Evaluating and Redesigning Imputation Methodologies for the 2015 American Housing Survey[1]

George R. Carter III[1]
Brian Shaffer[2]
[1]U.S. Department of Housing and Urban Development,
451 7th Street SW, Washington, DC 20410
[2]U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Abstract
The American Housing Survey (AHS) is the largest nationally representative survey of housing in the United States. It is conducted every two years by the U.S. Census Bureau for the Department of Housing and Urban Development (HUD) and has followed the same sample of housing units since 1985. The AHS imputes some items for non-response using hot-deck, cold deck, and regression-based methods. The AHS is currently undergoing a survey redesign and will introduce a new sample in 2015. As part of the redesign, the Census Bureau is evaluating current imputation methodologies and designing new approaches to impute items in 2015 and beyond. Utility costs are a financial housing characteristic that is imputed in the AHS. Utility cost data are currently adjusted and imputed using a regression-based method that utilizes utility bill and housing characteristics from the AHS and housing characteristics, consumption, and cost data from the 2001 Residential Energy Consumption Survey (RECS) adjusted for inflation. This research re-estimates utility models with data from the 2009 RECS and explores the implications of estimating utility costs without AHS utility billing data.

Key Words: imputation, variance, cold-deck imputation, hot-deck imputation, regression-based imputation, longitudinal

1. Introduction

Utility costs are an important component of housing costs, which also include rent or mortgage payments; garbage, trash, water, and sewage costs; real estate taxes; and other housing-related fees. Accurate measurement of housing costs is crucial to monitoring trends in affordable housing supply over time. Housing costs are incorporated into measures of housing burden, which examine the percentage of a household's income spent on housing. Housing burdens are a component of the U.S. Department of Housing and Urban Development's (HUD) "Worst Case Needs" measure,[2] which is used to estimate the number of households in the United States with the greatest housing needs. Utility cost data collected in the American Housing Survey (AHS) are adjusted and imputed using

[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. Department of Housing and Urban Development or the U.S. Census Bureau.

[2] One criteria for Worst Case Needs is having "severe rent burden," which is defined as unassisted renters with very low incomes (incomes not more than 50 percent of area median income (AMI)), who are paying more than half their income on housing.

regression models developed from the Residential Energy Consumption Survey (RECS). As part of the redesign of the AHS for 2015, imputation methods, including those used to estimate utility costs, are being evaluated by Census and HUD. The research discussed in this paper explores the implications of estimating utility costs in the AHS in the absence of billing data provided by respondents. In the remainder of section 1, we provide a brief overview of the AHS and 2015 redesign efforts. In section 2, we discuss how utility data are collected in the AHS. In section 3, we discuss how AHS utility data are edited and adjusted using data from the RECS. In section 4, the study design and model development are explained. Results are presented in section 5 and conclusions and next steps are explored in section 6.

1.1 American Housing Survey

The AHS is a key source for housing and utility cost data in the United States. The AHS is the largest, most comprehensive, longitudinal housing survey in the United States and is collected by the U.S. Census Bureau on behalf of the U.S. Department of Housing and Urban Development (HUD). The AHS employs two types of housing unit samples: a national sample and metropolitan area samples, which are surveyed on a rotating basis. In both types of sample, the same housing units are followed over time until a new sample is collected. Between two and twenty-nine metropolitan area samples have been surveyed in the years the Metro AHS has been fielded, depending on the availability of funding. The National AHS started in 1973 and has surveyed the same housing units since 1985, drawing additional sample to account for new construction. From 1973 to 1981, the National AHS was conducted annually. Since 1983, the National AHS has collected data every 2 years, in odd-numbered years. In 2007 and 2009, national and metropolitan area samples were surveyed in the same year. Starting in 2011, metropolitan area samples were merged with the national sample and collected simultaneously. Since 1997, the AHS has been collected via in-person and telephone interviews using an electronic questionnaire.

1.2 2015 AHS Redesign

In 2015, the AHS will retire the national sample it has surveyed since 1985 and all metro samples and will draw a new sample from the Census Master Address File. Along with the introduction of a new sample, the AHS is undergoing a redesign for 2015, balancing the goals of increasing data quality and reducing respondent burden. As part of the 2015 AHS redesign, HUD and the U.S. Census Bureau are evaluating all of the questions in the current AHS questionnaire and methods for imputing for non-response. Related to the goal of reducing respondent burden, the research discussed in this paper explores the implications of collecting less data on electric and gas utility costs in the AHS and relying more on modeling for the estimation of utility costs.

