Analysis of Housing Square Footage Estimates Reported by the American Housing Survey and the Survey Of Construction

Angela L. Delano

U.S. Census Bureau
4600 Silver Hill Road, Washington DC 20233
Angela.Lynn.Delano@census.gov

Disclaimer: *Any views expressed on (statistical, methodological, technical, or operational) issues are those of the author(s) and not necessarily those of the U.S. Census Bureau.*

Introduction

Government agencies, policy analysts, financial institutions, and manufacturers, to name a few, use data collected by the U.S. Census Bureau. This data helps measure and evaluate size, composition, and change occurring within the inventory of residential housing units available in the United States. One important measurement associated with this inventory relates to housing unit size. The size, or square footage, of a housing unit and the corresponding change of this measurement over time plays an important role in costs associated with buying residential homes (Sirmans & Macpherson, 2003) and heating residential homes (Energy Information Administration, 2012). Up until now there have been no studies analyzing housing square footage estimates reported by surveys conducted by U.S. Census Bureau.

This paper will examine housing square footage estimates reported by two U.S. Census Bureau conducted surveys – the American Housing Survey (AHS) and the Survey of Construction (SOC). These surveys are funded either in full or part by the Department of Housing and Urban Development (HUD). First, I will present a brief overview of the surveys. Next, I will discuss methods used to analyze housing square footage estimates reported by the surveys. Discussion of housing square footage estimates reported by the surveys will follow. Next, I will discuss issues associated with the collection and conversion of housing square footage estimates. These issues include differences between the surveys regarding sample sizes, survey designs, and square footage definitions. In addition, discussion of conversion issues specifically related to the SOC will also take place. Together these issues may suggest reasons for housing square footage differences reported by the AHS and the SOC. Finally, I will discuss updated square footage analysis results and offer proposed enhancements to the collection of housing square footage estimates by the AHS and the SOC.

AHS Background

AHS Survey

The AHS is the most comprehensive national housing survey in the U.S. The survey provides national and regional data on a wealth of topics related to the quality of U.S. housing. Topics covered by the survey include basic house characteristics (such as the number of bedrooms, bathrooms, and square footage), data on physical conditions, occupant demographics, and financial data such as home improvements, costs, and mortgages to name a few. The survey is sponsored by HUD and conducted by the U.S. Census Bureau.

AHS Design

The AHS is a longitudinal housing unit survey conducted biennially between May and September of odd-numbered years. The most recent AHS took place in 2011. The 2011 AHS consisted of a national sample of housing units and a supplemental sample of housing units selected from predefined metropolitan areas.

AHS Sample

Housing units selected for the AHS represent a cross section of all U.S. housing. Updates to the survey sample occur prior to each survey by adding newly constructed housing units and units discovered through coverage improvement efforts. Once selected, the survey returns to the same housing units every other year for data collection. Respondents for the survey consist of occupants of existing housing units as well as property owners of vacant housing units.

SOC Background

SOC Survey

The SOC is one of the most comprehensive new, privately-owned, residential housing survey in the U.S. The survey provides national and regional data on a range of topics. Topics covered by the survey include start and completion dates, house characteristics (such as the number of bedrooms, bathrooms, and square footage), and sales data (such as the sales date, pricing, and financing). Sales data provided only for new single-family houses. The SOC is partially funded by HUD and conducted by the U.S. Census Bureau.

SOC Design

The SOC is a housing unit survey conducted on a continuous monthly basis. The SOC consists of new, privately-owned, residential housing units sampled from predefined permit issuing and predefined non-permit issuing areas.

SOC Sample

Housing units selected for the SOC represent a cross section of all new, privately-owned, residential housing units constructed in the U.S. Each month a sample of new, privately-owned, residential housing units are created through one of two methods. The first method includes sampling of building permits in predetermined areas requiring a building permit to authorize new housing unit construction. Surveying of these predetermined permit issuing areas takes place on a monthly basis. The second method includes physical canvassing of predetermined areas that do not require a building permit to authorize new housing unit construction. Surveying of these predetermined non-permit issuing areas takes place on a monthly or quarterly basis. Reselection of new predetermined permit and non-permit issuing areas takes place approximately every 20 years.

Once a housing unit is selected, the survey returns to the same housing unit approximately every month from authorization of construction (or discovery via canvassing) through completion (or occupancy) and, if necessary, sale (for single-family houses). Respondents for the survey consist of builders, owners, and sales agents of new, privately-owned, residential housing units.

The AHS and the SOC are sample surveys. Estimates based on sample surveys may differ from estimates obtained from a complete census. The accuracy of an estimate based on a sample survey is subject to the joint effects of both sampling errors and nonsampling errors. For the purposes of this paper, I will not discuss these errors. See the Reference section for additional information regarding these surveys.

Methods

Data Sources

This paper examines housing square footage estimates reported by the AHS and the SOC. The data sources used to obtain these estimates consist of the 2011 AHS National/Metropolitan Public Use File (or PUF, version 1.2) and 2001-2013 SOC Internal Microdata Files.

AHS Data

The AHS PUF contains 186,448 observations. Analysis of housing square footage estimates occurred upon a subset of these observations. For the purposes of this paper, only detached, single housing unit buildings built from 2002-2011 with between 100-13,000 square feet were included in the analysis. Additional variables relating to geography (U.S., Northeast, Midwest, South, and West Region geographic levels), the type of foundation (basement, crawl space, slab, etc.), number of floors (1, 2, 3+), square footage (both originally reported and additional changes, if available), and weights were also taken into consideration. Excluded from analysis are observations containing missing data or square footage data outside 100-13,000 square feet. The final subset contained 10,212 observations.

