AMERICAN HOUSING SURVEY

Item Nonresponse and Imputation: 2015 to 2023

LAST UPDATED: SEPTEMBER 2024

U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
U.S. CENSUS BUREAU
U.S. DEPARTMENT OF COMMERCE

AMERICAN HOUSING SURVEY

Contents

AMERICAN HOUSING SURVEY

List of Exhibits

AMERICAN HOUSING SURVEY

1. Overview

The purpose of this document is to explain how the Department of Housing and Urban Development (HUD) and the U.S. Census Bureau (Census Bureau) implemented imputation strategies for item non-response in the American Housing Survey (AHS) from 2015 to 2023.

The AHS includes a multitude of questions covering a broad array of topics. In some instances, households refuse to respond to the entire survey. This is called unit nonresponse. HUD and the Census Bureau use a statistical technique called non-response adjustment to ensure that estimates produced from AHS data reflect both responding and nonresponding households.

When households chose to respond to the AHS, they generally provide answers to all survey questions. However, for some households, the Census Bureau is unable to obtain a response to every survey question. This is called item nonresponse. There are at least four reasons why item nonresponse occurs:

- A household chooses not to answer a question.
- A household did not know the answer to a question.
- A household is not asked the question because the household quit an interview early but answered enough questions to be considered a complete interview.
- A household was not eligible to be asked a question at the time of the interview but was edited in a way after collection that made the housing unit eligible to be asked the question.

A desirable outcome for many surveys is to have complete responses for each survey item (variable). When a survey variable includes missing responses, it can often be difficult for beginner or novice survey data users to adjust their analysis to account for the missing responses. To improve the usability of survey data, survey managers often “fill in” the missing data. This process is called imputation. Imputation improves the completeness and consistency of survey data and prepares the survey data for weighting and estimation processes.

HUD and the Census Bureau use imputation to fill in missing responses in the AHS. However, it is important to note that missing responses are imputed in the AHS, but only for certain variables. In fact, most AHS variables with missing responses are not imputed.

Section 2 of this help guide describes the general imputation techniques used in the AHS from 2015 to 2023. Section 3 describes the specific imputation technique for each public use file (PUF) microdata and Internal Use File (IUF) microdata variable and is organized by subsection (topic) and exhibit table (subtopic). Each table row indicates whether that variable can be found on the PUF or the IUF only.

2. Imputation Techniques in the AHS

There are numerous ways to impute missing responses and the method chosen often depends on the type of missing data. For the AHS, four different imputation techniques are used. The following section describes and summarizes the different imputation procedures.

2.1 Deductive Imputation

This method of imputation uses reported pieces of information to fill in missing values. For example, if payment amount on a mortgage is missing, but original loan amount, payment frequency, interest rate, and term are all reported we will use a financial formula to fill in the missing payment amount.

We also do a type of weighted coin flip to fill in some missing values for variables where we know some information about the population. For example, suppose an AHS respondent refused to answer the question about whether they owned or rented the home. We know from other data sources that about 60 percent of occupied housing units are owners and 40 percent are renters. We would then assign a 60 percent chance of filling in this missing observation as an owner (and a 40 percent chance that they are a renter.)

2.2 Cold Deck Imputation

The AHS is a longitudinal survey, meaning most housing units are surveyed every two years. The AHS includes questions about features of the housing unit. Some features of housing units that are part of the AHS sample are not expected to change over time. When a household does not answer a question about a feature of the housing unit and the question is about a feature that is not expected to change over time and there is a response to the question from a prior survey, the missing value is imputed with a value collected in a prior interview of the same housing unit. For example, having bodies of water near a housing unit is not expected to change. These data are filled in with prior year responses if one is available. This form of imputation is not the focus of this document, as it is taken care of by the survey instrument or consistency edits prior to other forms of imputation.

2.3 Hot Deck Imputation

In this method of imputation, a household with a missing value for an item (recipient) “borrows” a value from another household who provided a valid response for that item (donor). The hot deck imputation procedure is implemented in a way that attempts to match a recipient household with a donor household based on a common set of characteristics, referred to as the hot deck. Hot deck imputation is called “hot deck” because the donors and recipients are a part of the same data set. The deck is “hot” because it is currently being processed, as opposed to “cold deck” imputation where data is filled in with information from another source.

