APPENDIX K

Methodology for Analysis of Household Water and Wastewater Expenditures

Every ten years the U.S. Census Bureau conducts a census, for which every household (housing unit) in the nation is asked to complete a questionnaire. A randomly selected sample of one in six housing units receives *Form D-2*, a more detailed questionnaire referred to as "the long form." This questionnaire collects additional economic data, including household and personal income and expenditures. In the 2000 Census, question 45 asked what the annual costs of (expenditures on) different utilities and fuels were for the housing unit (house, apartment, or mobile home) in 1999. Water and sewer services combined were addressed in part "c" of the question.[195] For this part, the respondents could record an amount rounded to the nearest dollar, check an option stating that water and sewer service costs were included in their rent, or check an option stating that there were no charges to the housing unit for water and sewer services in 1999.

The Census Bureau does not make the raw data collected from the questionnaires available to the general public. However, in the Public Use Microdata Samples (PUMS), it does provide data from a stratified, random sample of housing units that responded to the long form.[196] Hence these samples contain records for a subsample of housing units on the characteristics of each unit and each person in it, and each microdata file is a stratified sample of the

[195] U.S. Census Bureau, *Census 2000 Form D-2* (Washington, D.C.: the Bureau, 2000), available as appendix D in *Public Use Microdata Sample 2000 Technical Data*, at www.census.gov/prod/cen2000/ doc/pums.pdf.

[196] Available from the Census Bureau at ftp://ftp2.census.gov/census_2000/datasets/PUMS/.

population that was created by subsampling the one-in-six sample of housing units that received the long form.[197] Housing-unit weights and person-level weights, used to indicate the number of households and people each respondent represents, are included for each record in the microdata samples.

Two versions of the microdata files are available: a 5 percent sample of all long-form respondents, from which the Census Bureau can create highly populated microdata files for small areas called Public Use Microdata Areas (PUMAs), and a 1 percent sample of all long-form respondents, from which the Census Bureau can create less populated microdata files for large areas called super–Public Use Microdata Areas (super–PUMAs). All states are split into super–PUMAs, which are split further into PUMAs. PUMAs and super–PUMAs never cross state boundaries.[198] Each PUMA is an area in the state that contains a minimum of 100,000 people. As a result of this threshold, PUMAs range in size from small parts of a metropolitan city to several contiguous counties in rural areas, depending on the location in the state. Super–PUMAs consist of one or more contiguous PUMAs, and they contain at least 400,000 people. Both the 1 percent and the 5 percent samples contain data on the level of the housing unit for all of a state's super–PUMAs, whereas only the 5 percent samples contain data on the level of the housing unit for the state's PUMAs. Nationwide the 5 percent sample files contain records for more than 14 million people and more than 5 million housing units. The 1 percent sample files, due to the lower sampling rate of the long form respondents, contain records only for more than

---

[197] U.S. Census Bureau, *Public Use Microdata Samples 2000 Technical Documentation* (Washington D.C.: the Bureau, 2000), available at www.census.gov/prod/cen2000/doc/pums.pdf.

[198] *Ibid.* Geographic Information System (GIS) shapefiles of PUMAs are available at www.census.gov/geo/www/cob/pu5_2000.html, and GIS shapefiles of super–PUMAS are available at www.census.gov/geo/www/cob/pu1_2000.html.

2.8 million people and more than 1 million housing units.[199] The 5 percent sample files contain a greater sample and provide the ability to conduct analysis at a smaller geographic region than the 1 percent sample files.

Methodology

The main power of the PUMS is that they give researchers the ability to analyze each housing unit's economic data separately and, using housing-unit weights appropriately, to produce regional estimates of expenditures and income that are not obtainable from the summaries produced by the Census Bureau. For this report, the University of North Carolina, Environmental Finance Center (UNCEFC) research team used STATA statistical software to analyze the data from the 5 percent microdata samples for the thirteen Appalachian states.[200] Using the dataset of housing unit level data, in which each record represents one household sampled for the 5 percent PUMS, six variables were retained:

- STATE: the state in which the housing unit is located, using the FIPS state code
- PUMA5: the PUMA in which the housing unit is located, using a state-level identifier
- PUMA1: the super-PUMA in which the housing unit is located, using a state-level identifier
- HWEIGHT: the weight indicating the number of housing units in the population represented by the record

---

[199] *Ibid.*

[200] For a description of the STATA software, visit www.stata.com. The microdata samples for the thirteen Appalachian state are available from ftp://ftp2.census.gov/census_2000/datasets/PUMS/FivePercent/.

- WATER: dollar payment for water and sewer services directly in 1999, or a code indicating the payment of these services through rent or no payment in 1999
- HINC: household income

Using the relationship between PUMAs and counties, the research team assigned each PUMA, and subsequently each housing unit, a dichotomous variable of 1 or 0 indicating whether or not any part of the PUMA was located inside the 410-county Appalachian area.[201] There are 699 PUMAs in the thirteen states; 184 are in Appalachia, including 28 that are partially in Appalachian counties and partially in non-Appalachian counties.[202]

To facilitate a comparison of the results of the present analysis with those of a similar national study that used a similar method, the research team dropped all households with less than $1,000 in income from the analysis.[203] The team assigned the remaining households to one of the following categories, on the basis of the coding of the WATER variable:[204]

- Households paying centralized systems directly for water and sewer services (records with an entry for WATER between 2 and 9,999)
- Households paying for water and sewer in their rent (records with an entry of 0 for WATER)

---

[201] Files showing the relationship between PUMAs and counties are available for each state at ftp://ftp2.census.gov/census_2000/datasets/PUMS/FivePercent/.

[202] Data from the 5 percent PUMAs for the thirteen states, compiled by UNCEFC.

[203] Scott J. Rubin, *The Cost of Water and Wastewater Service in the United States*, Rural Water Partnership Fund White Paper (Duncan, Okla.: National Rural Water Association, 2004). Rubin deleted households with less than $1,000 in income to focus the analysis on households with positive incomes and positive expenditures

.

[204] U.S. Census Bureau, *Public Use Microdata Samples 2000 Technical Documentation*, 7–33.

- Households that did not have a charge for water and sewer in 1999 (records with an entry of 1 for WATER)

Vacant housing units and group quarters were given a missing value for WATER by the Census Bureau in the microdata samples. The UNCEFC research team dropped these records before further analysis.

Using the housing-unit weights, the research team determined the total number and the proportions of housing units not paying for water and sewer services, paying for them directly and paying for them through rent, for all housing units in each of the thirteen Appalachian states as a whole, as well as in their Appalachian and non-Appalachian regions. For housing units paying directly for water and sewer services, the percentage of household income spent on these services in 1999 was calculated by dividing the cost of water and sewer services by the household income. Using the housing-unit weights again, the team determined the mean, the median, the standard deviation, the minimum and maximum cost of and percentage of household income spent on water and sewer services for each PUMA, for the Appalachian and non-Appalachian regions of each state, for each state as a whole, and for the entire Appalachian region.

Finally, the research team assigned households that paid directly for water and sewer services two dichotomous variables according to whether or not they spent more than 2.5 percent and 5 percent of their income on water and sewer services in 1999. The team then calculated the percentages of households that spent more than 2.5 percent and more than 5 percent of their income on water and sewer services for the Appalachian and non-Appalachian regions of each state, for each state as a whole, and for the entire Appalachian region.

The results of the analysis and their implications are discussed in chapter 6.