SOC Data

Upon initial creation, the 2001-2013 SOC Internal Microdata File contained 674,024 observations. Analysis of housing square footage estimates occurred upon a subset of these observations. For the purposes of this paper, only detached, single housing unit buildings completed from 2002-2011 with between 100-13,000 square feet were initially included in the analysis. Additional variables relating to "geography" (U.S., Northeast, Midwest, South, and West Region geographic levels), the type of foundation (basement, crawl space, slab, etc.), type of basement (finished or unfinished), number of floors (Split Level, 1, 2, 3+), square footage, square footage dimensions (interior or exterior), and weights were also taken into consideration. Excluded from analysis are observations containing missing data or square footage data outside 100-13,000 square feet. The final subset contained 190,233 observations.

Square Footage Estimates and Standard Errors

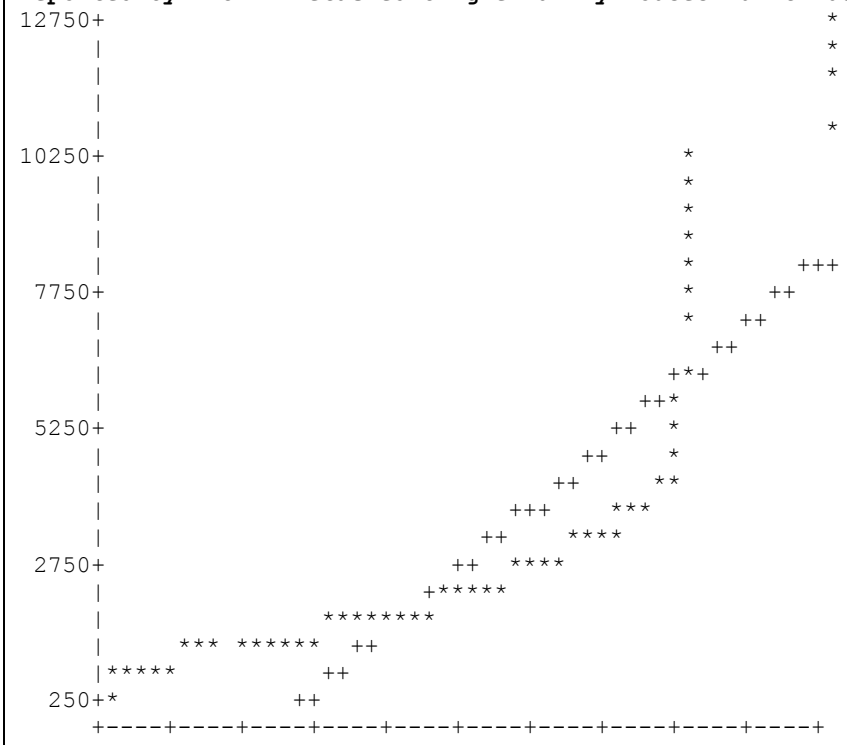
Weighted mean square footage estimates, and their associated standard errors, reported between the AHS and the SOC were calculated. Calculation of the estimates occurred at the U.S. and Regional geographic levels by year built. Corresponding estimates at lower geographic levels are not available. The SAS UNIVARIATE procedure produced the estimates and associated standard errors. See the Reference section for additional information regarding this procedure.

Significance Testing

Commonly used parametric statistical testing procedures assume data are independent and normally distributed. As the AHS and SOC are independent surveys, attention shifted to the data distributions. Normally distributed data should appear as a straight diagonal line on a probability plot. Figure 1 and Figure 2 on the following page display weighted normal probability plots for both the AHS and the SOC mean square footage in detached single-family houses built from 2002-2011, respectively. Both sets of data depart from this standard, indicating the data are not normally distributed. Thus, parametric statistical testing procedures are inappropriate.

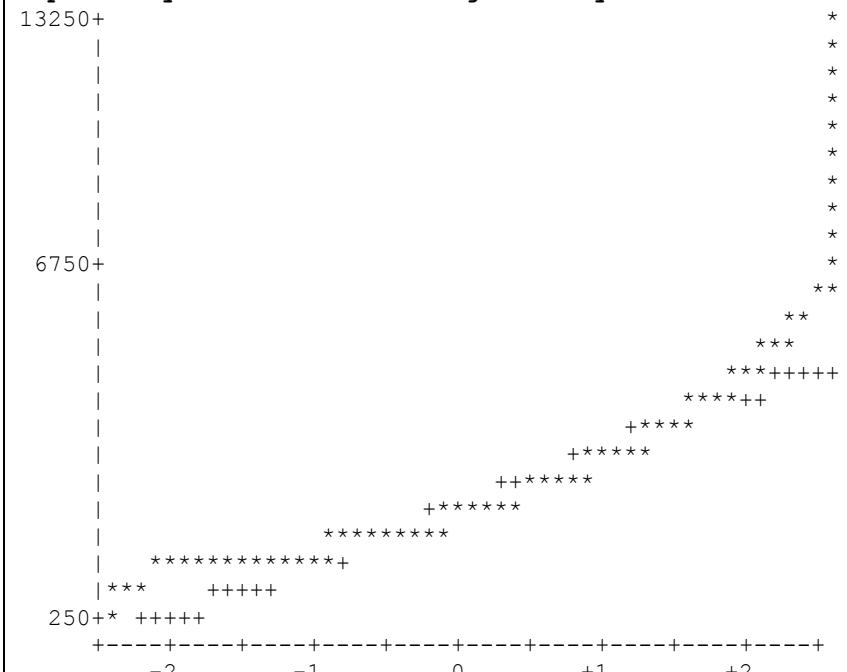
This paper employed the Wilcoxon Rank Sum non-parametric statistical test to test for statistical significance (Pappas & DePuy, 2004). The assumptions for this test require the data be independent and identically distributed, and that the shapes and spreads of the distributions are the same. Both the AHS and the SOC meet these conditions. Results from the Wilcoxon Rank Sum test will report on statistically significant differences (if present) between the underlying housing square footage distributions of the AHS and the SOC at a 90% confidence level.