2. Utility Cost Data Collection in the AHS

Utility costs in the AHS are self-reported. For monthly electric and unpiped gas costs, respondents are asked to refer to their bills when they report monthly electric and gas costs. In addition to electric and gas costs, questions are also asked about fuel oil, other fuels (i.e., coal, kerosene, wood, and solar), garbage and trash, and water and sewage. The series of utility questions in the AHS have a high respondent burden due to requiring the respondent to use records and to estimate costs when they don't have records.

To prepare respondents for the AHS survey, they are sent an advance letter describing the survey and a worksheet on utility costs, other housing costs, and mortgage costs that they

are asked to fill out. All housing units in the survey are sent the advance letter. In-person respondents who say they have not received an advance letter are provided one at the doorstep

2.1 Electric and Gas Cost Data Collection in the AHS

The goal of the electric and gas cost questions in the AHS are to collect billing data for the months of January, April, August, and December; months that Schwanz and Gorsak (2004) found most predictive of average monthly utility costs. Schwanz and Gorsak (2004) reported that, historically, a third of respondents in AHS provide billing data and the majority who do provide data for all four months. Using data from the 2007 AHS, Carter (2010) found that 39.5% of respondents report at least one month of electric bills and 35.9% report all four months of electric bills. The pattern was similar for gas bills with 40% of respondents reporting at least one month of bills and 36.5% of respondents reporting all four months of bills. Online billing and automatic bill pay options serve as barriers to respondents possessing physical bills. Even though respondents are asked to use their bills during the interview, until 2011, the AHS didn't ask respondents explicitly if they used bills or a worksheet in their interview. In the 2011 AHS, 20% used an electric bill in answering the survey, 18% used a gas bill, and 10% used the worksheet. Among those who used the worksheet, it is not known whether the respondent had the utility bill information filled out on the worksheet.

Respondents who don't have four months of billing data are asked other questions to allow for the estimation of utility costs. If the respondent doesn't have bills, doesn't know what the costs are on the bills, refuses to answer, or provides fewer than two valid amounts for the months, they are asked: "How much was your most recent bill?" and "What month was the bill for?" A subset of respondents are asked to estimate their average monthly costs of electricity. Respondents who use gas are asked similar cost questions as those who use electricity and are asked whether the gas comes from underground pipes or is bottled.

2.2 Other Utilities

In addition to electric and gas costs, the AHS collects information for fuel oil, other fuels (i.e., wood, coal, kerosene or any other fuel), garbage and trash collection, and water supply and sewage disposal. The level of detail collected these costs is less than that for electric and gas costs. Monthly billing data are not collected for these items. What is collected is whether the utility is used, whether it is billed separately or combined with another bill, and the total costs for the year. HUD and Census are exploring developing modeling for estimating bottled gas and fuel oil costs for the 2015 AHS.

3. Utility Data Editing, Estimation, and Imputation

After data collection, the utility data undergo consistency edits to check the data against relevant use and equipment variables and identify cases where utilities are billed separately. Since the AHS does not ask for billing data for all months and respondent recall of utility costs is not always accurate, the Utility Estimation System (UES) was created to estimate annual utility costs using regression models developed from the RECS. The advantage of using the RECS to model utility costs is that it collects administrative data from suppliers on actual billing amounts. The RECS also collects some similar housing characteristics to the AHS. This allows for the construction of models that can then be applied to the AHS. The 25 regression models that make up the UES can be found in Table 6 in the Appendix.

3.1 The Utility Estimation System (UES)

The UES is used to calculate annual estimates for households that use electricity and households that use underground piped natural gas. It consists of three steps: (1) calculating the utility cost estimate, (2) comparing the cost estimate to a cut-off value, and (3) aligning cost estimates to the RECS.

In the first step, cost estimates are calculated by applying one of a set of regression models to the data collected from the respondent. Utility cost regression models are created with the RECS. Since the RECS collects respondent monthly billing data directly from the supplier and asks the respondent questions that are also asked on the AHS, the Census Bureau can create regression models with the RECS and apply AHS responses to the regression parameters to obtain a total utility cost estimate for AHS respondents.