Before imputation, all records are sorted by an internal variable that contains some geographic information. This sorting keeps donor and recipient records geographically close to each other. The geography used to sort the records for imputation are state and county. After sorting is complete variables are imputed one at a time in a single pass through the data. For example, suppose a household reported their home was constructed in 2010 and has three bedrooms, but did not provide a valid response for number of bathrooms. In a hot deck procedure, the hot deck could be based on year built and number of bedrooms, such that the pool of potential donors most likely to be matched to the recipient are those with a similar year built and similar number of bedrooms, along with being close geographically.

In the AHS, the variables that define the hot deck are chosen because they are expected to be correlated, or more generally, they are associated, with the variable being imputed.

2.4 Special Imputation

For three specific groups of variables, specialized imputation procedures are applied. These groups are income, utilities, and home improvement. The special procedures are described in subsequent sections. Links to more in-depth documentation for Income and Utility special imputations can be found in the references section at the end of this document.

2.5 Variables with Imputed Inputs

Some variables in the AHS are created from other variables in the survey. We refer to these types of variables as recodes. For example, the recoded variable NUMADULTS (number of adults in the household) is created by summing the reported number of persons 18 years of age or over living in the unit. To ensure consistency, most of these recoded variables are determined after all edits and imputations to the various input variables are complete. For example, suppose a household had three persons on the roster, and all persons had missing values of age. During data processing, the missing age values would be imputed before any recoded variables based on age are determined (such as number of adults in the household). Recoded variables with imputed inputs are shown in the following tables with the imputed input variables indicated in parenthesis in the Imputation Method column.

3. Specific Imputation Strategy

The subsections in this section (3.1, 3.2, etc.) correspond to topics in the AHS PUF. Each subsection has one or more exhibit tables corresponding to subtopics within the subsection's topic. Also, it is important to note that the order in which variables are imputed is based on internal decisions by AHS managers and does not follow the topic/subtopic format.

Several variables are imputed using methods developed for the Current Population Survey (CPS). More information on these methods, referred to as CPS Imputations in the tables below, can be found at https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/imputation-of-unreported-data-items.html. Note, the AHS does not use the CPS longitudinal imputations methods.

3.1 Demographics Topic

Exhibit 3.1.1 Householder Demographics Subtopic

Variable	Description	Imputation Method	Availability
HHADLTKIDS	Number of the householder's unmarried children age 18 and over, living in this unit	Recoded from Imputed Inputs (see REL, AGE)	PUF
HHAGE	Age of householder	CPS Imputations[1] [2]	PUF
HHCITSHP	US citizenship of householder	CPS Imputations[1]	PUF
HHGRAD	Educational level of householder	CPS Imputations[1]	PUF
HHINUSYR	Year householder came to US	CPS Imputations[1]	PUF
HHMAR	Marital status of householder	CPS Imputations[1]	PUF
HHMOVE	Year householder moved in	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
HHNATVTY	Country of birth of householder	CPS Imputations[1]	PUF
HHMNTVTY	Country of birth of householder's mother	CPS Imputations[1]	PUF
HHFNTVTY	Country of birth of householder's father	CPS Imputations[1]	PUF
HHGEN	Householder's generational status	Recoded from Imputed Inputs (see HHCITSHP, HHMNTVTY, HHFNTVTY)	PUF

[1] For further information on the CPS Imputations, see https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/imputation-of-unreported-data-items.html.

[2] During data processing, cases in which there was no response for date of birth or estimated age undergo CPS age imputations. From 2015-2021, imputed AGE values were edited for consistency with the "under age 14" variable (as this variable is not used as an input for the CPS age imputations). For example, if the respondent indicated a person was under age 14, any imputed values of AGE that were greater than or equal to 14 years old were edited to be 13 years old. Alternatively, if the respondent indicated a person was not under age 14, any imputed values of AGE between 0-13 years old were edited to be 15 years old. In 2023, these consistency edits were inactivated, and there were no further edits to the imputed values of AGE.

