TECHNICAL APPENDIX: BUILDING THE METROTAX MODEL

This document provides an overview of the data and methods used to develop MetroTax, a federal income tax model created by the Brookings Metropolitan Policy Program. Using detailed microdata from the 2005 American Community Survey (ACS), the model is a static simulation used to estimate federal individual income tax filers, income tax liability, and eligibility for tax benefits such as the Earned Income Tax Credit (EITC), both under current law and under proposed expansions of the credit.

Data: About the ACS

The ACS is an annual national survey conducted by the U.S. Census Bureau encompassing roughly 3 million residents in 1.3 million households, or approximately 1 percent of the population.[1] It provides data on a number of socio-economic and demographic characteristics, such as age, marital status, family composition, and income by source.

The 2005 ACS was conducted on a monthly basis over the 2005 calendar year, but includes data that cover a 24-month period. This is due to the fact that respondents report items such as personal income and wages for the previous 12 months. For example, a respondent filling out the survey in March 2005 would report income stretching back to March 2004. To account for the differences and make dollar values comparable, the U.S. Census Bureau uses the Consumer Price Index (CPI) to adjust income items based on the month the survey was conducted to normalize dollar values to the 2005 calendar year.[2]

The MetroTax model makes use of a detailed subset of the 2005 ACS: the ACS Public Use Microdata Sample (PUMS). The PUMS presents the full range of responses reported on the individual questionnaire for a sample of housing units.[3] The 2005 ACS PUMS was selected using a systematic sampling method within each state and the District of Columbia, and provides a one-percent sample of the housing unit universe.[4]

There are two record types within the PUMS: the housing unit record and the person record (i.e., records for each individual within the housing unit). All identifying information is removed from the PUMS for the sake of confidentiality, and a serial number is assigned to link the person records to the appropriate housing unit. For the 2005 ACS, the PUMS contains 1,259,653 housing unit records and 2,913,796 person records.[5] The MetroTax model makes use of both types of records to identify household and individual characteristics relevant to defining tax units, filing status, and eligibility for the EITC. Moreover, the large sample size of the 2005 ACS PUMS affords an advantage over smaller surveys like the Current Population Survey, and permits us to examine sub-national geographies including cities, counties, metropolitan areas, and states.

Geographic Units of Analysis

The ACS identifies the state in which each PUMS record is located, but, for the sake of confidentiality, traditional sub-state geographies such as census tracts, cities, or counties are not identified. Rather, all records in the ACS PUMS are assigned to a Public Use Microdata Area (PUMA). Each PUMA contains a minimum population of 100,000, and, as a further precaution to protect the identities of ACS respondents, PUMA boundaries rarely conform to a single county or city.[6] Thus, to present data at the city and metropolitan levels, we first match individual PUMAs to these geographies where possible.

At the city level, we use information provided by the Integrated Public Use Microdata Series (IPUMS) to identify PUMAs that correspond with large cities in major metropolitan areas across the country.[7] To verify that the IPUMS-identified PUMAs align relatively closely with the city boundaries, we compared the 2005 population counts for the aggregated PUMAs against the total population of the city in that year. All 20 cities for which we report data have PUMA population counts within 5 percent of the city totals in 2005.

To aggregate PUMS data to the metropolitan level, we first use GIS (Geographic Information Systems) mapping software to select the PUMAs that most closely align with official metropolitan statistical area boundaries.[8] Next, we aggregate PUMAs to create the PUMA-equivalent of the 100 largest metropolitan statistical areas.[9] We then compare 2005 population totals for the aggregated PUMAs and the metropolitan areas. For the 100 largest metropolitan areas, population counts for the aggregated PUMAs are within 10 percent of the metropolitan area population counts in 2005.

Identifying Tax Units and Filing Status

Building on work performed at the Census Bureau and at the Urban-Brookings Tax Policy Center, the MetroTax model uses 2005 ACS PUMS data and tax year 2005 individual income tax filing rules and requirements to delineate tax units, identify probable dependents, and assign filing status.[10]

The fundamental unit of taxation in the U.S. federal income tax code is the family. Yet multiple tax units can be found within both households and families. Therefore, the model first uses ACS PUMS data from both the housing unit and person records to examine household relationships to the survey respondent (e.g., spouse, child, parent, other relative, roommate) and establish marital status. These relationship characteristics are used to identify families, related subfamilies, and unrelated individuals within each household and to identify potential tax filing units.[11]

After defining the relationships of household residents, the model uses data on age, school enrollment, disability status, and income to delineate additional potential tax filing units and identify and allocate dependents.

Dependents can include a range of individuals, such as children under 18 with little or no income, the severely or permanently disabled, or individuals over 65 with fixed incomes. The presence of dependents within a tax unit may affect filing status as well as the deductions, exemptions, and credits available to the tax filer.