In the next step, each cost estimate is compared to a set of values to avoid extremely low or high estimates. These cut-off values are provided at the Census Division level. If a cost estimate is outside the cut-off values, extreme individual bills are dropped, and a new cost estimate is calculated with the appropriate model. If no extreme individual bills are identified for a case that exceeds the cut-off, the cost estimate is obtained through hot deck imputation.

In the final step, cost estimates are adjusted to align to the RECS total utility cost. An adjustment factor is calculated such that the sum of the AHS cost estimates calculated with billing data added to the adjusted sum of the AHS cost estimates calculated without billing data is equal to the RECS average monthly cost multiplied by the AHS sample size. Cost estimates calculated with billing data are assumed to be correct; therefore, the factor is only applied to those estimates calculated without billing data.

3.2 History of Model Development

The electric and gas models have evolved over time. The U.S. Census Bureau developed the UES in 1993 based on a recommendation from the Office of Management and Budget (OMB). In this initial version, fifteen models were developed for each census division[3]; each model contained parameter estimates for billing months January, April, August, and December, corresponding with the amount of billing data obtained from the respondent. If only one month's billing data was provided, if only January and December electric bills, or if April and August gas bills were provided, additional parameter estimates corresponding with home heating equipment, water heating equipment, and housing and household characteristics were added to the model. Parameter estimates were calculated from the 1990 RECS and inflated to reflect 1993 utility costs. One inflation factor was calculated for electric costs, and another inflation factor was calculated for natural gas costs. Both inflation factors were calculated at the national level. In each AHS survey after 1993, the inflation factors and average monthly costs are updated.

In 2003, the system was updated based on research by Gorsak and Schwanz (2004). Parameter estimates were updated using data from the 1997 RECS. Additionally, two regression models were added: one model using average electric and gas costs estimated by the respondent, and one model assigning 0 to all regression parameters, thereby forcing the case through hot deck.

[3] Alaska was assigned to the New England census division. Hawaii was assigned to the West South Central census division.

In 2007, the parameter estimates were updated with data from the 2001 RECS. Four models were added. These models used billing data for March, May, June, and July to give the Census Bureau more flexibility in the type of data to accept from the respondent.

In 2009, four models corresponding with the remaining billing months (February, September, October, and November) were added to the system. The model from 2003, which assigned 0 to all regression parameters, was updated with 2001 RECS-based parameters for the housing and household characteristics data. This housing and household characteristics model (HHC) was applied to cases with no billing data and no estimate of average monthly costs. In this update to the system, the Census Bureau also split the Electric universe into two smaller universes: housing units that use electric only, and those that use electricity and gas. These universes were created for estimating electric utility cost. This update also saw the introduction of division-level inflation factors. As of 2013, the system updated in 2009 is still in use, with inflation factors and average monthly costs updated to reflect the current year of enumeration.

3.3 Limitations of using RECS

One limitation of using the RECS to model utility costs is that the models are dependent on getting current RECS data. Inflation factors are applied to the models to account for this. The RECS also has a relatively small sample size and limited publicly available geography. In our models, we model costs and consumption at the reportable domain level, which included states and groups of states. In the past, the RECS has released data at the regional and census divisional levels.

The RECS is currently undergoing a redesign and recently was the subject of a National Research Council review. In "Effective Tracking of Building Energy Use," a National Research Council (2012) panel made several recommendations to improve the RECS. They include: improved timeliness; revision of edit procedures; releasing prepublication estimates; increasing sample size to make more state data available; working more closely with suppliers; introducing a multimode approach for data collection (they already use a web survey for suppliers); and conducting an ongoing evaluation of the use of administrative records for data and imputation (e.g., square footage). HUD and Census are working with EIA to obtain preliminary data, finer geographic detail, and other data that may improve the estimation of utility costs in the AHS.

4. Present Study

4.1 Potential Changes for the 2015 AHS

Several options of redesigning utility cost models for the 2015 AHS were explored in our research.

1. Collect no utility data for electric and natural gas costs. Model utility costs with the UES, using a revised housing and household characteristics (HHCR) model.
2. Collect average monthly costs. Model utility costs using UES model for average monthly costs.
3. Adopt the American Community Survey (ACS) approach. Collect costs just for the last month. Model utility costs using UES models for last month's bill.
4. Continue with the current AHS approach. Ask for four months of bills and update UES as new RECS data become available.