HHOLDKIDS	Number of the householder's children age 6 through 17, living in this unit	Recoded from Imputed Inputs (see REL, AGE)	PUF
HHRACE	Race of householder	CPS Imputations[1]	PUF
HHRACEAS	Race of householder (Asian group)	CPS Imputations[1]	PUF
HHRACEPI	Race of householder (Native Hawaiian or other Pacific Islander group)	CPS Imputations[1]	PUF
HHSEX	Sex of householder	Missing values of HHSEX were deductively imputed based on information available on the household roster (e.g., if a female opposite-sex spouse of the reference person was present, then the sex of the reference person must be male). If there was not enough information on the roster to deduce a missing value of HHSEX it was instead allocated using CPS Imputations[1] [3] Note, the deductive approach was newly implemented in 2023 (for 2015-2021, missing values of SEX were imputed using only CPS Imputations)	PUF
HHSPAN	Spanish origin of householder	CPS Imputations[1]	PUF
HHYNGKIDS	Number of the householder's children under age 6, living in this unit	Recoded from Imputed Inputs (see REL, AGE)	PUF

Exhibit 3.1.2 Person Demographics Subtopic

Variable	Description	Imputation Method	Availability
AGE	Age of person	CPS Imputations[1]	PUF
CITSHP	US citizenship of person	CPS Imputations[1]	PUF
FAMLINE	Family number of person	Recoded from Imputed Input (see REL)	PUF
FAMREL	Family relationship of person	Recoded from Imputed Input (see REL)	PUF
FAMTYP	Family type of person	Recoded from Imputed Input (see REL)	PUF
GRAD	Educational level of person	CPS Imputations[1]	PUF
INUSYR	Year person came to the US	CPS Imputations[1]	PUF
MAR	Marital status of person	CPS Imputations[1]	PUF
MIL	Military status of person	CPS Imputations[1]	PUF
MLPA	Flag indicating this person served in September 2001 or later	CPS Imputations[1]	PUF
MLPB	Flag indicating this person served in from August 1990 - August 2001 (including Persian Gulf War)	CPS Imputations[1]	PUF
MLPCD	Flag indicating this person served in from May 1975 - July 1990	CPS Imputations[1]	PUF

[3] For HHSEX, the CPS Edits hot deck sorted the data file by household ID and allocated missing values of HHSEX by alternating between values of '1' (male) and '2' (female) until all missing values were allocated.

MLPE	Flag indicating this person served in Vietnam era (August 1964 - April 1975)	CPS Imputations[1]	PUF
MLPFG	Flag indicating this person served from February 1955 - July 1964	CPS Imputations[1]	PUF
MLPH	Flag indicating this person served in Korean War (July 1950 - January 1955)	CPS Imputations[1]	PUF
MLPI	Flag indicating this person served from January 1947 - June 1950[1]	CPS Imputations[1]	PUF
MLPJ	Flag indicating this person served in World War II (December 1941 - December 1946)	CPS Imputations[1]	PUF
MLPK	Flag indicating this person served in November 1941 or earlier	CPS Imputations[1]	PUF[4]
MOVE	Year person moved in	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
MOVM	Month person moved in	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	IUF Only
NATVTY	Country of birth of person	CPS Imputations[1]	PUF
RACE	Race of person	CPS Imputations[1]	PUF
RACEAS	Race of person (Asian group)	CPS Imputations[1]	PUF
RACEPI	Race of person (Native Hawaiian or other Pacific Islander group)	CPS Imputations[1]	PUF
REL	Relationship of person to reference person	CPS Imputations[1]	PUF
RESPONDENT	Flag indicating this person is the respondent	Starting in 2019, if the respondent's line number was missing, then the respondent was assumed to be the reference person.	PUF
SEX	Sex of person	Missing values of SEX were deductively imputed based on information available on the household roster (e.g., if a female opposite-sex spouse of the reference person was present, then the sex of the reference person must be male). If there was not enough information on the roster to deduce a missing value of SEX it was instead allocated using CPS Imputations[1] [5] Note, the deductive approach was newly implemented in 2023 (for 2015-2021, missing	PUF

[4] The MLPI variable is collapsed with the MLPK variable on the 2021 and later PUFs because the MLPK variable is IUF Only in 2021 and 2023.

