of these are included in our baseline results unless the state offers this type of credit like many states' EIC and/or a state child care credit based on some percentage of the Federal amount. Therefore we are implicitly modeling the impact of other credits in our liability. This is not to say that any given credit is unimportant. All credits have their own reasons for coming into use. But for our purposes, we needed to be able to model large-scale tax interactions across all fifty-one states, and we had to simplify the issues. This also applies to refundable credits, although the general ratio of refundable credits to non-refundable credits is low.

We identified six distinct types of non-refundable tax credits. They are the exemption credit, the child and dependent care credit, the child tax credit, low-income credits, family size credits, and an "other" category for all remaining state credits.

The exemption tax credits are provided to offset tax liability for the number of persons on a tax return and are used by states in lieu of providing exemption deductions. Exemption credits are applied to personal exemptions for the tax filer(s) and for non-personal exemptions for children, parents, or other dependents.

The child and dependent care tax credits are used to offset liability used in the care and welfare of children supported by the tax filer(s). These credits are typically provided as a percentage of the Federal child and dependent care credit and are usually non-refundable.

Child tax credits are provided explicitly for the children of the tax filer(s). These are both non-refundable and refundable at the Federal level, but are less common amongst the states and are often limited to being non-refundable for states.

Low-income tax credits in use at the states come in a variety of formats. They are generally designed to assist tax filers that earn income less than the Federal Poverty Levels (FPL). They are often intermingled with family size credits as these credits are often based initially on the FPL. In our current model there are over ten (10) types of low-income credits. New Mexico's Low-Income Comprehensive Tax Rebate is a good example of the low-income credit, as this is a comprehensive credit designed to take into account all taxes faced by low-income earners and not just the income tax specifically.

Family size credits are provided based on the size of a family, and in some cases on the income amounts available to the family. These can be considered very similar to a type of low-income credit, but some states do distinguish the family credit from low-income credits. Generally speaking, the family size is applicable to families with low to moderate incomes and increases in size relative to the number of members in the family.

Other state tax credits are just a residual amount of credits that were computed to account for all other amounts. Many states offer a plethora of tax credits, often targeted to small and/or unique interests for the state. There are many cases where we did not have available data to explicitly model these credits. Where we were able to deduce the size of the state's tax liablity and/or credit amounts not specified, we included all of these impacts as other state tax credits. Where we were unable to locate these estimates the other state tax credit amounts would be zero.

With that in mind, we explicitly modeled the following non-refundable tax credit programs in the following stacking order:

   I.   Exemption tax credits;
   II. Child and dependent care tax credits;
   III. Child tax credits;
   IV. Low-income tax credits;
   V.   Family size credits; and,
   VI. Other tax credits.

The stacking order of credits was determined somewhat arbitrarily although some attention was paid to the general order of the tax credits found on the tax forms. Since there is little conformity amongst the states in this regard, the stacking order tried to adhere to some of these rules but could not be determined definitely. As an example, many states that offer exemption tax credits offered those before computing any other credit, so those were included first in our stacking order. The remaining credits were done

based on some arbitrary combination of the location of the credit on the form and/or the size or prevalence of the credit.

The other type of tax credit is the refundable tax credit. For most states, this is the State version of the federal EIC. Some states take a percentage of the federal EIC, and many states that offer the EIC also allow it to be refundable against tax liability due. There are a few states that offer a non-refundable EIC without offering a refundable portion or they split it into two distinct credits. Any refundable credit that exceeds a filer's tax liability is shown as a negative tax liability in our results.

Some states offer other refundable tax credits, such as circuit-breaker property tax credits for the elderly and/or the poor, or refundable credits for historical buildings. As with other segments of our calculator, computing the effect of these programs was dependent upon having a suitable amount of data to accurately model the provision and having the tax credit be of sufficient size to warrant being modeled. The above noted tax credit for historical buildings is refundable in some locations, but it is often a very small net dollar figure because there are not many locations that are eligible to receive such a benefit.

The circuit breaker tax credits posed a unique problem amongst credits. First, they are not universally captured as part of the income tax system. A number of states allowed circuit breaker credit refund programs found outside of the income tax system. This would have led to an apples to oranges comparison to other states where the credits were included on the tax return. Second, there is some debate as to the applicability of local tax refunds for a state tax system. From a budgetary perspective, not all circuit breaker monies are considered state funds. Third, there is also some debate about whether the circuit breaker program is part of the income tax system or the property tax system. Therefore, it was decided to omit the impact of circuit breaker refund programs from the model.

For these types of targeted tax credits, we made a decision to not capture their impacts.

Notes on the Tax Parameters

The universal tax calculator operates through a combination of computer programming code and tax parameters designed to interact with the computer code. For any change/option available in the calculator, there likely exists a parameter that controls how a particular tax provision will be modeled.

Several different types of tax parameters exist, and this section provides a brief discussion on the types of available parameters for use in the calculator. The following list summarizes the major types of tax parameters:

I.   Scalar adjustments;
II.   Switches;
III.   Limits;
IV.   Phase out thresholds and percentages; and,
V.   Tax rates and brackets.