We present our evaluation of option 1 in this paper. We did not have the data to test options 2 and 3, as not all respondents are asked for their estimate of average monthly costs and their last month's bill. Schwanz and Gorsak (2004) noted that asking the respondent to estimate their average monthly costs places a high mental burden on respondents, as the task requires respondents to quickly remember their bills from the past year and convert them into a monthly value.

4.2 Study Design

The series of utilities cost questions in the AHS are complicated and many respondents do not have the billing data available to answer them. Since the utility cost questions are burdensome to respondents and costly to collect, Census and HUD conducted simulation studies to evaluate the impacts of respondents providing no monthly utility data. Regression models for electric and gas costs and consumption were estimated with public use data from the 2009 RECS.

The models were applied to data from 2009 AHS internal file to yield estimates for monthly electric and gas costs. Recoding variables was required in both the RECS and the AHS, so that the regression parameters could be adequately applied across surveys. We compared median estimates and frequency distributions of monthly electric and gas costs from our models to estimates from the internal 2009 AHS file and to estimates from the 2009 RECS to assess how close the models approximated costs at the Census Region level when compared to the RECS. Since the RECS is a gold standard for utility costs, we benchmarked our estimates to the 2009 RECS estimates.

4.3 Models

Since a model was developed in 2009 using only housing and household characteristics, we evaluated the predictive capability of this model when applied to all cases in the 2009 AHS National sample. We refer to this model as the HHC model. As the HHC model was developed with 2001 RECS data, and the 2009 RECS data were available, we created a revised HHC model (we call it the HHCR model) with the 2009 RECS data. The HHCR model incorporates Heating Degree Days and Cooling Degree Days[4] to account for variability in consumption. Since Heating Degree Days impact heating equipment usage and Cooling Degree Days impact air conditioning usage, this information was applied conditional on the presence of specific types of heating and cooling equipment. We also assessed the effects of recombining the electric-only and electric/gas-mixed universes. Our universe excludes units where the given utility is included in rent/condominium fee.

[4] A Degree Day is an index calculated based on the difference between the daily average temperature and 65 degrees Fahrenheit. Heating Degree Days are the negative differences and Cooling Degree Days are the positive differences.

Table 1. Variables in Models

Model	HHC	HHCR
Input Data	2001 RECS (inflated to 2009)	2009 RECS
Response Variable	Cost	ln(Consumption)
Independent Variables	Electric Heat	Heating Degree Days (if Electric Heat)
	Gas Heat	Heating Degree Days (if Gas Heat)
		Cooling Degree Days (if Central Air)
		Cooling Degree Days (if Wall Air Unit)
	Electric Hot Water	Electric Hot Water
	Gas Hot Water	Gas Hot Water
	Year Built	Year Built
	Type of Housing Unit (single=1, multi=2, mobile=3)	Single Unit (indicator)
		Multiunit (indicator)
		Mobile Home (indicator)
	Total Rooms	Total Rooms
	Total Bathrooms	Total Bathrooms
	Total Major Appliances	Total Major Appliances
	Number of Household Members	Number of Household Members

These variables are common to the RECS and the AHS. This lets us apply our coefficients from RECS to AHS.

4.3.1 Model Development

We attempted several enhancements in our development of the HHCR model to create a model with high predictive capabilities that would not require applying cut-offs or aligning to the RECS cost totals. First, we attempted to develop models at a more refined geographic level than Census Division. The RECS publishes data at the state (or group of states) level, which they call "reportable domain." Since a Census Division can cover a relatively large geographic area, analysis at the reportable domain level can account for some of the variability in cost at lower levels of geography.

Second, cost data are skewed. To address this, we modeled the natural logarithm of monthly cost as a function of the independent variables; more specifically, the natural log of cost + 1, to account for units that reported 0. During our evaluation, we noticed differences in average costs from state to state. Therefore, we attempted to estimate monthly consumption (natural log of monthly consumption + 1, since consumption data are also skewed) with our independent variables. Correlation coefficients supported this premise.

After iteratively removing outliers[5], we used stepwise regression to determine which variables to include and which variables to remove from each reportable domain's model.