Figure 1: Weighted Normal Probability Plot of Mean Square Footage Reported by AHS in Detached Single-Family Houses Built 2002-2011



Source: 2011 AHS National/Metropolitan Public Use File (v 1.2); Cases with missing data or square footage outside 100-13,000 removed from analysis

Figure 2: Weighted Normal Probability Plot of Mean Square Footage Reported by SOC in Detached Single-Family Houses Built 2002-2011



Source: 2001-2013 SOC Internal Microdata Files; Cases with missing data or square footage outside 100-13,000 removed from analysis

Housing Square Footage Estimates

Table 1 shows the number of cases analyzed for the AHS, the mean square footage reported, and associated standard errors in detached single-family houses built from 2002-2011 at the U.S. and Regional geographic levels.

Table1: Mean Square Footage Reported by AHS[1] in Detached Single-Family Houses

	Cases Analyzed					Square Footage					Standard Errors				
Year Built	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West
2002	1,283	69	307	484	423	2,942	3,952	3,076	2,811	2,760	2.2%	10.0%	4.5%	3.7%	3.7%
2003	1,336	80	294	546	416	2,849	3,045	2,905	2,876	2,676	2.1%	8.5%	4.6%	3.5%	2.7%
2004	1,386	79	311	543	453	2,736	3,480	3,021	2,676	2,398	1.9%	8.4%	4.4%	3.1%	2.7%
2005	1,693	91	318	727	557	2,854	3,031	3,186	2,697	2,826	1.9%	8.9%	4.5%	2.8%	3.0%
2006	1,510	81	279	710	440	2,900	3,356	3,130	2,799	2,827	1.9%	9.4%	4.8%	2.7%	3.5%
2007	1,171	54	208	572	337	2,970	4,087	3,531	2,679	3,009	2.4%	10.8%	5.8%	3.2%	4.5%
2008	793	64	129	367	233	2,995	2,783	2,885	3,050	3,043	2.8%	10.4%	7.6%	4.0%	4.7%
2009	465	38	97	204	126	2,958	3,882	2,785	3,044	2,591	3.8%	14.0%	8.6%	5.1%	8.2%
2010	424	30	96	199	99	2,502	2,752	2,055	2,624	2,523	3.8%	16.3%	4.8%	5.7%	8.0%
2011	151	6	39	68	38	3,281	3,106	3,985	3,190	2,802	6.0%	16.8%	12.8%	8.2%	12.5%
2002-11	**10,212**	**592**	**2,078**	**4,420**	**3,122**	**2,887**	**3,356**	**3,071**	**2,790**	**2,755**	**0.8%**	**3.4%**	**1.8%**	**1.1%**	**1.3%**

[1]Source: 2011 AHS National/Metropolitan Public Use File (v 1.2); Cases with missing data or square footage outside 100-13,000 removed from analysis

The overall mean square footage reported by the AHS over this time period ranges from 2,755 square feet in the West to 3,356 square feet in the Northeast. The associated standard errors are generally below 10% except for year 2011 and in the Northeast. Note the decreasing sample size in the Northeast over this time corresponds to an increasing standard error.

Similarly to Table 1, Table 2 shows the number of cases analyzed for the SOC, the mean square footage reported, and associated standard errors in detached single-family houses built from 2002-2011 at the U.S. and Regional geographic levels.

Table 2: Mean Square Footage Reported by SOC[1] in Detached Single-Family Houses

	Cases Analyzed					Square Footage					Standard Errors				
Year Built	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West
2002	24,029	2,931	4,792	10,575	5,731	2,392	2,676	2,330	2,370	2,389	0.3%	0.7%	0.6%	0.4%	0.5%
2003	23,584	2,658	4,616	10,416	5,894	2,406	2,617	2,338	2,402	2,406	0.2%	0.8%	0.6%	0.4%	0.5%
2004	24,383	2,785	5,020	10,624	5,954	2,422	2,684	2,352	2,414	2,418	0.2%	0.7%	0.6%	0.4%	0.5%
2005	23,714	2,541	4,797	10,677	5,699	2,519	2,718	2,438	2,534	2,498	0.3%	0.8%	0.7%	0.4%	0.6%
2006	22,854	2,036	4,229	10,932	5,657	2,569	2,850	2,419	2,578	2,576	0.3%	1.0%	0.7%	0.4%	0.5%
2007	20,144	1,652	3,411	10,087	4,994	2,627	2,777	2,451	2,655	2,629	0.3%	1.2%	0.8%	0.5%	0.6%
2008	15,816	1,255	2,763	8,084	3,714	2,635	2,847	2,404	2,673	2,649	0.4%	1.4%	1.0%	0.5%	0.8%
2009	11,556	1,014	2,009	5,939	2,594	2,548	2,761	2,355	2,565	2,569	0.5%	1.6%	1.1%	0.6%	0.9%
2010	12,621	1,070	2,362	6,515	2,674	2,451	2,778	2,327	2,438	2,452	0.4%	1.6%	1.0%	0.5%	0.9%
2011	11,532	980	2,096	5,923	2,533	2,566	2,728	2,409	2,612	2,505	0.4%	1.6%	1.0%	0.6%	0.8%
2002-11	**190,233**	**18,922**	**36,095**	**89,772**	**45,444**	**2,506**	**2,736**	**2,384**	**2,519**	**2,500**	**0.1%**	**0.3%**	**0.3%**	**0.2%**	**0.2%**

[1]Source: 2001-2013 SOC Internal Microdata Files; Cases with missing data or square footage outside 100-13,000 removed from analysis

For the SOC, the overall mean square footage reported over this time period ranges from 2,500 square feet in the West to 2,736 square feet in the Northeast. The associated standard errors are consistently below 2%. These reported square footages and associated standard errors are lower than those reported by the AHS. Note that the SOC sample sizes are higher than those from the AHS. Additional discussion regarding sample sizes will take place later.