All of these types of parameters are mathematical functions expressed in a tax-specific vocabulary. They can be combined together or found separate in the model. Essentially, they are the computable manifestation of the legal framework behind any proposed or existing tax policy.

Scalar adjustment

The first type of tax parameter is the scalar adjustment. This is a multiplicative factor assigned to a given data element. In most cases, this is a base data element from the database.

A simple example of a scalar adjustment is the amount of income attributable to wages. Wage income is often fully taxable and therefore would be included at 100% in the calculator for most states. Another (federal) example would be for Tier I Social Security Income, which is included at 50% of the base SSI, and Tier II that is included at 85%. An item could be doubled, which would indicate a scalar adjustment of "200%".

Most elements used in the model have some form of scalar adjustment inside of the tax calculator.

Switches

The second type of tax parameter is really a subset of the scalar adjustment noted above. A switch is merely a scalar adjustment with a value of either "0" or "1", denoting the ability to turn "on" or "off" a provision respectively.

An example of this parameter type is logic to allow a state to compute both a low-income credit and the State's version of the EITC. Since these programs often target the same or similar demographic subpopulations, one often makes the other unnecessary as well as the potential for abusive gaming of the tax system. Therefore, states often allow the taxpayer to elect one type of credit and not both. The tax calculator has the ability to model one or both of the provisions dependent on the value of the "switch" parameter.

Limits

The third type of parameter is the limit. This is a minimum ("floor") or maximum ("ceiling") amount tied to a tax provision. Most limits are maximum amounts that prevent the tax provision from exceeding a certain amount. Some limits prevent a tax provision from being reduced beyond a certain amount (the "floor"), but these are much less frequent, and this calculator is not currently using any "floor" limits.

Some states allow for deductions of pension income, and this would be represented by a limit parameter. In other words, a filer could omit their pension income up to a prescribed amount, but the value of their deduction/omission would be limited to the proscribed amount.

Phase-out thresholds and percentages

The fourth type of tax parameter is the phase out, which gradually reduces tax benefits for higher income filers. The phase-out is a specialty tax parameter with a tax-specific vocabulary.

Phase-outs utilize limits or income thresholds by filing status, a maximum level of phase-out based on a percentage reduction, and the rate of the phase-out. The combination of these parameters represents a phase-out range used by the calculator.

Phase-outs are not common at the State level, but a few states utilize them for itemized deductions. Some follow federal convention (Vermont, Montana) whereas the District of Columbia incorporated a simpler version for use after the federal Pease phase-out provision was eliminated (and subsequently returned after a "sunset" provision).

Tax rates and brackets

The final type of parameter is the tax rates and brackets. These are specialty combinations of limits and percentages used to compute the tax liability. Federal income taxes are computed on a graduated basis, whereby incomes are taxed at a progressively higher rate. Most U.S. states follow this convention, and those that do not either do not have any income taxes or are based on a "flat" or single tax rate. The calculator allows for a system of graduated rates and brackets to compute regular tax liability.

Tax rates and brackets are based on filing status. The ability to calculate the combined taxes of a married joint couple filing on a single return is available. Because this calculator operates on a single year basis, there is no computable indexing of brackets. All amounts represent a single tax year.

Understanding the interactions of tax parameters with the calculator

The above represents a simplistic overview of the types of tax provisions. However, it serves as an introduction to the use of the tax parameters and how they operate within the universal tax calculator.

This description of the calculator provides a summary of the major components and issues involved in its construction. Significant detail has been omitted for brevity. For more detail on a particular state, see the Tax Parameter Summary table.

Appendix A – Summary Statistics for 2012 Production File

Table A-1 Record Counts by State

State	Unweighted		Weighted	
	Number	Percent	Number	Percent
California	28,848	9.5%	18,591,362	11.8%
Texas	17,050	5.6%	12,080,787	7.7%
New York	13,596	4.5%	10,194,017	6.5%
Florida	12,448	4.1%	9,817,047	6.2%
Illinois	9,341	3.1%	6,581,282	4.2%
Pennsylvania	8,832	2.9%	6,734,867	4.3%
Ohio	7,738	2.6%	6,018,283	3.8%
Maryland	7,533	2.5%	3,023,315	1.9%
Minnesota	6,871	2.3%	2,794,334	1.8%
Michigan	6,827	2.3%	5,040,598	3.2%
Connecticut	6,743	2.2%	1,842,471	1.2%
Colorado	6,459	2.1%	2,549,435	1.6%
Virginia	6,381	2.1%	3,968,737	2.5%
Georgia	6,377	2.1%	4,810,027	3.1%
New Jersey	6,369	2.1%	4,454,990	2.8%
North Carolina	5,848	1.9%	4,757,790	3.0%
New Hampshire	5,751	1.9%	699,258	0.4%
Hawaii	5,736	1.9%	680,211	0.4%
Wisconsin	5,608	1.9%	3,002,926	1.9%
Iowa	5,476	1.8%	1,600,475	1.0%
Maine	5,233	1.7%	733,108	0.5%
District of Columbia	5,085	1.7%	382,451	0.2%
Washington	5,060	1.7%	3,552,601	2.3%
Rhode Island	4,991	1.6%	577,215	0.4%
Nevada	4,808	1.6%	1,364,551	0.9%
Delaware	4,745	1.6%	462,224	0.3%