After finalizing our models, we applied our new regression parameters to the recoded 2009 AHS data. After transforming our estimates from the natural log scale back to our base consumption, we multiplied our monthly consumption estimate by each state's average price per unit of consumption to obtain the unit's average monthly cost. To obtain fuel prices, we obtained price and consumption data from the RECS website for each state. We then calculated a weighted average across the 12 months prior to the 2009 AHS enumeration (May 2008 through April 2009). Finally, we multiplied the average by a factor to put the value into a unit of measure equivalent to the RECS public use data. For electricity, the factor was 0.01 to convert cents to dollars. For gas, the factor was 0.1 to convert dollars per 1,000 cubic feet to dollars per 100 cubic feet.

Next, we summarized our results and compared to the RECS. In our initial evaluation, we observed our estimated electric costs were concentrating towards the center of our distributions. This finding led us to split the electric universe back into the two components created in 2009 (electric-only and electric/gas mix). After we split the electric universe, the distribution of the combined data set overlapped the RECS distribution more closely.

5. Results

5.1 Evaluation Methods

The RECS and the AHS are independent samples; we cannot compare each unit's estimated monthly cost to the actual cost. Therefore, our comparisons must be at an aggregate level. We evaluated the quality of our estimates by how closely our distributions matched the medians and distributions calculated from the RECS.

We calculated quantiles for all units, medians for selected characteristics, and categorical percent distributions for the UES estimates from the 2009 AHS, estimates generated by applying the current Housing and Household Characteristics (HHC) model applied to all cases in the 2009 AHS, and estimates generated by applying the HHCR model applied to all cases in the 2009 AHS. We compared all three of our estimated distributions to the distributions from the 2009 RECS.

We estimated the variation in our quantiles, medians, and categorical percent distributions attributed to repeated sampling for each survey using replicate weights. We obtained a pooled standard error estimate and performed t tests on the null hypothesis of zero difference between each distribution and the RECS distribution, with a two-sided α=0.1.

5.2 Quantiles for All Units

Tables 2 and 3 show the national electric and natural gas quantiles of average monthly costs (in dollars). The UES-derived values were not topcoded for publication, and therefore values at or above the 97.5th percentile were suppressed for confidentiality. At the national level, all differences between evaluation medians and the RECS median were statistically significant, for both electric costs and natural gas costs. However, the national distributions calculated with the proposed model followed the RECS visually. The electric median for the HHCR model was lower than the 2009 RECS while the UES and HHC medians were

[5] Outliers were determined based on a negative impact on the regression model. We defined an outlier as a value with a large Cook's D and a small COVRATIO.

higher. Gas medians for all three AHS models were higher than the 2009 RECS gas median.

Table 2. National Electric Quantiles (in dollars)

	2009 AHS			2009 RECS
Quantile	UES	HHC	HHCR	
0% Min	3	-30	0	0
1%	21	13	14	18
5%	37	42	36	30
10%	48	55	45	40
25% Q1	71	77	63	62
50% Median	108	106	90	97
75% Q3	159	147	128	144
90%	224	193	169	204
95%	273	220	198	245
99%	*	273	268	356
100% Max	*	404	1097	1587

Source: 2009 American Housing Survey; 2009 Residential Energy Consumption Survey

Table 3. National Natural Gas Quantiles (in dollars)

	2009 AHS			2009 RECS
Quantile	UES	HHC	HHCR	
0% Min	2	-70	0	1
1%	11	3	7	4
5%	22	32	22	11
10%	31	47	31	20
25% Q1	48	68	48	38
50% Median	73	90	71	61
75% Q3	111	119	98	90
90%	162	149	130	126
95%	203	165	154	155
99%	*	200	210	226
100% Max	*	413	590	530

Source: 2009 American Housing Survey; 2009 Residential Energy Consumption Survey

5.3 Medians for Selected Housing Characteristics

Even if the model estimates from the UES, HHC, and HHCR matched the 2009 RECS perfectly, we would want to know if medians of average utility costs were similar across selected types of housing units. For this reason, we compared electric and natural gas monthly costs for subsets of the total housing population. Median electric and natural gas average monthly costs (in dollars) were calculated for the following characteristics for the UES, the HHC, and the HHCR models and were compared to the 2009 RECS:

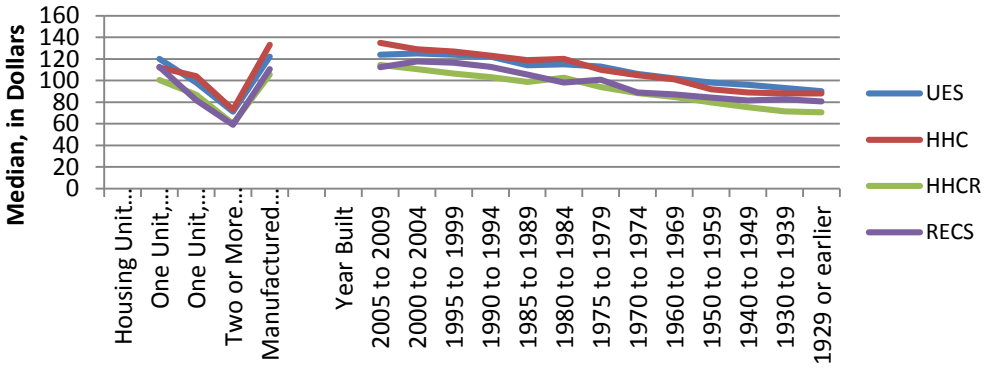
- Housing Unit Type (one unit attached, one unit detached, building with two or more units, manufactured/mobile home)
- Tenure (owner, renter)
- Year Built
- Square Feet
- Total Number of Rooms
- Bedrooms
- Complete Bathrooms
- Householder Characteristics (Black, White, Hispanic, Elderly, and Married)
- Education Level of Householder
- Number of Household Members
- Household Income
- Below Poverty
- Subsidized Housing (owned by Public Housing Authority, receive government subsidy)

Altogether, there were 75 categories in all of these characteristics.

5.3.1 Electric

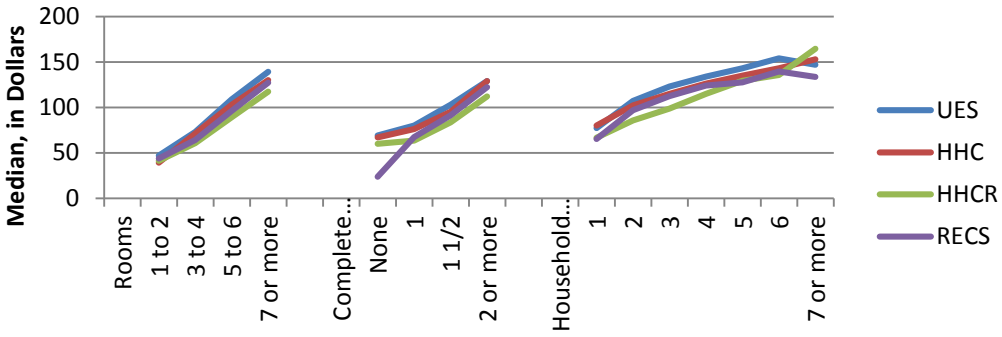
Seven of the medians for the 75 categories calculated with the UES model were not significantly different from RECS. Using the HHC model, 14 out of 75 calculated were not significantly different from RECS. Using the HHCR model, 27 out of 75 calculated were not significantly different from RECS. Additionally, the UES and the HHC models overestimated the medians of most of these characteristics if the difference was significant, while the HHCR model underestimated the medians. The following tables provide a graphical representation of the medians of select characteristics. We show these characteristics because they were included as terms in the regression models.

Figure 1. Median Costs for Electric by Housing Unit Type and Year Built



Source: 2009 American Housing Survey; 2009 Residential Energy Consumption Survey

Figure 2. Median Costs for Electric by Numbers of Rooms, Bathrooms, and Household Members

