Synchronizing Survey Edit and Imputation Systems: A Look at the American Housing Survey's Utility System

Mark Gorsak and Dennis Schwanz, U.S. Census Bureau[1]
Mark Gorsak, U.S. Census Bureau, 4700 Silver Hill Road,Washington, D.C. 20233-1912

Key Words: Utilities ; American Housing Survey ; Psychology of Survey Response

Introduction

Surveys use various methods to control nonsampling error to improve the quality of the survey data. The Current Population Survey uses characteristics as age, race, sex, and origin in their population controls to correct bias among these demographic groups (West and Robinson, 2002). To reduce item nonresponse for financial variables such as income or household assets, surveys capture interval-censored observations (e.g., $10,000 - $14,999) whenever the respondent is unable or refuses to provide a precise response. Then a survey like the Health and Retirement Survey would apply multiple imputations methods (Heeringa, 1993).

The American Housing Survey (AHS) captures electric and natural gas costs from the respondent either by recording the monthly bills from specific months, (January, April, August, and December) or asking for a monthly average over the prior twelve months. The AHS uses data from the Residential Energy Consumption survey (RECS) and applies the findings from the RECS to the AHS utility data. If the respondent does not provide any utility data, then the AHS imputation system determines an annual utility cost by matching to the primary variables from the RECS findings. The application of the RECS data to the three methods in the AHS utility system creates common thread between the RECS and the AHS.

Overview

The AHS runs an edit for the utility system that uses information provided by the respondent multiplied by factors generated from the RECS. The first edit includes monthly utility bills from the respondent. Each amount is multiplied by a factor to calculate an annual cost. When the respondent does not provide any utility bills, the second edit uses an average monthly amount provided by the respondent. The average monthly amount along with other characteristics related to the housing unit are multiplied by factors to calculate an annual cost. It should be noted that when the respondent provides monthly bills, an average monthly amount from the respondent is not asked.

The questions about utilities place a burden on the respondent. By providing monthly bills, the respondent needs to maintain accurate records over a period of time. Keeping an electric or natural gas bill over six months old may not be important to the respondent. Historically, about 30% of the respondents provide electric or natural gas bills. When respondents do provide monthly utility bills, the majority of the respondents provide four months of bills.

When the respondent does not provide monthly utility bills, the AHS tries to capture an average monthly cost over the past 12 months. This type of question also places a burden on the respondent. The respondent needs to think

[1] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

about past utility bills. Theoretically, the respondent needs to add 12 bills and then divide the total by 12. Realistically, the respondent may remember one or two bills and then decide what to do with the amounts. By providing an average monthly amount, the respondent has a psychology burden during the time period of the question. Historically, about 50% or the respondents provide an average monthly amount. Though this method is the most popular among respondents, there is a greater psychology burden placed on the respondent by answering an average monthly amount than by giving four monthly bills.

The AHS utility edit utilizes other demographic and household information correlated with utility costs to address for any possible error occurring from the psychology of the survey response.

This paper outlines the AHS utility process by providing some background on the RECS and the AHS and then describing the response processes for the utility costs. The paper concludes with describing the different methodologies used in the edit of annual utility cost and some results from the 2003 AHS.

Residential Energy Consumption Survey

The RECS is a national statistical survey that collects energy-related data for occupied housing units. The RECS began in 1978 and had ten more enumerations, most recently in 2001. In the 2001 RECS, data were collected from about 5,000 housing units statistically selected to represent the 107.0 million housing units in the United States.

RECS data is available for the nine Census divisions. A Census division is the next geographic area smaller than a region and bigger than a state. The RECS provides information on the use of energy in residential housing units.

This information includes:
- the physical characteristics of the housing unit,
- the demographic characteristics of the household,
- the types of fuels used, and
- energy consumption and expenditures data on electricity and natural gas.

There are three different modes of collection for the RECS data.
- On-site 40-minute personal interviews conducted in the housing unit.
- Telephone interviews with rental agents of rented housing units that have any of their energy use included in their rent.
- questionnaires mailed to the housing units' energy suppliers asking them to provide the units' actual energy consumption amounts and expenditures

For detailed information about the RECS, go to the home page for the RECS under the Energy Information Administration web site, http://www.eia.doe.gov/emeu/recs/contents.html.