[5] For SEX, the CPS Imputations sorted the data file by household ID and person line number and then allocated missing values of SEX by alternating between values of '1' (male) and '2' (female) until all missing values were allocated.

		values of SEX were imputed using only CPS Edits, hot deck)	
SPAN	Spanish origin of person	CPS Imputations[1]	PUF

Exhibit 3.1.3 Total Household Demographics Subtopic

Variable	Description	Imputation Method	Availability
GRANDHH	Flag indicating grandparent headed household with no parent present	Recoded from Imputed Inputs (see REL)	PUF
HSHLDTYPE	Type of household	Recoded from Imputed Inputs (see REL, MAR, SEX)	PUF
MILHH	Type of military household	Recoded from Imputed Inputs (see MIL and the MLP variables)	PUF
MULTIGEN	Type of multigenerational household	Recoded from Imputed Input (see REL)	PUF
NUMADULTS	Number of persons age 18 and over living in this unit	Recoded from Imputed Input (see AGE)	PUF
NUMELDERS	Number of persons age 65 and over, living in this unit	Recoded from Imputed Input (see AGE)	PUF
NUMNONREL	Number of nonrelatives living in this unit	Recoded from Imputed Input (see REL)	PUF
NUMOLDKIDS	Number of children age 6 through 17, living in this unit	Recoded from Imputed Input (see AGE)	PUF
NUMSECFAM	Number of secondary families living in this unit	Recoded from Imputed Inputs (see REL)	PUF
NUMSUBFAM	Number of subfamilies living in this unit	Recoded from Imputed Inputs (see REL)	PUF
NUMVETS	Number of veterans living in this unit	Recoded from Imputed Inputs (see AGE, MIL)	PUF
NUMYNGKIDS	Number of children under age 6, living in this unit	Recoded from Imputed Input (see AGE)	PUF
PARTNER	Flag indicating unmarried partner household	Recoded from Imputed Inputs (see REL, SEX)	PUF
SAMESEXHH	Flag indicating same sex married household	Recoded from Imputed Inputs (see REL, SEX)	PUF

3.2 Equipment and Appliances Topic

Exhibit 3.2.1 Appliances Subtopic

Variable	Description	Imputation Method	
COOKFUEL	Type of fuel used most for cooking	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
COOKTYPE	Type of cooking equipment	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF
DISHWASH	Flag indicating unit has a working dishwasher	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
DRYER	Type of clothes dryer	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF

FRIDGE	Flag indicating unit has a working refrigerator	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF
KITCHSINK	Flag indicating unit has a kitchen sink	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF
WASHER	Flag indicating unit has a working washing machine	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF

Exhibit 3.2.2 Cooling Subtopic

Variable	Description	Imputation Method	Availability
ACPRIMARY	Type of primary air conditioning	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
ACSECNDRY	Type of secondary air conditioning	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF

Exhibit 3.2.3 Heating Subtopic

Variable	Description	Imputation Method	Availability
FIREPLACE	Flag indicating unit has a useable fireplace	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
HEATFUEL	Type of main heating fuel	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
HEATTYPE	Type of main heating equipment	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF

Exhibit 3.2.4 Water and Sewer Subtopic

Variable	Description	Imputation Method	Availability
HOTWATER	Type of hot water system	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF
SEWTYPE	Type of sewer system	Imputation for a component of this recoded variable began in 2017. The imputation is model-based and uses structure type and housing unit density as predictors.	PUF
WATSOURCE	Source of water for unit	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF

3.3 Food Security Topic (2015 only)

Exhibit 3.3.1 Food Security Subtopic

Variable	Description	Imputation Method	Availability
FSEATLESS	Flag indicating adults in the household ate less than they felt they should because there wasn't enough money for food, in the last 30 days	Deductive based on the proportions of reported data.	PUF
FSHUNGRY	Flag indicating adults in the household were hungry due to not having money for food, in the last 30 days	Cold deck based on the proportions of reported data.	PUF
FSLOSTWGT	Flag indicating adults in the household lost weight due to not having money for food, in the last 30 days	Cold deck based on the proportions of reported data.	PUF

3.4 Home Improvement Topic

Exhibit 3.4.1 General Subtopic

Variable	Description	Imputation Method	Availability
REMODAMT	Total cost of home improvement jobs in last two years	Recoded using imputed inputs (See JOBCOST).	PUF

Exhibit 3.4.2 Job Specific Subtopic

Variable	Description	Imputation Method	Availability
JOBCOST	Cost of home improvement job	Hot deck imputation using weighted median of combined home improvement job costs, median of job cost by job type, age of householder, household income, current market value of unit, and whether job was do-it-yourself.	PUF
JOBDIY	Flag indicating job was do-it-yourself	Logistic regression model using home improvement job type, age of householder, household income, current market value of unit, urban/rural status, and structure type.	PUF

3.5 Housing Costs Topic

Starting in 2021, Census and HUD added a new imputation for MORTSTAT (Number and type of mortgages or similar debts) and updated existing imputation methodology for MARKETVAL (Current market value of unit) and RENT (Monthly rent amount).