After identifying all potential tax filing units and allocating any dependents within the household, the model assigns each tax unit a filing status as follows:

- *Non-filer*—potential tax filing units with incomes below the tax year 2005 IRS income thresholds for which filing a tax return is required

- *Married couple filing jointly*—tax units where both spouses are present in the household and the reported joint income amount requires them to file a return in tax year 2005 (Note: this is a simplifying assumption. All married couples living together are assumed to file jointly. The model does not account for cohabiting married couples who may choose to file separately.)

- *Head of household*—tax units where the tax filer is unmarried or legally separated, resides with one or more identified dependents, and has an income that requires him/her to file in tax year 2005

- *Married filing separately*—tax units where either: (a) the spouse is absent; or (b) the couple is separated and does not have any dependents; and where reported income requires the tax unit to file in tax year 2005

- *Single*—unmarried tax filers without dependents and with incomes that require them to file a tax return in tax year 2005.

Identifying the Undocumented

Before estimating eligibility for the EITC, the model takes an additional step to identify tax units that likely contain undocumented immigrants. Non-citizen low-income earners and their dependents must be legal immigrants with Social Security numbers valid for employment in order to be eligible for the credit. To determine citizenship and documented status, MetroTax employs a residual method devised by Jeff Passel of the Pew Hispanic Center.[12]

Our interpretation of Passel's algorithm begins by first identifying non-citizens based on country of birth, citizenship, and year of entry into the United States. In most cases, to become a naturalized citizen, a foreign-born individual must be at least 18 years old and reside in the United States for at least five years. The model assumes that foreign-born residents who entered the country in or before 1980 are here legally. For those who entered after 1980, the model assigns the individual one of 21 different immigrant statuses based, in part, on their country of origin. These categories include naturalized citizens, refugees,

married couples where at least one spouse has an authorized status, H1-B visa holders (e.g., hi-tech workers), au pairs, and those with green cards or legal permanent status. Refugees are distinguished by country of origin and particular periods of entry. For example, immigrants from the former Soviet Union who entered the country from 1980 to 1982 and from 1987 to 1997, if not otherwise indicated, are more than likely in the country legally and most likely arrived as refugees.

Characteristics such as occupation and school enrollment status are also used to determine legal status of the foreign-born population. For example, individuals and families who receive public assistance (e.g., food stamps, rent vouchers) or have indicated service in the U.S. Armed Forces are likely legal residents. High school and international university exchange students are identified by age, educational attainment, and school enrollment status. Foreign-born professionals such as doctors, nurses, professors, and software engineers are assumed to have authorized status.

Establishing Federal, State, and Local EITC Eligibility

Once tax units have been established and undocumented tax units have been excluded, the MetroTax model assesses eligibility for the federal EITC using tax year 2005 IRS guidelines (Table A).

TABLE A.
Structure of the EITC by Number of Qualifying Children, Tax Year 2005[13]

Tax Year 2005	Credit rate (percent)	Minimum income for maximum credit	Maximum credit	Phaseout rate (percent)	Phase-out range*	
					Beginning income	Ending income
No children	7.65	5,220	399	7.65	6,530	11,750
One child	34	7,830	2,662	15.98	14,370	31,030
Two children	40	11,000	4,400	21.06	14,370	35,263

 *These income levels represent the phase-out range for unmarried filers. In tax year 2005, the values for married couples filing jointly were $2,000 higher over the phase-out range.

Tax filers with qualifying children are eligible for the largest credits. To be deemed a qualifying child for purposes of claiming the EITC, the child must be the son, daughter, stepchild, or foster child of the claimant (or a descendant of one of these relations); or the child must be the brother, sister, stepbrother, or stepsister of the claimant (or a descendant of any of them). In addition, the child must be one of the following: (a) under the age of 19; (b) under the age of 24 and a student; or (c) totally and permanently disabled.

Note that filers must also meet a residency test to qualify for the federal EITC; children must live with the filer for at least half of the year in order for the filer to claim the child. This information is not available through the ACS. Using the PUMS data, we can only determine where the child was living at the time the survey was taken. Therefore, the model assumes the parent or guardian with whom the child was living at the time of the survey fulfills the residency requirement.

Using both household and person records from the ACS PUMS, the model assesses relationship, age, disability, and school enrollment data within each tax filing unit to determine the number of qualifying children. Then, based on the number of qualifying children present in the tax unit and the tax filing unit's earned income for tax year 2005, the model determines EITC eligibility and the total credit amount for which the tax filer is eligible. Table B presents the final MetroTax counts for the tax filing and EITC-eligible universe as compared to total returns filed, and total returns receiving the EITC, as reported by the IRS Statistics of Income division (SOI) for tax year 2005.