Turning our attention to the comparison between these two surveys, Table 3 shows the percent differences (calculated by the difference between the AHS and the SOC estimates with respect to the SOC estimate) and the associated statistical significances in detached single-family houses built from 2002-2011 at the U.S. and Regional geographic levels.

Table 3: Percent Difference in Mean Square Footage Reported by AHS [1] and SOC [2] in Detached Single-Family Houses

	Percent Difference (AHS-SOC / SOC)					Statistical Significance Probabilities				
Year Built	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West
2002	23%	48%	32%	19%	16%	0.0001 *	0.5890	0.0001 *	0.6688	0.0724 *
2003	18%	16%	24%	20%	11%	0.0208 *	0.9566	0.0001 *	0.9028	0.0656 *
2004	13%	30%	28%	11%	-1%	0.0001 *	0.1954	0.0001 *	0.0074 *	0.0881 *
2005	13%	12%	31%	6%	13%	0.0840 *	0.1886	0.0001 *	0.5475	0.4971
2006	13%	18%	29%	9%	10%	0.6327	0.3874	0.0001 *	0.0989 *	0.7870
2007	13%	47%	44%	1%	14%	0.3750	0.2690	0.0001 *	0.0010 *	0.2474
2008	14%	-2%	20%	14%	15%	0.8933	0.0553 *	0.0068 *	0.0493 *	0.2580
2009	16%	41%	18%	19%	1%	0.4338	0.9168	0.0039 *	0.7406	0.1631
2010	2%	-1%	-12%	8%	3%	0.8308	0.3519	0.2298	0.8253	0.5046
2011	28%	14%	65%	22%	12%	0.0981 *	0.8337	0.0145 *	0.8734	0.3485
2002-11	**15%**	**23%**	**29%**	**11%**	**10%**	**0.0001 ***	**0.6033**	**0.0001 ***	**0.2407**	**0.0389 ***

[1] Source: 2011 AHS National/Metropolitan Public Use File (v 1.2); Cases with missing data or square footage outside 100-13,000 removed from analysis

[2] Source: 2001-2013 SOC Internal Microdata Files; Cases with missing data or square footage outside 100-13,000 removed from analysis

* Statistically significant differences in the underlying square footage distributions of the AHS and the SOC at a 90% confidence level

The majority of percent differences between the AHS and the SOC are positive. This indicates the AHS mean square footage is larger than the corresponding SOC square footage. Also, note how the statistical significance probabilities are more significant for earlier years versus recent years. These results indicate there are statistically significant differences in the underlying housing square footage distributions between the two surveys.

This poses an interesting question. Why are there more statistically significant differences between the surveys in earlier years as opposed to recent years? Can you compare estimates obtained from the surveys? Well the answer is yes! However to accomplish this one must take into account issues associated with housing square footage estimates.

Housing Square Footage Issues

This paper will discuss four issues associated with housing square footage estimates. These issues include differences between the AHS and the SOC regarding sample sizes, survey designs, and square footage definitions. An additional issue also includes conversion issues specifically related to the SOC. Together these issues suggest reasons for housing square footage estimate differences reported between the surveys.

Sample Size Issues

Table 4 on the following page shows sample sizes used from the AHS and the SOC to analyze mean square footage reported in detached single-family houses built from 2002-2011 at the U.S. and Regional geographic levels.

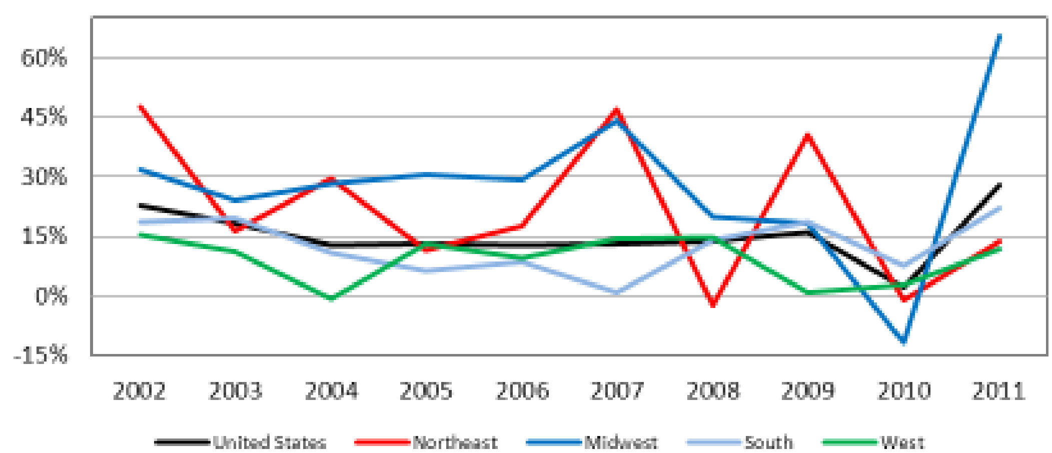
Overall, the AHS contains far less observations than the SOC (10,212 AHS observations to 190,233 SOC observations). Within the AHS, sample sizes for the Northeast geographical area (592 observations) and for 2011 as a whole (151 observations) are particularly small. It takes a while to update the AHS sample with new residential construction. This lag in updating the AHS sample with new residential construction limits available comparisons over recent years to the SOC (which is entirely comprised of new residential construction).