MORTSTAT Imputation

From 2015 to 2019, the AHS treated owner-occupied units that did not report whether they had a mortgage as owned free and clear. Census found that the lack of imputation for this variable could cause underreporting of the number of mortgaged units and the monthly total housing cost amount since some homeowners who did not report their mortgage status likely had a mortgage.

In 2021, Census developed a logistic regression model using stepwise selection to predict whether a unit was mortgaged at the time of interview for owner-occupied units missing this information. The list of input variables used in the modeling process included:

- Householder characteristics: age, income, education, gender, race and ethnicity
- Years in the unit
- Acquisition information: bought unit outright, received unit as a gift or inheritance, had a down payment, first time home buyer
- Unit details: structure type, number of bedrooms
- Lack of homeowner's insurance

Units predicted to be mortgaged were given one loan on the mortgage data set. These loans are classified as regular loans following AHS convention for mortgages without a reported loan type. The respondents for households given loans due to the imputation were never asked the mortgage questions. Therefore, the mortgage level variables for the imputed loans are either missing or imputed for variables with an imputation in place. This results in higher overall housing costs for units with imputed mortgages, since the mortgage payment is imputed for those units that were imputed to have a mortgage.

Hot Deck Imputation Updates

During the imputation process for current market value (MARKETVAL) and amount paid in rent (RENT), not reported values are first assigned to receive a donor from a "high value" or "low value" group on an alternating basis. Then the units are passed through the standard hot deck and receive values from their high or low value group donor.

Since at least 1997, the imputation had been using a static formula to define "high" and "low" market value and rent groups that did not fully account for changes in home value, rent, and inflation. Over time, this resulted in a much larger share of donors making up the "high" value donor pool.

In order to better align the high and low value groups used for the imputation with current values, Census developed a method for revising the formula using linear regression to calculate a more reasonable cutoff based on the number of bedrooms. This results in imputed values for market value and rent that were generally higher than previous survey waves, but also more in line with the reported distributions. This method was used to update the formula in 2021 and again in 2023.

Utilities Imputation

In 2021, the Census Bureau updated the Utility Estimation models using the Energy Information Administration's 2015 Residential Energy Consumption Survey (RECS), the most recently available data at the time of development. Additionally, Census made modifications to bring the RECS and AHS distributions closer to each other. For more information on the differences in data and methodologies between the 2009 RECS and the 2015 RECS modeling, see Historical Changes.

Exhibit 3.5.1 Owner's Purchase, Value, and Land Debt Subtopic

Variable	Description	Imputation Method	Availability
MARKETVAL	Current market value of unit	Hot deck imputation using structure type, tenure, property value or rent, number of	PUF

		bedrooms, vacancy status, and demographic information (e.g., age, race).	
MORTSTAT	Number and type of mortgages or similar debts.	Units not reporting whether they had a mortgage were imputed by a logistic regression using householder and housing unit characteristics associated with whether a unit is mortgaged. This imputation was added in 2021.	IUF Only

Exhibit 3.5.2 Total Housing Cost Subtopic

Variable	Description	Imputation Method	Availability
INSURAMT	Monthly homeowner or renter insurance amount	Deductive imputation and/or hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race).	PUF
MORTAMT	Monthly total mortgage amount	Recoded using imputed inputs (See INSURAMT, PMTAMT, PROTAXAMT)	PUF
PROTAXAMT	Monthly property tax amount	Deductive imputation.	PUF
RENT	Monthly rent amount	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race).	PUF
TOTHCAMT	Monthly total housing costs	Recoded using imputed inputs (See INSURAMT, MORTAMT, PROTAXAMT, RENT, UTILAMT) See https://www.huduser.gov/portal/sites/default/files/pdf/American-Housing-Survey.pdf.	PUF
UTILAMT	Monthly total utility amount	Recoded using imputed inputs (See ELECAMT, GASAMT, OILAMT, OTHERAMT, TRASHAMT, WATERAMT) See https://www.huduser.gov/portal/sites/default/files/pdf/American-Housing-Survey.pdf.	PUF