TABLE B.
IRS Tax Returns Filed, and those Receiving EITC, versus MetroTax Tax Filing Units Identified, and those Eligible for EITC, Tax Year 2005

	IRS (SOI)	MetroTax		
	Tax Returns Filed	**Tax Units Identified**	**Upper Bound**	**Lower Bound**
Total	132,844,632	130,708,116	130,932,767	130,483,465
EITC	22,053,667	23,154,501	23,275,550	23,033,452

Once federal EITC eligibility and credit amounts have been determined, the model next assesses eligibility for state-level credits. As of tax year 2008, 19 states and the District of Columbia offer refundable state EITCs that build on the federal version of the credit. In addition, New York City and Montgomery County, MD offer local versions of the credit. MetroTax uses eligibility parameters in each state and local jurisdiction to calculate the additional benefit available to eligible tax filers through these credits.[14]

Finally, to model the effects of various proposals to expand the federal EITC, we adjust the eligibility parameters used in the model as outlined in the paper and recalculate federal, state, and local eligibility and credit amounts.

Model Limitations and Areas for Further Research

This section notes a number of caveats to the analysis that should be taken into account when interpreting model results, and highlights areas for future research and model development.

Sampling and Non-Sampling Error

The ACS is a mostly self-reported survey, so populations can be underrepresented, overrepresented, or misrepresented, and incomes may be under-reported compared to what is reported on tax returns.[15] All results are subject as well to sampling and non-sampling error. Sampling error in the ACS results from the use of probability sampling to select the households that receive the questionnaire. In any sample, there is the chance the subset of the universe selected does not reflect the universe. Thus, ACS estimates include a measure of the standard error. Standard errors can be used to produce confidence intervals and estimate the margin of error, or the range in which the "true" estimate likely falls. In addition, as the PUMS is a sub-sample of the ACS sample, the PUMS is subject to additional sampling error, which can produce larger standard errors than would be obtained from the full ACS data set.[16] Table C presents examples at the national, state, and local levels of the 90-percent confidence intervals that accompany estimates derived using the MetroTax model.

TABLE C.
90-percent Confidence Intervals for Representative Estimates at National, State, and Metropolitan Levels, MetroTax Model, Tax Year 2005

Total Tax Units				
	Estimate	Upper Bound	Lower Bound	Margin of Error (+/-)
United States	130,708,116	130,932,767	130,483,465	224,651
Tennessee	2,626,060	2,654,578	2,597,542	28,518
Phoenix-Mesa-Scottsdale, AZ MSA	1,758,216	1,780,281	1,736,151	22,065
EITC-Eligible Tax Units				
	Estimate	Upper Bound	Lower Bound	Margin of Error (+/-)
United States	23,154,501	23,275,550	23,033,452	121,049
Tennessee	557,416	576,873	537,959	19,457
Phoenix-Mesa-Scottsdale, AZ MSA	279,329	292,397	266,261	13,061

Sample Weights

To adjust or "inflate" the PUMS sample counts to reflect the entire population, ACS provides a set of weights for each record type. In reality, most tax units fall somewhere between individuals and households; therefore, neither the housing unit nor the person weights provides the perfect mechanism for inflating the tax unit analysis to reflect the total universe. MetroTax applies the person weights, which may slightly over-estimate tax units, but future research should explore the feasibility of creating a hybrid weighting system that more closely reflects the tax unit level of analysis.

Data Availability

As noted previously, some data needed to affirmatively establish EITC eligibility, such as the presence of children in the home for more than 6 months during the tax year, and limitations on investment income, are not reported completely in the ACS. The ACS also lacks information necessary to estimate tax liability accurately. For instance, information on deductions from income for health and retirement contributions is not available in the ACS. Without this information the model's ability to establish tax liability and the use and benefit of non-refundable tax credits is limited. One result is that the current analysis makes no estimate of the effects of the refundable portion of the Child Tax Credit, claimed by many of the same families who benefit from the EITC. Future iterations of the model will attempt to adjust for these items using additional federal datasets such as IRS microdata, the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), and the Consumer Expenditure Survey (CES).

EITC Participation Rates

Estimates of EITC eligibility from the ACS model should not be compared to actual IRS EITC payments, at any level of geography, to calculate EITC participation rates. To calculate an EITC participation rate, one must match data that establish eligibility for the credit to data that identify whether those same eligible filers claimed the credit.[17] This model lacks the information necessary to identify which tax units actually filed a tax return, the number that claimed the EITC, or the number of units which may have claimed the credit in error. Each of these components is necessary to determine actual participation rates; therefore, the results of this model should only be interpreted as a first-order indicator of eligibility for the EITC.