To address sample size issues, updated square footage analysis will be limited to detached single-family houses built from 2002-2010. Excluded from analysis are data from 2011.

Survey Design Issues

Figure 3 charts the mean square footage percent difference (calculated by the difference between the AHS and the SOC estimate with respect to the SOC estimate) in detached single-family houses built from 2002-2011 at the U.S. and Regional geographic levels.

Figure 3: Mean Square Footage Percent Difference (AHS[1] - SOC[2]/ SOC) in Detached Single-Family Houses Built 2002-2011



[1]Source: 2011 AHS National/Metropolitan Public Use File (v 1.2); Cases with missing data or square footage outside 100-13,000 removed from analysis
[2]Source: 2001-2013 SOC Internal Microdata Files; Cases with missing data or square footage outside 100-13,000 removed from analysis

Table 4: Cases A

Year Built	United States									
2002	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2003	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2004	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2005	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2006	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2007	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2008	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2009	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2010	424	30	96	199	99	12,621	1,070	2,362	6,515	2,674
2011	151	6	39	68	38	11,532	980	2,096	5,923	2,533
2002-11	**10,212**	**592**	**2,078**	**4,420**	**3,122**	**190,233**	**18,922**	**36,095**	**89,772**	**45,444**

[1]Source: 2011 AHS National/Metropolitan Public Use File (v 1.2); Cases with missing data or square footage outside 100-13,000 removed from analysis

[2]Source: 2001-2013 SOC Internal Microdata Files; Cases with missing data or square footage outside 100-13,000 removed from analysis

The pattern of alignment for years 2002-2006 (earlier years) is similar, with almost all percentages showing positive differences. Percent differences for year 2002 are also higher than for years 2003-2006. Positive differences indicate the square footage reported by the AHS is higher than the corresponding SOC square footage. This makes sense given the AHS is a longitudinal survey. The AHS interviews occupants of existing housing units every other year, and square footages may in fact change over time (for example, if an addition to a house is completed).

Patterns for years 2007-2011 (recent years) are more varied. Remember the SOC is a monthly survey. The SOC interviews builders, owners, and/or sales agents of new, privately-owned, residential housing units from authorization through completion (or occupancy). For the SOC, square footage is a snapshot taken at the time a house is completed. So in recent years, single-family detached houses are relatively newer, thus the square footage may be more likely to match up to the square footage available immediately after house construction. This difference in survey design limits available comparisons over earlier years between the surveys as the available square footage in existing housing units change, due to construction of additions, for example.

To address survey design issues, updated square footage analysis will be limited to detached single-family houses built from 2007-2010. In addition, updated square footage analysis will obtain original AHS square footage estimates where applicable.

Square Footage Definition Issues

The housing square footage definitions used by the AHS and the SOC are different. The AHS and the SOC handle basements, and their associated square footages, differently. Specifically:

AHS housing square footage: *"…excluded from the calculation of square footage are unfinished attics, carports, attached garages, porches that are not protected from weather (such as screened porches), and mobile home hitches. Both finished and unfinished basements are included. Square footage is based on respondent's estimate of the size of the unit."*

SOC housing square footage: *"Square footage of floor area, defined as all completed finished floor space, including space in basements and attics with finished walls, floors, and ceilings. This does not include a garage, carport, porch, unfinished attic or utility room. Unfinished areas of the basement are excluded."*

There are also differences in how the surveys capture basement information. The AHS only captures presence of a basement as part of the physical house structure. Whether or not the basement is finished does not influence collection of the square footage, as both finished and unfinished areas of the basement should be included in the square footage. The SOC too captures presence of a basement as part of the physical house structure. Where a basement is identified, the SOC proceeds to capture whether the basement is finished (all or part) or unfinished. Only finished areas of a basement should be included in the square footage.

In essence, the AHS is capturing living space (where all areas of the basement are included) while the SOC is capturing floor space (where only finished areas of the basement are included). Based on these definitional differences, comparisons between the surveys must take into account both the presence and type of basement.

To address square footage definitional issues, updated square footage analysis will assume unfinished housing square footage is primarily located in the basement. Analysis will break out structures based upon presence and type of basement.

External Conversion Issues

As noted above, the AHS collects and estimates housing square footage *"…based on respondent's estimate of the size of the unit."* This respondent varies by survey. Remember the AHS respondent consists of mainly occupants or property owners. Occupants in particular may be estimating the size of the unit based upon memory, thus increasing

difficulty in obtaining accurate estimates. The SOC respondent, on the other hand, consists of the builder, owner, or sales agent. These respondents may have access to more accurate information in the way of construction blue prints or floor plans, thus decreasing difficulty in obtaining accurate estimates.

In addition, the SOC also collects the dimension (interior, exterior, or unknown) upon which the square footage is based. For estimating, the SOC converts all reported square footages to exterior dimensions through the following procedures:

- Square footages based on exterior dimensions not increased.
- Square footages based on interior dimensions increased 8%.
- Square footages based on unknown dimensions increased 4%.