The American Housing Survey

The primary purpose of the AHS is to provide a current and ongoing series of information on the size and composition of the housing inventory and measure the type of changes. The AHS consists of interviews conducted every other year at a sample of housing units. The housing units are selected by scientific sampling methods from geographic areas throughout the Nation. There are approximately 55,000 housing units for the national survey (AHS-N) and 4,000 housing units per metropolitan area for the metropolitan survey (AHS-MS). The AHS-N is enumerated every other year. The last enumeration was in 2003. The AHS-MS is enumerated every four years in a metropolitan area.

Interviewers collect the AHS data using laptop computers by personal visits and telephone interviews from their homes. The interview lasts about 45 minutes.

Topics included in the AHS:
- household demographics
- housing modifications, needs, problems
- housing costs, including utilities,
- and energy usage.

The AHS incorporated the 1990 RECS beginning in 1993 and the 1997 RECS in 2003. The monthly billing data from the 2001 RECS was not available for the AHS in 2003.

For detailed information about the AHS, go to the U.S. Census Bureau home page for the American Housing Survey under the web site: http://www.census.gov/hhes/www/ahs.html

Utility Cost Response Process

There are five stages of the response process for a respondent: encoding, comprehension, information retrieval, judgment, communication (Biemer and Lyberg, 2003). The information retrieval stage varies for the respondent in the AHS. The respondent uses record look-up when applying a response to the monthly billing data question and recall with rough approximation when thinking about a monthly average. The amount of effort required for accurate recall would vary for the respondent. It would take several seconds to retrieve from memory twelve electric or gas bills. There may be use of short term and long term memory for a utility response. The memory may recover fragments of the information (Sudman, Bradburn, Schwarz, 1996). Then the respondent would use a rough approximation strategy for answering the question.. The determinants for selection are the motivation of the respondent and the accessibility of the episodic memory (Tourangeau, Rips, Rasinski, 2000). If the determinant tax the respondent too much, then the respondent could telescope information into the collection period and possibly think about the most salient bill from that period and provide that as an answer.

The 2001 AHS–N showed an unusually high number for monthly gas bills over $200 for respondents providing an average monthly cost. It is thought that the respondents only think about the winter months when responding and do not factor the low usage of gas during the summer months. This could support the notion that respondents recover a fragment of information and that fragment would be the expensive bills relative to the other bills. Higher gas usage occurs mainly during the winter months while the AHS-N interview period occurs during the summer months. Do the respondents consider the summer months when averaging gas bills?

From month to month, electric bills remain more constant than gas bills. The respondents could think back two or three months and then determine an average for electricity for that time period.

The information retrieval process changes for utility costs when using monthly billing data. The respondent relies on records like a checkbook or the actual bills. The record look-up would be less taxing on the respondent. The respondent has the ability to quickly search through their records and provide any needed utility data. The AHS notifies the respondent about the utility cost information via a letter prior to conducting the interview. In the AHS–N, when a respondent provides monthly bills, the respondent provides all four months about 90% of the time for both electricity and natural gas. There is no need to average the data when providing monthly billing data.

The respondent burden on utility costs from the RECS is non-existent. The utility cost data

from the RECS comes directly from the utility company. The RECS respondent provides consent in the survey to retrieve the utility cost from the utility company. One of the main reasons to model the AHS data after the RECS data is that the RECS utility cost data comes directly from the source providing the actual bill.

Methodology for Utilities in the AHS

There are three methods to generate an electric or natural gas annual cost in the AHS.

1. Monthly Billing Data
2. Average Monthly Cost
3. Impute for Non-Response

There are limits for the utility estimates. Estimates generated within the AHS are subject to a floor and ceiling limit based on Census division. These floor and ceiling limits were established using the RECS data and then adjusted by the inflation factor.

An inflation factor for the household characteristics adjusts for the difference between years when generating the coefficients from the RECS and the enumeration year of the AHS.

After generating all the utility costs in the AHS, there is an adjustment factor using the averages by Census divisions from the RECS. This adjustment allows the estimates at an aggregate level from the AHS to be comparable to the estimates from the RECS.