Exhibit 3.5.3 Utilities Subtopic

Variable	Description	Imputation Method	Availability
ELECAMT	Monthly electric amount	Model-based imputation. See https://www.huduser.gov/portal/sites/default/files/pdf/American-Housing-Survey.pdf. This variable is an input to the housing adequacy variable.	PUF
GASAMT	Monthly gas amount	Model-based imputation. See https://www.huduser.gov/portal/sites/default/files/pdf/American-Housing-Survey.pdf.	PUF
OILAMT	Monthly oil amount	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race).	PUF
OTHERAMT	Monthly amount for other fuels	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race).	PUF
TRASHAMT	Monthly trash amount	Hot deck imputation using structure type, tenure, property value or rent, number of	PUF

		bedrooms, vacancy status, and demographic information (e.g., age, race).	
WATERAMT	Monthly water amount	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race).	PUF

3.6 Housing Problems Topic

Exhibit 3.6.1 Electrical Problems Subtopic

Variable	Description	Imputation Method	Availability
FUSEBLOW	Number of times fuses blown or circuit breakers tripped in last 3 months	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race). This variable was not imputed prior to 2017. This variable is an input to the housing adequacy variable.	PUF
NOWIRE	Flag indicating electrical wiring concealed by walls	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF
PLUGS	Flag indicating every room has working electrical plug	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF

Exhibit 3.6.2 General Subtopic

Variable	Description	Imputation Method	Availability
ADEQUACY	Type of housing adequacy	Recoded using imputed inputs (See other subtopics in this section and section 3.10 for imputed input variables used to created ADEQUACY)	PUF
UPKEEP	Type of upkeep problems	Recoded using imputed inputs (See, LEAKO, LEAKI, FLOORHOLE, WALLCRACK, PAINTPEEL, RODENT)	PUF

Exhibit 3.6.3 Heating Problems Subtopic

Variable	Description	Imputation Method	Availability
COLD	Flag indicating unit was uncomfortably cold for 24 hours or more last winter	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race). This variable was not imputed prior to 2017. This variable is an input to the housing adequacy variable.	PUF
COLDEQ	Flag indicating unit as uncomfortably cold for 24 hours or more last winter because the main heating equipment broke down	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race). This variable was not imputed prior to 2017.	PUF

Exhibit 3.6.4 Leaking and Mold Subtopic

Variable	Description	Imputation Method	Availability
LEAKI	Flag indicating inside water leaks in last 12 months	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race). This variable was not imputed prior to 2017. This variable is an input to the housing adequacy variable.	PUF
LEAKO	Flag indicating outside water leaks in last 12 months	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race). This variable was not imputed prior to 2017. This variable is an input to the housing adequacy variable.	PUF

Exhibit 3.6.5 Pests Subtopic

Variable	Description	Imputation Method	Availability
ROACH	Number of times there were signs of live or dead cockroaches in last 12 months	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race).	PUF
RODENT	Number of times there was signs of rodents in last 12 months	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF

Exhibit 3.6.6 Plumbing Problems Subtopic

Variable	Description	Imputation Method	Availability
NOTOIL	Flag indicating if unit had any toilet breakdowns in last 3 months	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race).	PUF
SEWBREAK	Number of sewer breakdowns within last 3 months that last 6 hours or more	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race).	PUF

Exhibit 3.6.7 Structural Problems Subtopic

Variable	Description	Imputation Method	Availability
FLOORHOLE	Flag indicating floor has holes	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF
PAINTPEEL	Flag indicating interior area of peeling paint larger than 8 x 11	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF
WALLCRACK	Flag indicating inside walls or ceilings have open holes or cracks	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age,	PUF

		race). This variable is an input to the housing adequacy variable.	

3.7 Income Topic

Income variables in the AHS have been imputed in the same way since 2005, including on the most recent sample first drawn in 2015. As developed by Susin (2005), the AHS uses a hybrid of a regression model and a hot deck procedure to impute the variables that represent the twelve types of income. For types of income that must have values greater than or equal to zero, the imputation had two steps, as defined below:

Step 1: Impute the presence of income.