Table A-1 Record Counts by State (Continued)

State	Unweighted		Weighted	
	Number	Percent	Number	Percent
Missouri	4,735	1.6%	3,136,568	2.0%
Massachusetts	4,619	1.5%	3,562,578	2.3%
South Dakota	4,561	1.5%	428,558	0.3%
Nebraska	4,557	1.5%	939,082	0.6%
Indiana	4,247	1.4%	3,156,418	2.0%
Kansas	4,227	1.4%	1,440,581	0.9%
Vermont	4,177	1.4%	352,379	0.2%
Kentucky	4,168	1.4%	2,259,227	1.4%
Oregon	4,128	1.4%	2,018,339	1.3%
Wyoming	3,941	1.3%	292,376	0.2%
South Carolina	3,874	1.3%	2,300,571	1.5%
Arizona	3,833	1.3%	3,245,682	2.1%
Tennessee	3,832	1.3%	3,240,482	2.1%
Alaska	3,801	1.3%	362,308	0.2%
Oklahoma	3,621	1.2%	1,845,307	1.2%
North Dakota	3,559	1.2%	351,258	0.2%
Alabama	3,328	1.1%	2,368,258	1.5%
Utah	3,314	1.1%	1,245,344	0.8%
Idaho	3,074	1.0%	731,416	0.5%
Arkansas	2,991	1.0%	1,497,033	1.0%
West Virginia	2,933	1.0%	947,585	0.6%
New Mexico	2,898	1.0%	994,732	0.6%
Mississippi	2,852	0.9%	1,421,424	0.9%
Louisiana	2,838	0.9%	2,172,569	1.4%
Montana	2,676	0.9%	527,621	0.3%
Total, All States	302,538	100.0%	157,152,058	100.0%

Source: **State Modeling Project**

State Codes

1	=	"Alabama"
2	=	"Alaska"
4	=	"Arizona"
5	=	"Arkansas"
6	=	"California"
8	=	"Colorado"
9	=	"Connecticut"
10	=	"Delaware"
11	=	"District of Columbia"
12	=	"Florida"
13	=	"Georgia"
15	=	"Hawaii"
16	=	"Idaho"
17	=	"Illinois"
18	=	"Indiana"
19	=	"Iowa"
20	=	"Kansas"
21	=	"Kentucky"
22	=	"Louisiana"
23	=	"Maine"
24	=	"Maryland"
25	=	"Massachusetts"
26	=	"Michigan"
27	=	"Minnesota"
28	=	"Mississippi"
29	=	"Missouri"
30	=	"Montana"
31	=	"Nebraska"
32	=	"Nevada"
33	=	"New Hampshire"
34	=	"New Jersey"
35	=	"New Mexico"
36	=	"New York"
37	=	"North Carolina"
38	=	"North Dakota"
39	=	"Ohio"
40	=	"Oklahoma"
41	=	"Oregon"
42	=	"Pennsylvania"
44	=	"Rhode Island"
45	=	"South Carolina"
46	=	"South Dakota"
47	=	"Tennessee"
48	=	"Texas"
49	=	"Utah"
50	=	"Vermont"
51	=	"Virginia"
53	=	"Washington"
54	=	"West Virginia"
55	=	"Wisconsin"
56	=	"Wyoming"

Appendix B – Estimation Results from the PUF

Table B-1A

```
Logistic regression                              Number of obs    =      150047
                                                 Wald chi2(7)     =    10406.54
                                                 Prob > chi2      =      0.0000
Log pseudolikelihood =  -28484183                Pseudo R2        =      0.2424

------------------------------------------------------------------------------
             |               Robust
    hasgains |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnintst |   .1261299   .0067827    18.60   0.000     .1128361    .1394237
       lndbe |   .2923718   .0054882    53.27   0.000     .2816151    .3031284
  lnpensions |  -.0011029   .0041789    -0.26   0.792    -.0092935    .0070876
       mtxrt |   2.546112    .233159    10.92   0.000     2.089129    3.003096
        mars |   .5033161   .0463062    10.87   0.000     .4125576    .5940746
     famsize |  -.0672656   .0190194    -3.54   0.000     -.104543   -.0299881
       agede |   -.020226   .0327087    -0.62   0.536    -.0843339    .0438819
       _cons |  -4.188864   .0543172   -77.12   0.000    -4.295324   -4.082405
------------------------------------------------------------------------------
```

Table B-1B

```
      Source |       SS           df       MS              Number of obs =    51475
-------------+------------------------------            F(  7, 51467) =   842.86
       Model |  33122.9407         7   4731.84867        Prob > F      =   0.0000
    Residual |  288938.695     51467   5.61405745        R-squared     =   0.1028
-------------+------------------------------            Adj R-squared =   0.1027
       Total |  322061.636     51474   6.25678276        Root MSE      =   2.3694

------------------------------------------------------------------------------
     lngains |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnintst |   .2081099   .0037859    54.97   0.000     .2006894    .2155304
       lndbe |   .0520541   .0032867    15.84   0.000     .0456121     .058496
  lnpensions |  -.0409517   .0025444   -16.09   0.000    -.0459387   -.0359647
       mtxrt |   2.834971   .1726536    16.42   0.000     2.496568    3.173374
        mars |   .2913975   .0303094     9.61   0.000     .2319907    .3508042
     famsize |   .0301172   .0121332     2.48   0.013      .006336    .0538985
       agede |   .1055183   .0203204     5.19   0.000     .0656901    .1453466
       _cons |   5.714152   .0413883   138.06   0.000     5.633031    5.795274
------------------------------------------------------------------------------
```