Monthly Billing Data

The 1993 AHS-N used the 1990 RECS to generate regression coefficients to estimate the utility costs based on the monthly billing data. The 2003 AHS-N updated those terms using the 1997 RECS.

For respondents providing two or more months of utility costs, the AHS determines an annual cost estimate based solely on the billing data by using regression terms derived from the RECS. Basically, the annual cost estimate is a weighted average of the months inputted into the model. Each Census division has coefficient terms for each month.

There are two exceptions to the weighted average solely based on two or more months of utility data. The annual estimates from the January and December electric data and the April and August natural gas data also utilize household characteristics. These household characteristics are the same as those used for one billing month and average monthly cost methods. The research for electricity estimates showed that billing data from two seasons help estimate the annual electricity bill. The December and January months only cover the winter season. The research for natural gas estimates showed that billing data from the winter season is essential in estimating the annual natural gas bill. The April and August months do not cover the winter season.

Households providing only one billing month use eight household characteristics along with the monthly bill to estimate the annual utility cost. The household characteristics include home heating fuel source, water heating fuel source, type of structure, the year the structure was built, and the number of rooms, bathrooms, appliances, and household members. Appliances include refrigerator, an oven, a cooking range, a dishwasher, a washing machine and a clothes dryer.

Again, there are separate regression terms by Census division. These terms use an inflation factor to adjust for the difference between the year of the source data and the year of enumeration.

Average Monthly Cost

Prior to the 2003 AHS-N, the average monthly billing method used a factor of 12 to estimate the annual utility cost.

Beginning with the 2003 AHS-N, the estimate derived from the average monthly bill provided by the respondent uses the nine household characteristics and a factor ranging from 6 to 8 with the average monthly bill.

The information retrieval to determine an average monthly cost is more demanding on the respondent. In a short amount of time, the respondent needs to think about the time period of the previous twelve months, twelve electric or gas bills, and then divide by 12. Questions could arise in the mind about seasonal effects or a salient bill or no bill at all. The methodology compensates for the demand on the response by applying data items less cognitively difficult on the respondent. The information retrieval is not as demanding when thinking about the home or water heating methods or the number of rooms or bathrooms in the house.

Impute for Non-Response

Prior to the 2003 AHS-N, the imputation method used the nearest neighbor method based on tenure and number of bedrooms. The nearest neighbor method would search for processed records containing the same tenure value and number of bedrooms.

Research from the 1990 and the 1997 RECS data showed that Census division along with the eight household characteristics to be highly correlated with utility costs. Due to this stability, the AHS incorporated three household characteristics beginning with the 2003 AHS-N into the imputation method when determining the annual utility costs.

The imputation method in the 2003 AHS-N uses the nearest neighbor method matching on three household characteristics within a Census division: The household characteristics are home heating fuel source, water heating fuel source, and the total number of rooms. The fuel sources and the number of rooms are highly correlated with utility costs. Also, the number of rooms provides demarcation between small and large housing units. There are a total of 180 cells for each utility in the imputation method. There are nine Census divisions, electric or natural gas home heating, electric or natural gas water heating, and five categories for total rooms, less than 4, 4, 5, 6, and 7 or more rooms.

The nearest neighbor method would search for processed records containing the same values for these characteristics. If there are no processed records, then the unit would receive the median value from the previous enumeration adjusted for inflation.

Results from the 2003 AHS-N

Table 1 shows the breakdown of housing units using each method when calculating electric and gas utility costs for the 2003 AHS-N.

Table1. Number and Percentage of Records by Utility Cost Method

Utility Cost Method	Electric		Gas	
Monthly Billing	14528	**32**%	8590	**30**%
Four Months	13344	29%	7630	27%
Three Months	435	1%	501	2%
Two Months	428	1%	282	1%
One Month	321	1%	177	>1%
Avg. Monthly Billing	24402	**53**%	12820	**45**%
Imputation	6960	**15**%	6940	**25**%
Total	45890		28350	

The average monthly billing method is the most common method for determining electric and gas utilities, 53% and 45%, respectively. This method is the most demanding on the

respondent cognitively.