Separate OLS regression models are estimated for each of the twelve types of income. The model predicts the probability of the presence of any income of the given type. While the outcome of the OLS regression is the predicted probability of having income for each observation ranging from 0 to 1, the 0/1 imputed values are then selected given the estimated probability of the presence of income.

Step 2: Impute income amount.

First, a regression model is estimated for each income type that predicts the given income amount using the donor cases. Second, the donors and the recipients are split into into hot deck cells using the predicted values from the regression. Susin (2005) suggested forming cells with about 500 records to ensure that each cell had plenty of donors. Third, the recipients are imputed with the hot deck using the cells formed by the predicted values. Since the records are also sorted with respect to geography, the last donor will typically be geographically close to the recipient.

Susin (2005) suggests "triangular regression-based hot deck" which allows the inclusion of predictive variables that are subject to non-response. Triangularity indicates a set of sequential, triangular equations. For example, receipt of the first income type is imputed based on only the selected regressors such as age, sex, and region. Receipt of the second income type is imputed using the regressors and the first income type and so on. Triangular regression overcomes a problem of circular dependence since all the income variables may have missing values. That is, "social security cannot be imputed until earning incomes are computed and vise versa." For completeness, we mention that the models of Step 1 are not triangular in that they do not use the previously imputed variables. A link to the paper describing the methodology can be found in the References section of this document.

Exhibit 3.7.1 Person Income Subtopic

Variable	Description	Imputation Method	Availability
INTP	Person's interest, dividends, and net rental income (past 12 months)	Hot deck imputation based on a regression model using demographic variables, work status, and various others.	PUF
OIP	Person's other income (past 12 months)	Hot deck imputation based on a regression model using demographic variables, work status, and various others.	PUF

PAP	Person's public assistance income (past 12 months)	Hot deck imputation based on a regression model using demographic variables, work status, and various others.	PUF
RETP	Person's retirement income (past 12 months)	Hot deck imputation based on a regression model using demographic variables, work status, and various others.	PUF
SEMP	Person's self-employment income (past 12 months)	Hot deck imputation based on a regression model using demographic variables, work status, and various others.	PUF
SSIP	Person's Supplemental Security Income (past 12 months)	Hot deck imputation based on a regression model using demographic variables, work status, and various others.	PUF
SSP	Person's Social Security income (past 12 months)	Hot deck imputation based on a regression model using demographic variables, work status, and various others.	PUF
WAGP	Person's wages or salary income (past 12 months)	Hot deck imputation based on a regression model using demographic variables, work status, and various others.	PUF

Exhibit 3.7.2 Total Household Income Subtopic

Variable	Description	Imputation Method	Availability
FINCP	Family income (past 12 months)	Recoded using imputed inputs (See INTP, OIP, PAP, RETP, SEMP, SSIP, SSP, WAGP)	PUF
HINCP	Household income (past 12 months)	Recoded using imputed inputs (See INTP, OIP, PAP, RETP, SEMP, SSIP, SSP, WAGP)	PUF

3.8 Mortgage Details Topic

Starting in 2021 HUD and Census redesigned and streamlined the mortgage details content. This impacted the list of variables that could be publicly released. Specifically, variables AMMORT and PMTONLY and their imputations have been part of internal processing for many years. However, 2021 is the first time they appeared on the PUF in recent cycles.

HUD and Census also introduced a new imputation for predicting whether a unit is mortgaged or not at the time of interview for owner-occupied households that do not report this information. This resulted in the creation of new loans on the mortgage data set. See the MORTSTAT Imputation section of the Housing Costs topic for more information on this process and its impact on mortgage details variables.