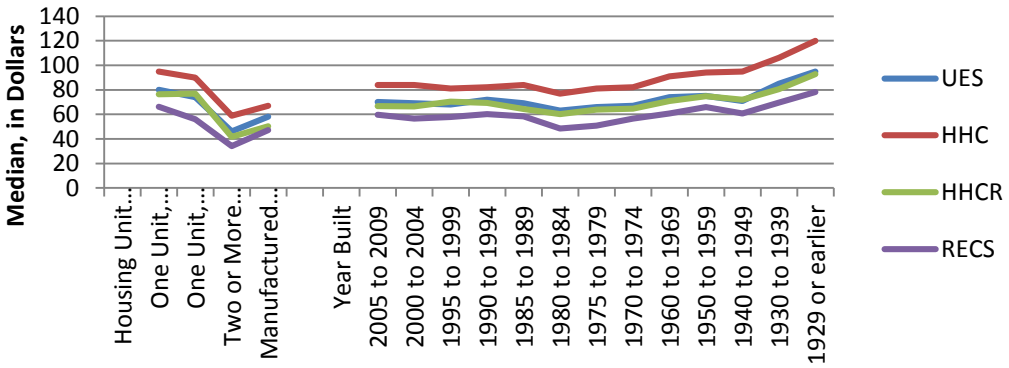


Source: 2009 American Housing Survey; 2009 Residential Energy Consumption Survey

5.3.2 Natural Gas

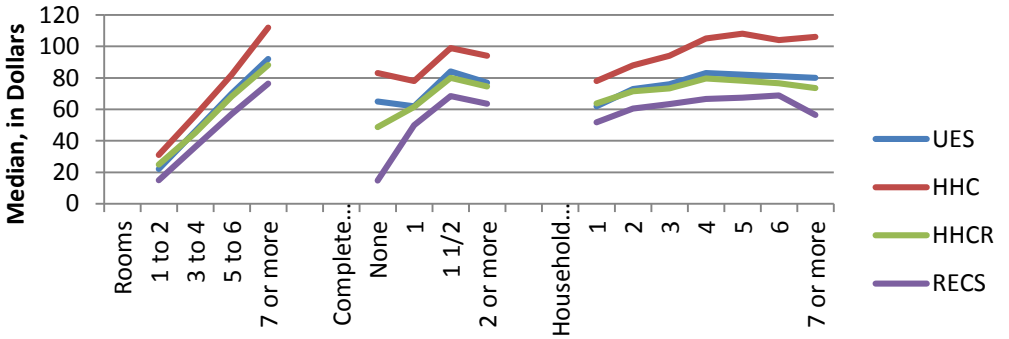
The '0 bedrooms' characteristic was not evaluated because the sample contained only 2 units in the Natural Gas universe. Of the remaining 74 characteristics, all of the medians calculated with the UES or the HHC model were significantly different from the RECS. Using the HHCR model, only 5 characteristics' medians were not significantly different from the RECS. All three models consistently overestimated the RECS medians. The following tables provide a graphical representation of the medians of select characteristics. We show these characteristics because they were included as terms in the regression models.

Figure 3. Median Costs for Natural Gas by Housing Unit Type and Year Built



Source: 2009 American Housing Survey; 2009 Residential Energy Consumption Survey

Figure 4. Median Costs for Natural Gas by Numbers of Rooms, Bathrooms, and Household Members



Source: 2009 American Housing Survey; 2009 Residential Energy Consumption Survey

5.4 Categorical Percent Distributions

Data analysts often use categorical percent distributions to reduce noise in distributions to make them more interpretable. The AHS National publication (HUD and Census, 2009) provides categorical distributions of the number of housing units paying different monthly costs for electricity and monthly costs for gas. Each of these tables contains seven groups to provide the user with a distribution of housing units' monthly utility expenses, grouped into the following categories: Less than $25, $25 to $49, $50 to $74, $75 to $99, $100 to $149, $150 to $199, and $200 or more.

For our analysis, we converted the counts to percentages to account for differences in the number of housing units estimated by RECS and those estimated by AHS. We then performed statistical testing on the differences in percentages between our three models

and the RECS. Groups with non-significant differences are denoted with an asterisk (*) in tables 4 and 5.

Table 4. National Electric Cost Distributions

	2009 AHS			2009 RECS
	UES	HHC	HHCR	
Less than $25	1.6%	2.2%	1.8%	2.7%
$25 to $49	9.4%	5.1%	11.6%	13.4%
$50 to $74	16.7%	16.5%	22.9%	18.2%
$75 to $99	17.6%*	21.2%	21.5%	17.4%
$100 to $149	26.6%	31.1%	26.3%*	25.3%
$150 to $199	14.3%	15.4%	11.0%	12.3%
$200 or more	13.8%	8.6%	4.9%	10.7%

Source: 2009 American Housing Survey; 2009 Residential Energy Consumption Survey

At the national level, one cell calculated from the UES ($75 to $99) was not significantly different from the RECS. All cells calculated from the HHC model were significantly different from the RECS. One cell calculated from the HHCR model ($100 to $149) was not significantly different from the RECS. Overall, no one model outperformed the others when replicating the distribution calculated from the RECS.