Table B-2A

```
Logistic regression                              Number of obs    =       96571
                                                 Wald chi2(5)     =     2240.21
                                                 Prob > chi2      =      0.0000
Log pseudolikelihood = -4183060.1                Pseudo R2        =      0.1600

------------------------------------------------------------------------------
             |               Robust
      hassep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |    1.00606   .0230993    43.55   0.000     .9607859    1.051334
       mtxrt |   2.203826    .898308     2.45   0.014     .4431744    3.964477
        mars |   .3537818   .0884371     4.00   0.000     .1804483    .5271153
     famsize |  -.1841447   .0265683    -6.93   0.000    -.2362175   -.1320719
       agede |   -.600306   .0881553    -6.81   0.000    -.7730872   -.4275247
       _cons |  -14.59491   .2998766   -48.67   0.000    -15.18265   -14.00716
------------------------------------------------------------------------------
```

Table B-2B

```
      Source |       SS          df       MS              Number of obs =   11500
-------------+------------------------------              F(  5, 11494) =  598.20
       Model | 3491.21855         5   698.243711          Prob > F      =  0.0000
    Residual | 13416.2556     11494   1.16723991          R-squared     =  0.2065
-------------+------------------------------              Adj R-squared =  0.2061
       Total | 16907.4741     11499     1.470343          Root MSE      =  1.0804


------------------------------------------------------------------------------
       lnsep |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   .4612973   .0092044    50.12   0.000     .4432551    .4793395
       mtxrt |  -.0625158    .334101    -0.19   0.852    -.7174107    .5923792
        mars |  -.2515508    .032092    -7.84   0.000    -.3144565    -.188645
     famsize |   .0515633   .0105577     4.88   0.000     .0308683    .0722582
       agede |  -.1070076   .0274535    -3.90   0.000     -.160821   -.0531941
       _cons |    3.68409   .1052346    35.01   0.000     3.477812    3.890367
```

Table B-3A

```
Logistic regression                           Number of obs   =      96571
                                              Wald chi2(5)    =     355.53
                                              Prob > chi2     =     0.0000
Log pseudolikelihood =  -11245131             Pseudo R2       =     0.0228


------------------------------------------------------------------------------
             |               Robust
  hassehealth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   .2749375   .0214332    12.83   0.000     .2329293    .3169457
       mtxrt |   -1.34596   .3261351    -4.13   0.000    -1.985173   -.7067468
        mars |   .3205253   .0562314     5.70   0.000     .2103139    .4307368
     famsize |  -.1552902   .0192964    -8.05   0.000    -.1931104     -.11747
       agede |  -.3188473   .0439173    -7.26   0.000    -.4049236   -.2327711
       _cons |  -4.471596   .1971543   -22.68   0.000    -4.858012   -4.085181
------------------------------------------------------------------------------
```

Table B-3B

```
      Source |       SS          df       MS              Number of obs =   23313
-------------+------------------------------              F(  5, 23307) =  773.61
       Model | 3307.68416         5   661.536832          Prob > F      =  0.0000
    Residual | 19930.6273     23307   .855134823          R-squared     =  0.1423
-------------+------------------------------              Adj R-squared =  0.1422
       Total | 23238.3115     23312   .99683903           Root MSE      =  .92474


------------------------------------------------------------------------------
   lnsehealth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   .1158923   .0040707    28.47   0.000     .1079135    .1238712
       mtxrt |  -.2671862   .0914992    -2.92   0.004    -.4465307   -.0878418
        mars |   .3870376   .0175922    22.00   0.000     .3525557    .4215195
     famsize |   .0896552    .006248    14.35   0.000     .0774088    .1019016
       agede |  -.1516015   .0132659   -11.43   0.000    -.1776035   -.1255996
       _cons |   6.423675   .0386008   166.41   0.000     6.348015    6.499335
```

45

Table B-4A

```
Logistic regression                              Number of obs  =     150047
                                                 Wald chi2(5)   =     911.15
                                                 Prob > chi2    =     0.0000
Log pseudolikelihood = -14681957                 Pseudo R2      =     0.0603

------------------------------------------------------------------------------
             |             Robust
      hasira |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   .3029358   .0296898    10.20   0.000     .244745     .3611267
       mtxrt |   4.045252   .5742608     7.04   0.000    2.919722    5.170783
        mars |   1.199263   .0832203    14.41   0.000    1.036154    1.362372
     famsize |  -.2944323   .0316807    -9.29   0.000   -.3565253   -.2323393
       agede |  -.3679672   .0492988    -7.46   0.000   -.4645912   -.2713433
       _cons |  -7.475749   .2703289   -27.65   0.000   -8.005583   -6.945914
------------------------------------------------------------------------------
```