The monthly billing method shows that if respondents provide their utility information, then they give four months of costs the vast majority of the time, 92% for electricity and 89% for gas. Since information retrieval through record look-up is not demanding on the respondents, the respondents, when motivated to use records, supply the months requested.

Utility cost estimates from the imputation method occur 15% of the time for electricity and 25% for gas. An interview period during the summer months probably plays an important role when collecting gas data. Since gas is primarily used during the winter months, the effort to recall this usage either through records or by memory recall during the summer is demanding on the respondent. The imputation method addresses the need for winter months by using home heating and the number of rooms within the household.

The utility costs should show that the more energy a household uses then there are higher utility costs for that household. Each method should show this trend. Using electricity as an example, there are home heating and water heating with electricity. There are four categories to consider for energy usage.

Cat 1. Home Heating and Water Heating
Cat 2. Home Heating Only
Cat 3. Water Heating Only
Cat 4.Neither Home Heating and Water Heating

Tables 2 and 3 present the means for the four categories by Census division. Heating a home and water with electricity should cost more than not heating a home or water with electricity. Heating only a home or heating only water within a home with electricity should be between the other two categories. Tables 2 and 3 generally show this trend.

There is a difference between the means using the monthly billing records and the average monthly amount from the respondent categorized by home and water heating. This is primarily due to the adjustment factor that allows the aggregate level at the Census division for the AHS to equal the aggregate level at the Census division for the RECS.

The adjustment factor is only applied to the estimates using the average monthly amount and the imputation method. The adjustment factors range from 0.56 to 0.87 depending upon the Census division. Table 3 shows the adjustment factors. Records estimated with monthly billing records do not need an adjustment to the RECS. These records are believed to be factual since they use actual billing data; so, they do not necessitate an adjustment to the RECS. Records estimated by an average monthly cost from the respondent are more subjective.

There may be another cause to the differences between the two methods. People answering the utilities by looking up records may differ from people using an average monthly estimate. Further research will explore this possibility as well as an adjustment factor that incorporates the estimates from the monthly billing records.

Future Work

The AHS team is redesigning the housing cost portion of the instrument. We plan to introduce a new question to collect the most recent utility bill from the respondent. A response to a recent monthly bill question should be cognitively less difficult than a response to an average monthly bill question. It will also allow the respondent some more time to think about utility costs before receiving the average monthly billing question. There is an expectation the recent billing month question would increase the item response rate for utility cost and allow the respondent more time to develop a strategy in

responding to the average monthly billing question.

Acknowledgments

The authors would like to thank Nancy Mathiowetz and Roger Tourangeau for inspiring ideas and insightful comments when developing the methodology. The authors also appreciate the efforts and contributions in implementing the methodology.

References

Biemer, Paul P. and Lars E. Lyberg. (2003). *Introduction to Survey Quality*. pp 123 - 148.

Heeringa, S.G. (1993). "Imputation of Item Missing Data in the Health and Retirement Survey." Proceedings of the Survey Methods Section, American Statistical Association. Pp. 107-116.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. (1996) *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. pp 55 - 184.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. (2000). *The Psychology of Survey Response*. pp 136 - 164.

West, Kirsten K. and J. Gregory Robinson. (2002). "The Use of Independent Population Controls to Calibrate Surveys." Paper presented at the 2002 International Conference on Improving Surveys, Copenhagen, Denmark, August 25-28, 2002.

Table 2. Mean Costs for Electric Monthly Billing Method

CD	Cat. 1	Cat. 2	Cat. 3	Cat. 4
1	103	*	89	70
2	128	82	92	80
3	93	*	76	67
4	104	89	92	73
5	119	112	98	89
6	118	104	97	88
7	129	118	109	105
8	108	80	75	72
9	91	69	75	71

* The number has been suppressed because it is statistically unreliable.

Table 3. Mean Costs for the Average Monthly Bill for Electric After the Application of the Adjustment Factor

CD	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Adj. Factor
1	134	86	79	62	0.87
2	102	86	73	65	0.70
3	79	76	59	59	0.78
4	78	64	57	54	0.64
5	86	77	64	67	0.72
6	80	71	65	64	0.67
7	80	75	72	65	0.63
8	61	55	43	46	0.56
9	66	62	66	59	0.62