Table 5. National Natural Gas Cost Distributions

	2009 AHS			2009 RECS
	UES	HHC	HHCR	
Less than $25	6.3%	3.2%	6.4%	13.1%
$25 to $49	20.4%	8.1%	20.5%	25.1%
$50 to $74	24.3%*	20.5%	26.0%*	25.0%
$75 to $99	17.5%*	27.3%	22.9%	16.9%
$100 to $149	19.0%	31.1%	18.6%	14.2%
$150 to $199	7.1%	8.7%	4.3%*	3.9%
$200 or more	5.3%	1.1%	1.4%*	1.8%

Source: 2009 American Housing Survey; 2009 Residential Energy Consumption Survey

At the national level, two cells calculated from the UES ($50 to $74 and $75 to $99) were not significantly different from the RECS. Three cells calculated from the HHCR ($50 to $74, $150 to $199, and $200 or more) were not significantly different from the RECS. Although the UES and the HHCR model outperformed the HHC model, neither one outperformed the other consistently to match the RECS distribution.

5.4.1 Regional Distribution Comparisons – Electric

We performed significance testing on the RECS at the regional level and identified how closely the UES, the HHC, and the HHCR estimates matched the distributions of the RECS.

Across all four regions, eight of 28 categories summarized with UES were not significantly different from RECS. Seven of 28 groups summarized with HHC model were not significantly different from RECS. Five of 28 groups summarized with the HHCR model were not significantly different from RECS.

5.4.2 Regional Distribution Comparisons – Natural Gas

Across all four regions, six of the 28 groups summarized with UES were not significantly different from RECS. Three of the 28 groups summarized with HHC model were not significantly different from RECS. Eight of the 28 groups summarized with the HHCR model were not significantly different from RECS.

6. Conclusions

The HHCR model outperformed the UES in estimating aggregate electric monthly costs, but underestimated the median electric average monthly cost at the national level. Splitting the electric universe into electric-only and electric/gas mix subsets enabled us to capture the segment of the population that consumes less electricity. Using a natural logarithm transformation on consumption helped us to capture a skewed consumption distribution, but it did not adequately capture the higher percentiles. The HHCR model overestimated the median natural gas average monthly cost at the national level.

Future research on the utility models by HUD and Census will include investigations into more effective ways to estimate electric costs for units that consume larger amounts of electricity and the exploration of more effective methods for estimating monthly costs for units that consume smaller amounts of natural gas. We are looking into improving the model performance by incorporating data on monthly heating degree days and cooling degree days, refining the specification of appliances, using income to better model nonlinearities, and using supplier prices averages to better capture geographic variations in costs. We plan to re-estimate the UES using data from the 2009 RECS to provide a more appropriate comparison system with which to evaluate the HHCR approach. In addition, we plan to estimate models of oil and bottled costs in the AHS using models developed from the RECS. Finally, as utility costs are a component of housing costs, we are evaluating the effects of changes in estimation methodology on estimates of total housing costs.

Acknowledgments

The authors would like to like to thank Stephen Ash, Tamara Cole, Arthur Cresce, Ruth Ann Killion, and Ernest Lawley of Census, Shawn Bucholtz of HUD, and James Berry, Edgardo Cureg, and Eileen O'Brien of the Energy Information Administration for their helpful comments on the research

References

Carter, George. 2010. "The Use of Utility Records as an Information Retrieval Aid in the American Housing Survey." Proceedings of the American Statistical Association, Survey Research Methods Section.

National Research Council. 2012. Effective Tracking of Building Energy Use: Improving the Commercial Buildings and Residential Energy Consumption Surveys. Washington, DC: The National Academies Press.

U.S. Department of Housing and Urban Development and U.S. Census Bureau. 2009. *American Housing Survey for the United States: 2009*.

Appendix

Table 6. Billing Data Used In Each UES Model

Model	Month			
1	January	April	August	December
2	January		August	December
3	January	April		December
4	January	April	August	
5		April	August	December
6	January	April		
7	January		August	
8	January			December
9		April	August	
10		April		December
11			August	December
12-23	Individual months			
24	Average monthly data needed			
25	No monthly data			