Table B-4B

```
      Source |       SS       df       MS             Number of obs =      6032
-------------+------------------------------          F(  5,  6026) =    109.98
       Model |  471.134855      5   94.226971          Prob > F      =    0.0000
    Residual |  5162.87088   6026   .856765827         R-squared     =    0.0836
-------------+------------------------------          Adj R-squared =    0.0829
       Total |  5634.00573   6031   .934174387         Root MSE      =    .92562

------------------------------------------------------------------------------
       lnira |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   .1615735   .0124577    12.97   0.000    .1371519     .185995
       mtxrt |  -1.080339   .2850585    -3.79   0.000   -1.639156   -.5215226
        mars |   .3977286   .0338476    11.75   0.000    .3313752     .464082
     famsize |  -.1247386   .0130845    -9.53   0.000   -.1503888   -.0990884
       agede |   .2099949    .030002     7.00   0.000    .1511802    .2688095
       _cons |   6.128966   .1200072    51.07   0.000    5.893709    6.364223
------------------------------------------------------------------------------
```

Table B-5A

```
Logistic regression                              Number of obs  =     150047
                                                 Wald chi2(5)   =    1116.97
                                                 Prob > chi2    =     0.0000
Log pseudolikelihood = -27262661                 Pseudo R2      =     0.0586

------------------------------------------------------------------------------
             |             Robust
 hasstloanint |    Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |    .118125   .0162361     7.28   0.000    .0863029    .1499471
       mtxrt |   5.236319   .4138389    12.65   0.000     4.42521    6.047428
        mars |   .6284081   .0546494    11.50   0.000    .5212973    .7355189
     famsize |  -.1237507   .0203366    -6.09   0.000   -.1636097   -.0838917
       agede |  -1.965511   .0962696   -20.42   0.000   -2.154196   -1.776826
       _cons |  -4.795339   .1392383   -34.44   0.000   -5.068241   -4.522437
------------------------------------------------------------------------------
```

Table B-5B

```
      Source |       SS           df       MS            Number of obs =     3504
-------------+------------------------------           F(  5,  3498) =     7.15
       Model |  66.6432453          5  13.3286491        Prob > F      =   0.0000
    Residual |  6523.90366       3498  1.86503821        R-squared     =   0.0101
-------------+------------------------------           Adj R-squared =   0.0087
       Total |   6590.5469       3503  1.88140077        Root MSE      =   1.3657
```

```
------------------------------------------------------------------------------
   lnstloanint |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |    .047585   .0325637     1.46   0.144    -.0162608    .1114307
       mtxrt |     .49329   .5429405     0.91   0.364    -.5712221    1.557802
        mars |   .2809594   .0717726     3.91   0.000     .1402389    .4216798
     famsize |  -.0480428   .0257188    -1.87   0.062    -.0984681    .0023826
       agede |  -.2248277   .1450349    -1.55   0.121    -.5091892    .0595338
       _cons |   5.011366   .2945613    17.01   0.000     4.433837    5.588896
------------------------------------------------------------------------------
```

Table B-6

```
Tobit regression                                 Number of obs   =     150047
                                                 LR chi2(5)      =   12276.79
                                                 Prob > chi2     =     0.0000
Log likelihood = -1301794.9                      Pseudo R2       =     0.0047
```

```
------------------------------------------------------------------------------
          cc |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   14634.42   201.6009    72.59   0.000     14239.28    15029.55
       mtxrt |   26897.52   3420.868     7.86   0.000     20192.69    33602.35
        mars |   3165.557   493.9757     6.41   0.000     2197.374    4133.739
     famsize |  -517.9668   179.6765    -2.88   0.004    -870.1292   -165.8044
       agede |  -8538.678   447.2587   -19.09   0.000    -9415.296    -7662.06
       _cons |  -181614.5   1936.906   -93.77   0.000    -185410.8   -177818.2
-------------+----------------------------------------------------------------
      /sigma |   48774.59   113.4698                      48552.19    48996.99
------------------------------------------------------------------------------
  Obs. summary:      57341  left-censored observations at cc<=0
                     92706      uncensored observations
                         0  right-censored observations
```

Table B-7

```
Tobit regression                                 Number of obs   =     150047
                                                 LR chi2(5)      =    7807.73
                                                 Prob > chi2     =     0.0000
Log likelihood = -818575.17                      Pseudo R2       =     0.0047
```

```
------------------------------------------------------------------------------
          md |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   3613.914    65.5355    55.14   0.000     3485.466    3742.363
       mtxrt |   14005.11   1137.967    12.31   0.000     11774.71     16235.5
        mars |  -179.4062   159.5777    -1.12   0.261    -492.1752    133.3629
     famsize |  -342.5032   58.53501    -5.85   0.000    -457.2306   -227.7757
       agede |  -3636.234   152.2393   -23.88   0.000     -3934.62   -3337.848
       _cons |   -47874.9   626.8397   -76.38   0.000    -49103.49    -46646.3
-------------+----------------------------------------------------------------
      /sigma |   14403.01   41.73344                      14321.22    14484.81
------------------------------------------------------------------------------
  Obs. summary:      88442  left-censored observations at md<=0
                     61605      uncensored observations
                         0  right-censored observations
```