For example, if the AHS respondent reports the square footage as 2,000 square feet, the AHS estimate is 2,000 square feet. If the SOC respondent reports the square footage as 2,000 square feet based on interior dimensions, the SOC estimate is 2,160 square feet.

For more information, Table 5 shows the number of cases analyzed for the SOC, the number of these cases reporting square footages based on exterior dimensions, and the resulting percentage of cases being converted to exterior dimensions for detached single-family houses built from 2002-2011 at the U.S. and Regional geographic levels.

At the U.S. geographic level, the percentage of cases converting to exterior dimensions has climbed steadily from 62% in 2002 to 72% in 2011. This conversion process in effect serves to increase the housing square footage estimates reported by the SOC, thus potentially altering comparisons between the surveys.

Table 5: Cases Analyzed for Mean Square Footage Reported in Detached Single-Family Houses

	Cases Analyzed for SOC[1]					Cases Reporting Exterior Dimensions					Percent Cases Converting Dimensions				
Year	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West
2002	24,029	2,931	4,792	10,575	5,731	9,011	1,074	2,152	3,973	1,812	62%	63%	55%	62%	68%
2003	23,584	2,658	4,616	10,416	5,894	8,510	1,025	2,063	3,610	1,812	64%	61%	55%	65%	69%
2004	24,383	2,785	5,020	10,624	5,954	9,146	1,167	2,282	3,959	1,738	62%	58%	55%	63%	71%
2005	23,714	2,541	4,797	10,677	5,699	8,232	1,039	2,055	3,642	1,496	65%	59%	57%	66%	74%
2006	22,854	2,036	4,229	10,932	5,657	6,775	689	1,405	3,467	1,214	70%	66%	67%	68%	79%
2007	20,144	1,652	3,411	10,087	4,994	5,345	512	966	2,857	1,010	73%	69%	72%	72%	80%
2008	15,816	1,255	2,763	8,084	3,714	4,323	425	765	2,246	887	73%	66%	72%	72%	76%
2009	11,556	1,014	2,009	5,939	2,594	3,002	283	527	1,642	550	74%	72%	74%	72%	79%
2010	12,621	1,070	2,362	6,519	2,674	3,421	311	584	1,847	679	73%	71%	75%	72%	75%
2011	11,532	980	2,096	5,923	2,533	3,194	296	565	1,634	699	72%	70%	73%	72%	72%
2002-11	**190,233**	**18,922**	**36,095**	**89,776**	**45,444**	**60,959**	**6,821**	**13,364**	**28,877**	**11,897**	**68%**	**64%**	**63%**	**68%**	**74%**

[1]Source: 2001-2013 SOC Internal Microdata Files; Cases with missing data or square footage outside 100-13,000 removed from analysis

To address external conversion issues, updated square footage analysis will assume the SOC housing square footage estimate is collected directly from the respondent (similarly to the AHS) with no conversion to exterior dimensions applied.

Updated Square Footage Analysis

Now I will rerun the original square footage analysis taking into account the assumptions noted above when addressing issues associated with housing square footage. To recap, the updated analysis will:

- Remove recent year (2011) comparisons given lag in updating AHS sample with new residential construction.
- Remove earlier year (2002-2006) comparisons given available square footage changes over time (i.e. additions).
- Break out basement presence and type given differences in how surveys define square footage.
- Remove exterior square footage conversion applied to the SOC square footage estimates.

Table 6 shows the updated number of cases analyzed for the AHS, the mean square footage reported, and associated standard errors in detached single-family houses by basement presence built from 2007-2010 at the U.S. and Regional geographic levels.

First, note there are not as many detached single-family houses with basements as there are houses without basements in the AHS (827 observations with a basement to 2,026 observations without a basement). There are also a limited number of houses without basements in the Northeast (20 observations). Remember smaller sample sizes correspond to larger standard errors.

Table 6: Mean Square Footage Reported by AHS[1] in Detached Single-Family Houses by Basement

	Cases Analyzed					Square Footage					Standard Errors				
Year	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West
With Basement															
2007	307	48	156	61	42	3,958	4,387	3,513	3,883	5,050	4%	11%	6%	10%	12%
2008	239	57	102	44	36	3,614	3,061	3,186	4,378	4,540	5%	10%	8%	10%	11%
2009	148	35	73	28	12	3,709	3,653	3,261	4,226	3,890	6%	14%	9%	11%	24%
2010	133	26	76	19	12	3,036	3,081	2,229	4,596	3,555	7%	17%	5%	19%	25%
2007-10	**827**	**166**	**407**	**152**	**102**	**3,686**	**3,599**	**3,170**	**4,175**	**4,514**	**3%**	**6%**	**4%**	**6%**	**8%**
No Basement															
2007	864	6	52	511	295	2,623	2,422	3,567	2,481	2,709	3%	14%	14%	3%	4%
2008	554	7	27	323	197	2,681	1,371	1,738	2,752	2,798	3%	22%	9%	4%	5%
2009	317	3	24	176	114	2,590	6,075	1,613	2,733	2,429	5%	54%	13%	6%	9%
2010	291	4	20	180	87	2,279	1,393	1,472	2,375	2,265	4%	7%	6%	5%	7%
2007-10	**2,026**	**20**	**123**	**1,190**	**693**	**2,590**	**2,265**	**2,658**	**2,568**	**2,640**	**2%**	**23%**	**10%**	**2%**	**3%**

[1]Source: 2011 AHS National/Metropolitan Public Use File (v 1.2); Cases with missing data or square footage outside 100-13,000 removed from analysis
First interview square footage determined where applicable.