Also note that, because the model is an eligibility model rather than one designed to predict program participation, actual increases in the number of filers receiving the EITC under the proposed expansions of the credit, and in EITC dollars received, will depend upon the rates at which eligible (and ineligible) workers and families participate in the program.

Impact of Policy Proposals

Because the MetroTax model is static, the estimates presented here reflect the effects of current law and proposals as they would have affected metropolitan tax filers in 2005. Changes in population and the economy since then, as well as changes in behavior that may be induced by changes in tax law, are not reflected in this analysis.

Conclusion

Despite these caveats, the size of the ACS PUMS sample and the breadth of information available through the survey mean that results derived from the MetroTax model offer a reasonably robust snapshot of probable tax filers and EITC eligibility at the metropolitan and state levels.[18] As future years of ACS microdata become available, MetroTax will be updated to reflect the most recent data, and, with each iteration of the model, any changes in tax law will also be incorporated. Future iterations of this research will seek to further refine and develop the MetroTax model to address the limitations highlighted above and to expand the capacity of the model to examine the impact of tax policies at the state and local level.[19] For detailed nationwide effects of these proposals, readers should refer to estimates generated by more complete, sophisticated models such as that operated by the Urban-Brookings Tax Policy Center.[20]

---

[1] For general background on the ACS, see **www.census.gov/acs/www**

[2] Kirby G. Posey, Edward Welniak, and Charles Nelson, "Income in the American Community Survey: Comparisons to Census 2000" (U.S. Census Bureau, 2003).

[3] For more information on the ACS PUMS, see **www.census.gov/acs/www/Products/PUMS**

[4] For more information, see "PUMS Accuracy of the Data (2005)" **www.census.gov/acs/www/Downloads/2005/AccuracyPUMS.pdf**

[5] Ibid.

[6] To find maps of PUMA boundaries within each state and the District of Columbia, see **www.census.gov/geo/www/maps/puma5pct.htm**

[7] For more information on the Minnesota Population Center's Integrated Public Use Microdata Series, see **http://usa.ipums.org/usa/**

[8] We use metropolitan statistical area boundaries as defined by the U.S. Office of Management and Budget in December of 2003. See **www.census.gov/population/estimates/metro-city/0312mcsa.txt**

[9] We identify the 100 largest metropolitan statistical areas based on 2005 employment figures. Employment totals come from the Bureau of Economic Analysis 2005 Wage and Salary Employment data.

[10] Amy O'Hara, "New Methods for Simulating CPS Taxes" (U.S. Census Bureau, 2004); Jeff Rohaly, Adam Carasso, and Mohammed Adeel Saleem, "The Urban-Brookings Tax Policy Center Microsimulation Model: Documentation and Methodology for Version (0304)" (Washington: Urban Institute, 2004); Internal Revenue Service, "Forms, Instructions, and Publications for TY 2005" **www.irs.gov/formspubs/article/0,,id=158421,00.html**

[11] Related subfamilies are defined by the U.S. Census Bureau as, "A married couple (with or without children) or a single parent with one or more never-married children under the age of 18, residing with and related to the householder, but not including the householder or the householder's spouse." For example, when a grown child moves back to their parental home with their own children or spouse, they qualify as a subfamily. See **http://factfinder.census.gov/home/en/epss/glossary_s.html**

[12] The routine to estimate foreign-born tax filers' legal status is based on work by Jeff Passel and others; see Jeff Passel, Jennifer Van Hook, and Frank Bean, "Estimates of the Legal and Unauthorized Foreign-Born Population for the United States and Selected States, based on Census 2000 (Sabresystems Inc., 2004).

[13] See Tax Policy Center, "Historical EITC Parameters" **www.taxpolicycenter.org/taxfacts/displayafact.cfm?Docid=36**

[14] For state-level EITC eligibility parameters, see **www.taxpolicycenter.org/taxfacts/displayafact.cfm?Docid=293**

[15] Posey, Welniak, and Nelson, "Income in the American Community Survey."

[16] For more information, see "PUMS Accuracy of the Data (2005)."

[17] Alan Berube, "Earned Income Credit Participation—What We (Don't) Know" (Washington: Brookings Institution, 2005).

[18] For instance, the correlations at the metropolitan level between tax filers and EITC-eligible filers estimated by the MetroTax model, and tax returns and EITC recipients reported by the IRS, are 0.999.

[19] With the release of this paper, the MetroTax model SAS code is available under the GNU General Public License (please contact David Park at **dpark@brookings.edu** for more information). Our hope is to build an open source collaborative community around MetroTax and related work and its use as a policy analysis tool at the metropolitan, state, and national levels.

[20] See Tax Policy Center, "Overview of the Tax Policy Center Microsimulation Model" (2007).