Table B-8

```
Tobit regression                                Number of obs   =     150047
                                                LR chi2(5)      =     2387.13
                                                Prob > chi2     =      0.0000
Log likelihood = -295703.63                     Pseudo R2       =      0.0040


------------------------------------------------------------------------------
        omd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
   lnincome |   558.0362   96.25965     5.80   0.000     369.3692    746.7032
      mtxrt |   51708.55   2351.649    21.99   0.000     47099.37    56317.74
       mars |  -4748.166   342.9781   -13.84   0.000    -5420.396   -4075.936
    famsize |  -303.6491   128.9181    -2.36   0.019     -556.326   -50.97218
      agede |  -10758.75   401.0123   -26.83   0.000    -11544.72   -9972.771
      _cons |  -34781.24   890.1069   -39.08   0.000    -36525.83   -33036.65
------------+-----------------------------------------------------------------
     /sigma |   23027.41   125.8092                      22780.82    23273.99
------------------------------------------------------------------------------
  Obs. summary:     131400   left-censored observations at omd<=0
                     18647      uncensored observations
                         0 right-censored observations
```


Table B-9A

```
Logistic regression                             Number of obs   =      62315
                                                Wald chi2(5)    =     627.57
                                                Prob > chi2     =     0.0000
Log pseudolikelihood =  -18126369               Pseudo R2       =     0.0601


------------------------------------------------------------------------------
            |               Robust
     hascce |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
   lnincome |   .2369502   .0230701    10.27   0.000     .1917336    .2821667
      mtxrt |   8.742061   .5791367    15.09   0.000     7.606974    9.877148
       mars |  -.2147757   .0633642    -3.39   0.001    -.3389673   -.0905841
    famsize |    .001337   .0241872     0.06   0.956    -.0460691    .0487431
      agede |  -1.476587   .1691629    -8.73   0.000     -1.80814   -1.145034
      _cons |   -5.76535   .2334774   -24.69   0.000    -6.222957   -5.307742
------------------------------------------------------------------------------
```


Table B-9B

```
      Source |       SS       df       MS              Number of obs =    5727
-------------+------------------------------           F(  5,  5721) =   18.39
       Model |  74.1457437      5  14.8291487           Prob > F      =  0.0000
    Residual |   4612.4588   5721  .806232966           R-squared     =  0.0158
-------------+------------------------------           Adj R-squared =  0.0150
       Total |  4686.60454   5726  .818477915           Root MSE      =   .8979


------------------------------------------------------------------------------
      lncce |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
   lnincome |   .0830192   .0202218     4.11   0.000     .0433768    .1226616
      mtxrt |   .8354984   .3289696     2.54   0.011     .1905933    1.480403
       mars |  -.3111789   .0362475    -8.58   0.000    -.3822377   -.2401201
    famsize |   .0565548   .0145591     3.88   0.000     .0280135    .0850961
      agede |  -.2624056   .1289049    -2.04   0.042     -.515108   -.0097031
      _cons |   6.480914   .1954486    33.16   0.000      6.09776    6.864067
------------------------------------------------------------------------------
```

Appendix C – Estimation Results from the SCF

Table C-1A

```
      Source |       SS          df       MS              Number of obs =   16370
-------------+------------------------------              F(  4, 16365) = 1200.70
       Model |  4005.57403        4  1001.39351           Prob > F      = 0.0000
    Residual |  13648.4973    16365  .834005332           R-squared     = 0.2269
-------------+------------------------------              Adj R-squared = 0.2267
       Total |  17654.0713    16369   1.0785064           Root MSE      = .91324

-------------+----------------------------------------------------------------
     LNVALUE |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |    .006494   .0004937    13.15   0.000     .0055264    .0074617
     FAMSIZE |   .0170197   .0068144     2.50   0.013     .0036627    .0303767
     MARRIED |   .1150217   .0190411     6.04   0.000      .077699    .1523444
    LNINCOME |   .4372681   .0071581    61.09   0.000     .4232374    .4512988
       _cons |   6.753875   .0850365    79.42   0.000     6.587194    6.920556
```

Table C-1B

```
      Source |       SS          df       MS              Number of obs =   16192
-------------+------------------------------              F(  4, 16187) = 2283.27
       Model |  640.581382        4  160.145345           Prob > F      = 0.0000
    Residual |  1135.33515    16187  .070138701           R-squared     = 0.3607
-------------+------------------------------              Adj R-squared = 0.3605
       Total |  1775.91653    16191  .109685413           Root MSE      = .26484

-------------+----------------------------------------------------------------
       RATIO |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |  -.0115935   .0001439   -80.55   0.000    -.0118756   -.0113114
     FAMSIZE |   .0138109   .0019829     6.96   0.000     .0099241    .0176977
     MARRIED |  -.0336637   .0055484    -6.07   0.000    -.0445392   -.0227883
    LNINCOME |   .0163805   .0020835     7.86   0.000     .0122966    .0204644
       _cons |   .8048336   .0247734    32.49   0.000      .756275    .8533921
-------------+----------------------------------------------------------------
```