Of more interest are the mean square footages themselves. The change in square footage ranges between detached single-family houses with and without a basement is evident. The mean square footages in houses with a basement reported by the AHS over this time are almost entirely over 3,000 square feet. The square footage ranges for houses without a basement are usually smaller. This result clearly shows additional square footage is available in houses with a basement. This may or may not indicate a corresponding increase in house quality.

Table 7 on the following page shows the updated number of cases analyzed for the SOC, the mean square footage reported, and associated standard errors in detached single-family houses by basement presence and type built from 2007-2010 at the U.S. and Regional geographic levels.

Table 7: Mean Square Footage Reported by SOC[1] in Detached Single-Family Houses by Basement

	Cases Analyzed					Square Footage					Standard Errors				
Year	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West
Finished Basement															
2007	1,423	180	587	320	336	3,633	3,641	3,357	3,856	3,836	1%	3%	2%	3%	2%
2008	1,252	121	504	311	316	3,699	3,727	3,253	4,171	3,930	1%	5%	2%	2%	3%
2009	849	105	333	205	206	3,642	3,615	3,235	4,265	3,717	2%	5%	2%	3%	3%
2010	880	85	420	201	174	3,404	3,429	3,281	3,588	3,443	2%	6%	2%	3%	3%
2007-10	**4,404**	**491**	**1,844**	**1,037**	**1,032**	**3,616**	**3,626**	**3,293**	**3,976**	**3,781**	**1%**	**2%**	**1%**	**1%**	**1%**
Unfinished Basement															
2007	4,888	1,172	2,184	839	693	2,410	2,527	2,213	2,855	2,232	1%	1%	1%	2%	1%
2008	3,789	888	1,771	617	513	2,397	2,590	2,148	2,902	2,326	1%	2%	1%	2%	2%
2009	2,732	705	1,245	416	366	2,342	2,535	2,137	2,700	2,193	1%	2%	1%	2%	2%
2010	3,139	808	1,455	470	406	2,303	2,543	2,104	2,478	2,248	1%	1%	1%	2%	2%
2007-10	**14,548**	**3,573**	**6,655**	**2,342**	**1,978**	**2,378**	**2,547**	**2,164**	**2,789**	**2,253**	**0%**	**1%**	**1%**	**1%**	**1%**
No Basement															
2007	13,833	300	640	8,928	3,965	2,411	2,420	1,815	2,448	2,406	0%	3%	1%	0%	1%
2008	10,775	246	488	7,156	2,885	2,403	2,568	1,828	2,446	2,369	0%	3%	2%	1%	1%
2009	7,975	204	431	5,318	2,022	2,305	2,385	1,736	2,347	2,309	1%	4%	2%	1%	1%
2010	8,602	177	487	5,844	2,094	2,217	2,548	1,643	2,262	2,191	0%	5%	2%	1%	1%
2007-10	**41,185**	**927**	**2,046**	**27,246**	**10,966**	**2,361**	**2,475**	**1,770**	**2,401**	**2,352**	**0%**	**2%**	**1%**	**0%**	**0%**

[1]Source: 2001-2013 SOC Internal Microdata Files; Cases with missing data or square footage outside 100-13,000 removed from analysis
No external square footage conversion factor applied.

First, as discussed earlier, the SOC breaks down detached single-family houses not just by basement presence, but also by basement type. That is, whether the basement is finished (partially or fully) or entirely unfinished. It is interesting to note almost 70% of the SOC observations over this time do not contain a basement (41,185 observations with no basement from a total of 60,137 observations). Also notable, where there is a basement, it is more likely to be an unfinished basement (4,404 observations with a finished basement versus 14,548 observations with an unfinished basement). Remember for the SOC only finished areas of a basement should be included in the square footage, unlike for the AHS where all basement areas are included in the square footage.

Again, of more interest are the mean square footages themselves. Based upon the AHS findings discussed earlier, one would expect square footage ranges for detached single-family houses with a finished basement to be higher than for houses without a basement. This is true given that the mean square footages in houses with a finished basement reported by the SOC over this time are entirely over 3,200 square feet. The square footage ranges for houses without a basement are smaller, all less than 2,600 square feet. This result clearly indicates additional square footage is available in houses with a basement.

Of equal (or maybe more) interest, are the mean square footages for detached single-family houses with unfinished basements. Here, the SOC does not capture any area of the basement in the square footage, so in theory one would expect square footage ranges for houses with unfinished basements to be lower than for houses with a finished basement. This is true. However, the results also show a slight difference between houses with an unfinished basement when compared to houses with no basement (2,164 – 2,789 square feet in houses with an unfinished basement over 2007-2010 versus 1,770 – 2,475 square feet in houses with no basement over the same time). This may indicate more of a fundamental difference between the quality of houses with and without basements. At any rate, houses with an unfinished basement as reported in the SOC do not match up well with the AHS (where only the presence of a basement is captured). As a result, further comparison between the surveys will remove this particular SOC breakout.

Finally, let us now turn our attention to the updated comparison between the surveys. Table 8 on the following page shows the percent differences (calculated by the difference between the AHS and the SOC estimates with respect to the SOC estimate) and the associated statistical significance in detached single-family houses by basement presence built from 2007-2010 at the U.S. and Regional geographic levels.