Exhibit 3.8.1 Mortgage Origination Subtopic

Variable	Description	Imputation Method	Availability
INTRATE	Interest rate of mortgage	Hot deck imputation using number of years paid on mortgage.	PUF
MORTTERM	Term of mortgage	Hot deck imputation using number of years paid on mortgage.	IUF Only
AMMORT	Initial amount borrowed or refinanced for the mortgage	Hot deck imputation using the amount borrowed as a percentage of the current market value of the unit.	PUF

Exhibit 3.8.2 Current Payment Details Subtopic

Variable	Description	Imputation Method	Availability
PMTAMT	Amount of mortgage payment	Deductive imputation.	PUF
PMTONLY	Amount of mortgage payment paid towards principal and interest only	Deductive imputation.	PUF

3.9 Neighborhood Features Topic

Exhibit 3.9.1 General Subtopic

Variable	Description	Imputation Method	Availability
TPARK	Number of mobile homes in group	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	IUF Only

3.10 Occupancy and Tenure Topic

Exhibit 3.10.1 Tenure Subtopic

Variable	Description	Imputation Method	Availability
CONDO	Flag indicating unit is part of a condominium	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race). This variable is an input to the housing adequacy variable.	PUF
COOP	Flag indicating unit is part of a cooperative	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	IUF Only
MGRONSITE	Flag indicating there is an owner or property manager onsite	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
TENURE	Owner or renter status of unit	Deductive imputation.	PUF

Exhibit 3.10.2 Vacancy Characteristics Subtopic

Variable	Description	Imputation Method	Availability
VACANCY	Vacancy status	Deductive imputation. This variable is an input to the housing adequacy variable.	PUF
VACMONTHS	Number of months unit has been vacant	Hot deck imputation using structure type and vacancy status..	PUF

3.11 Structural Topic

Exhibit 3.11.1 Exterior Features Subtopic

Variable	Description	Imputation Method	Availability
LOTSIZE	Lot size (acres)	Cold deck using administrative records if available. Otherwise, hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
LOTSIZE_IUF	Lot size (acres)	Cold deck using administrative records if available. Otherwise, hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	IUF Only
PORCH	Flag indicating unit has porch or deck or balcony or patio	Deductive imputation.	PUF

Exhibit 3.11.2 General Subtopic

Variable	Description	Imputation Method	Availability
BLD	Type of housing unit	Hot deck imputation based on number of floors, tenure, and vacancy status.	PUF
NUNITS	Number of units in building	Hot deck imputation based on number of floors, tenure, and vacancy status.	IUF Only
YRBUILT	Year unit was built	Cold deck if a value is available from administrative records. Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race).	PUF
YRBUILT_IUF	Year unit was built	Cold deck if a value is available from administrative records. Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, vacancy status, and demographic information (e.g., age, race).	IUF Only

Exhibit 3.11.3 Interior Features Subtopic

Variable	Description	Imputation Method	Availability
BATHEXCLU	Flag indicating the unit's bathroom facilities are for the exclusive use of the household	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race). This variable was not imputed prior to 2017. This variable is an input to the housing adequacy variable.	PUF
BATHROOMS	Number of bathrooms in unit	The components of this recoded variable are imputed using 2 different hot decks. One has bins defined by structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race). The other has bins defined by structure type, number of occupants, tenure, and vacancy status. This variable is an input to the housing adequacy variable.	PUF
BEDROOMS	Number of bedrooms in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	PUF
DENS	Number of dens or libraries in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	IUF Only
DINING	Number of dining rooms in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	PUF
FAMROOMS	Number of family rooms, great rooms, or TV rooms in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	IUF Only
FINROOMS	Number of other finished rooms in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	PUF
FOUNDTYPE	Type of foundation	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
KITCHENS	Number of kitchens in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	PUF
LAUNDY	Number of laundry or utility rooms in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	PUF
LIVING	Number of living rooms in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	IUF Only
OTHFN	Number of other finished rooms in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	IUF Only
RECROOMS	Number of recreation rooms in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	IUF Only

STORIES	Number of stories in building	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF
STORIES_IUF	Number of stories in building	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	IUF Only
TOTROOMS	Number of rooms in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	PUF
UFINROOMS	Number of other unfinished rooms in unit	Hot deck imputation based on structure type, number of occupants, tenure, and vacancy status.	PUF
UNITFLOORS	Number of floors in unit	Hot deck imputation using structure type, tenure, property value or rent, number of bedrooms, and demographic information (e.g., age, race).	PUF

4. References

Susin, Scott. 2005. “Imputation via Triangular Regression-Based Hot Deck.” U.S. Census Bureau. https://www.census.gov/programs-surveys/ahs/research/publications/hotdeck.html.

U.S. Department of Housing and Urban Development

Office of Policy Development and Research

Washington, DC 20410-6000