Table C-2A

```
Logistic regression                                  Number of obs   =     30075
                                                     Wald chi2(4)    =   1228.23
                                                     Prob > chi2     =    0.0000
Log pseudolikelihood =  -56201379                    Pseudo R2       =    0.0617

-------------------------------------------------------------------------------
             |               Robust
      haslicv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |    .341101   .0183904    18.55   0.000     .3050565    .3771456
     married |   .2100427   .0401739     5.23   0.000     .1313032    .2887822
        agep |   .0438124    .006962     6.29   0.000     .0301671    .0574576
      agesqr |    -.00011   .0000611    -1.80   0.072    -.0002298    9.79e-06
       _cons |  -4.939295   .1888383   -26.16   0.000    -5.309412   -4.569179
-------------------------------------------------------------------------------
```

Table C-2B

```
      Source |       SS          df       MS              Number of obs =    6656
-------------+------------------------------              F(  4,  6651) =  410.43
       Model |  4670.11231        4  1167.52808           Prob > F      = 0.0000
    Residual |  18919.7215     6651  2.84464313           R-squared     = 0.1980
-------------+------------------------------              Adj R-squared = 0.1975
       Total |  23589.8338     6655  3.54467825           Root MSE      = 1.6866
```

```
------------------------------------------------------------------------------
      lnlicv |      Coef.   Std. Err.      t     P>|t|    [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   .7454299    .024682     30.20   0.000     .6970453    .7938145
     married |   .0571546   .0496082      1.15   0.249    -.0400934    .1544027
        agep |   .0731039   .0079085      9.24   0.000     .0576008     .088607
      agesqr |  -.0004259   .0000691     -6.16   0.000    -.0005614   -.0002903
       _cons |   2.933293   .2172761     13.50   0.000     2.507362    3.359224
------------------------------------------------------------------------------
```

Table C-3A

```
Logistic regression                             Number of obs  =      30075
                                                Wald chi2(4)   =    3134.98
                                                Prob > chi2    =     0.0000
Log pseudolikelihood = -66091834                Pseudo R2      =     0.1318


------------------------------------------------------------------------------
             |              Robust
    hasdbplan |      Coef.   Std. Err.      z     P>|z|    [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   .4552265   .0172949     26.32   0.000     .4213292    .4891238
     married |   .3065755    .036033      8.51   0.000     .2359521    .3771989
        agep |   .0524427   .0065932      7.95   0.000     .0395203    .0653651
      agesqr |   -.000023   .0000583     -0.40   0.693    -.0001373    .0000912
       _cons |  -5.523435   .1780742    -31.02   0.000    -5.872454   -5.174416
------------------------------------------------------------------------------
```

Table C-3B

```
      Source |       SS       df       MS              Number of obs =     8546
-------------+------------------------------           F(  4,  8541) =    61.02
       Model |  342.591012     4   85.647753           Prob > F      =   0.0000
    Residual |  11987.2598  8541  1.40349606           R-squared     =   0.0278
-------------+------------------------------           Adj R-squared =   0.0273
       Total |  12329.8508  8545  1.44293164           Root MSE      =   1.1847


------------------------------------------------------------------------------
     lnplanb |      Coef.   Std. Err.      t     P>|t|    [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   .1762792   .0186353      9.46   0.000     .1397495     .212809
     married |   .0276782   .0307059      0.90   0.367    -.0325129    .0878692
        agep |    .032099   .0051343      6.25   0.000     .0220345    .0421636
      agesqr |  -.0002969    .000044     -6.75   0.000    -.0003832   -.0002107
       _cons |  -1.182855   .1508313     -7.84   0.000    -1.478521   -.8871889
------------------------------------------------------------------------------
```

Table C-4A

```
Logistic regression                             Number of obs  =      30075
                                                Wald chi2(4)   =    3683.25
                                                Prob > chi2    =     0.0000
Log pseudolikelihood = -59742568                Pseudo R2      =     0.2215


------------------------------------------------------------------------------
             |              Robust
    hasdcplan |      Coef.   Std. Err.      z     P>|z|    [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnincome |   1.178294   .0232223     50.74   0.000     1.132779    1.223809
     married |  -.0361422   .0372436     -0.97   0.332    -.1091382    .0368539
        agep |   .1354681   .0080484     16.83   0.000     .1196937    .1512426
      agesqr |  -.0017106   .0000831    -20.57   0.000    -.0018736   -.0015477
       _cons |  -7.628465   .1966028    -38.80   0.000      -8.0138   -7.243131
------------------------------------------------------------------------------
```