Table 8: Percent Difference in Mean Square Footage Reported by AHS [1] and SOC [2] in Detached Single-Family Houses by Basement

	Percent Difference (AHS - SOC / SOC)					Statistical Significance Probabilities				
Year Built	United States	North-east	Mid-west	South	West	United States	North-east	Mid-west	South	West
	AHS Basement / SOC Finished Basement									
2007	9%	20%	5%	1%	32%	0.0001 *	0.0362 *	0.0848 *	0.0351 *	0.2067
2008	-2%	-18%	-2%	5%	16%	0.0001 *	0.0001 *	0.0483 *	0.0280 *	0.2148
2009	2%	1%	1%	-1%	5%	0.0001 *	0.0474 *	0.4163	0.0026 *	0.1450
2010	-11%	-10%	-32%	28%	3%	0.0001 *	0.0775 *	0.0001 *	0.0795 *	0.3389
2007-10	**2%**	**-1%**	**-4%**	**5%**	**19%**	0.0001 *	0.0001 *	0.0001 *	0.0001 *	0.0233 *
	AHS No Basement / SOC No Basement									
2007	9%	0%	97%	1%	13%	0.9057	0.8686	0.0909 *	0.5422	0.5638
2008	12%	-47%	-5%	13%	18%	0.2209	0.0235 *	0.7675	0.9783	0.0135 *
2009	12%	155%	-7%	16%	5%	0.2798	0.5766	0.7443	0.0895 *	0.9364
2010	3%	-45%	-10%	5%	3%	0.0766 *	0.0142 *	0.1913	0.0867 *	0.1133
2007-10	**10%**	**-8%**	**50%**	**7%**	**12%**	0.0094 *	0.0126 *	0.0408 *	0.1381	0.0059 *

[1] Source: 2011 AHS National/Metropolitan Public Use File (v 1.2); Cases with missing data or square footage outside 100-13,000 removed from analysis
First interview square footage determined where applicable.

[2] Source: 2001-2013 SOC Internal Microdata Files; Cases with missing data or square footage outside 100-13,000 removed from analysis
No external square footage conversion factor applied.

* Statistically significant differences in the underlying square footage distributions of the AHS and the SOC at a 90% confidence level

Note how the percent differences between the surveys for detached single-family houses with basements are now different than those obtained from the initial analysis. Quite a few differences are now smaller (less than 6% in magnitude). This indicates estimates reported from the AHS and the SOC are now more similar. The differences are also no longer exclusively positive, indicating the SOC mean square footage may be larger than the corresponding AHS square footage. This result is interesting as it is unlikely a new housing unit will lose square footage so quickly after construction. In fact, this result may hit more specifically on the difficulty respondents from both surveys have in estimating square footage.

Also, note the statistical significance probabilities. In the initial analysis, most statistical significance took place in earlier years. In the updated analysis, significant differences are noted in recent years, but occurring more often amongst detached single-family houses with basements. These results indicate there are statistically significant differences in the underlying housing square footage distributions between the two surveys for houses with basements. This makes sense given the SOC square footage estimates for houses with a basement may in fact be less than the corresponding AHS estimate. Remember the SOC does not capture unfinished areas of the basement (which may be more likely to remain unfinished due to the presence of a washer, dryer, and/or water heater). These areas should be included in the AHS estimate, thus making the AHS square footage larger than the corresponding SOC square footage.

Proposed Enhancements

In summary, interpretation of housing square footage estimates reported by the AHS and the SOC is not a straightforward process. To interpret these results correctly, addressing issues associated with housing square footage estimates is necessary. These issues are not all encompassing, nor are they constrained to any one particular survey area. From when data are collected – to what types of data are collected – to what, if any, post-processing procedures are applied – the occurrence and impact of issues must be taken into consideration.

This paper, in addition to identifying and addressing housing square footage issues, proposes two enhancements to the AHS and the SOC. If enacted, these enhancements should assist with future comparisons between the surveys. The first enhancement regards data collection for the AHS. The second enhancement regards additional research for the SOC.

Proposed AHS Enhancement – Capture of additional information regarding basement type would assist with comparisons between the AHS and the SOC. Specifically indicating whether a basement is finished (fully or partially) or unfinished, as currently collected in the SOC, would be useful.

Proposed SOC Enhancement – Aside from differences in house structure, conversion to exterior dimensions applied to reporting of housing square footage for the SOC needs further research. For the purposes of this paper, all assumptions addressing housing square footage issues were applied collectively. The individual effect of "marking up" housing square footage estimates reported by the SOC was not addressed individually. Additional research here is necessary.

Acknowledgement

I would like to recognize the following persons who helped make this paper possible. First, I would like to recognize Erica Filipek and members of the Residential Construction Branch for their review and comments. I would also like to recognize Bonnie Kegan for her guidance regarding statistical testing procedures. Thank you!

References

Sirmans, G. Stacy, MacDonald, Lynn, Macpherson, David A., and Zietz, Emily N. "*The Value of Housing Characteristics: A Meta Analysis,"* Paper Presented at the Mid-Year meeting of the American Real Estate and Urban Economics Association, 2005.

U.S. Energy Information Administration. *The Impact of Increasing Home Size on Energy Demand*. Available at: http://www.eia.gov/consumption/residential/reports/2009/square-footage.cfm.
Released: 04/19/12.

U.S. Census Bureau. *New Residential Construction.* Available at: http://www.census.gov/construction/nrc.
Accessed: 10/30/13.

U.S. Census Bureau. *American Housing Survey (AHS).* Available at: http://www.census.gov/housing/ahs.
Accessed: 10/30/13.

SAS Business Analytics and Business Intelligence Software. The UNIVARIATE Procedure. Available at:
http://support.sas.com/documentation/cdl/en/procstat/63104/HTML/default/viewer.htm#procstat_univariate_sect007.htm
Accessed: 09/30/13.

Pappas, Paul A. and DePuy, Venita. "*An Overview of Non-parametric Tests in SAS: When, Why, and How"* Paper TU04 Presented at the SouthEast SAS Users Group Conference (SESUG 2004).