Table C-4B

```
      Source |       SS       df       MS              Number of obs =   10542
-------------+------------------------------           F(  4, 10537) =  180.73
       Model |  1190.00431        4  297.501077         Prob > F      =  0.0000
    Residual |  17344.8332    10537  1.64608838         R-squared     =  0.0642
-------------+------------------------------           Adj R-squared =  0.0638
       Total |  18534.8375    10541  1.75835666         Root MSE      =   1.283


------------------------------------------------------------------------------
      lnplanb |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnincome |   .1838149   .0173555     10.59   0.000     .1497948     .217835
      married |   -.016566   .0303154     -0.55   0.585      -.07599     .042858
         agep |   -.075218    .006242    -12.05   0.000    -.0874535   -.0629825
       agesqr |   .0009828   .0000643     15.28   0.000     .0008567    .0011089
        _cons |   .8295436   .1449176      5.72   0.000     .5454776     1.11361
------------------------------------------------------------------------------
```

Table C-5

```
      Source |       SS       df       MS              Number of obs =   30075
-------------+------------------------------           F(  4, 30070) = 6723.34
       Model |  82099.5277        4  20524.8819         Prob > F      =  0.0000
    Residual |  91797.0711    30070  3.05277922         R-squared     =  0.4721
-------------+------------------------------           Adj R-squared =  0.4720
       Total |  173896.599    30074  5.78229031         Root MSE      =   1.7472


------------------------------------------------------------------------------
     lnwealth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnincome |   1.387882   .0121035    114.67   0.000     1.364158    1.411605
      married |   .1642545   .0232151      7.08   0.000     .1187519    .2097571
         agep |   .0450189   .0034468     13.06   0.000     .0382631    .0517747
       agesqr |   .0000562   .0000321      1.75   0.080    -6.70e-06    .0001191
        _cons |  -3.940298   .0853743    -46.15   0.000    -4.107635   -3.772961
------------------------------------------------------------------------------
```

51

Appendix D – Estimation Results from the MEPS

Table D-1

Estimation Results for Total Health Care Spending [a/]

(Weighted Least Squares)

Variable	Coefficient	Standard Error	T-Ratio
Female	0.2634	0.0173	15.22
Uninsured	-0.6769	0.0386	-17.55
HasSomePrivateCoverage	0.0382	0.0278	1.38
Age	0.0193	0.0014	13.85
Age Squared	0.0001	0.0000	5.53
Northeast	0.0601	0.0267	2.25
Midwest	0.1139	0.0254	4.48
South	0.0291	0.0232	1.25
Poor Health	0.9288	0.0289	32.18
Family Income-Poor	-0.2569	0.0334	-7.69
Family Income-Near Poor	-0.1671	0.0462	-3.62
Family Income-Low Income	-0.1824	0.0287	-6.35
Family Income-Middle Income	-0.1154	0.0207	-5.57
Intercept	5.6903	0.0394	144.47
Adjusted R-Square	0.234		
F-Value	660.540		

[a/] Dependent variable is the natural logarithm of total health care expenses for those individuals reporting positive expenditures. Family income designations relate to the Federal Poverty Level.
Source: 2001 Medical Expenditure Survey (MEPS), Full-year Consolidated File

Appendix E – Description of Variables Used in the Imputations

Variable Name	Description
INCOME	Total Income (CPS Definition)
HASGAINS	Binary Variable Equal to 1 if Taxpayer Reports Capital Gains
CGAINS	Capital Gains Reported on 1040
MARS	Marital Status (1=Married Filing Joint, 0=Otherwise)
LNGAINS	Natural Logarithm of Capital Gains
LNINCOME	Natural Logarithm of Income
INTST	Interest Income
DBE	Dividend Income (Before Exclusion)
LNINTST	Natural Logarithm of Interest Income
LNDBE	Natural Logarithm of Dividend Income
DIVSHR	Share of Income Attributable to Dividend Income
INTSHR	Share of Income Attributable to Interest Income
MTXRT	Marginal Tax Rate
PENSIONS	Pension Income
LNPENSIONS	Natural Logarithm of Pension Income
HASSEP	Has a Self-Employed Retirement Plan
LNSEP	Natural Logarithm of SEP Contribution
HASSEHEALTH	Has Deduction for Self-Employed Health Plan
LNSEHEALTH	Natural Logarithm of Self-employed health deduction
HASIRA	Has deduction for an Individual Retirement Account (IRA)
LNIRA	Natural Logarithm of IRA deduction
HASSTLOANINT	Reports a deduction for Student Loan Interest
LNSTLOANINT	Natural Logarithm of deduction for Student Loan Interest
CC	Amount of Itemized Deductions for Charitable Contributions
MD	Amount of Miscellaneous Itemized Deductions (Unlimited)
OMD	Amount of Other Miscellaneous Itemized Deductions (Limited)
HASCCE	Reports deduction for Child and Dependent Care Expenses
LNCCE	Natural Logarithm of Child and Dependent Care Expenses
LNVALUE	Natural Logarithm of Home Value
RATIO	Ratio of Home Equity to Home Value
HASLICV	Has cash value life insurance
LNLICV	Natural Logarithm of the amount of cash value life insurance
HASDBPLAN	Is covered by a Defined Benefit Pension plan
LNPLANB	Natural Logarithm of DB plan balance
HASDCPLAN	Is covered by a Defined Contribution Pension plan
LNPLANC	Natural Logarithm of DC plan balance
LNWEALTH	Natural Logarithm of Total Net Worth
FAMILYSIZE	Family Size
AGEDE	Number of Aged/Blind Exemptions