

# **Measuring Aggregate Housing Wealth: New Insights from an Automated Valuation Model**

Joshua H. Gallin, Eric Nielsen, Raven Molloy, Paul Smith, Kamila Sommer

March 2018

## **Abstract**

We construct a new measure of aggregate U.S. housing wealth based on Zillow's Automated Valuation Model (AVM). AVMs offer advantages over other methods because they are based on recent market transaction prices, utilize large datasets which include property characteristics and local geographic variables, and are updated frequently. However, using Zillow's AVM to measure aggregate housing wealth requires overcoming several challenges related to the representativeness of the Zillow sample. We propose methods that address these challenges and generate a new estimate of aggregate U.S. housing wealth from 2001 to 2016. This new measure provides insights into some of the disadvantages of other approaches to measuring housing wealth. Specifically, with respect to the owner valuations typically used in survey data, it appears that homeowners were slow to recognize the drop in housing wealth during the financial crisis and that their estimates of this drop were unrealistically small. At the same time, repeat-sales price indexes appear to overstate the extent of the drop in value between 2006 and 2011 and overstate the recovery thereafter.

Incomplete draft. Please do not cite without permission of the authors.

We thank Zillow for providing the data and for very helpful discussions about its construction, and we thank Max Miller and Hannah Hall for excellent research assistance. All errors remain our own.

The analysis and conclusions set forth here are those of the authors and do not indicate concurrence by other members of the research staff, the Board of Governors, or the Federal Reserve System.

## I. Introduction

Housing wealth is a major component of household balance sheets. According to the 2013 Survey of Consumer Finances, the value of the primary residence represented about two-thirds of a typical household's total assets.<sup>1</sup> Because housing is such a large part of households' balance sheets and is often used to secure a loan, it plays a key role in households' savings and consumption decisions. As a result, changes in housing wealth affect a wide range of macroeconomic outcomes, including consumer spending, economic growth, business cycles, mortgage lending, wealth inequality, economic mobility, business formation, investment in education, geographic mobility, and tax policy.

Nonetheless, housing wealth is quite difficult to measure, which hampers empirical research on the role it plays in the economy. The best measure of a home's value is a price from a recent arm's-length market transaction. But transactions for a given home are typically infrequent, with years or even decades between sales. Moreover, the heterogeneity of housing and the endogeneity of the decision to sell make it problematic to impute values of recently-sold homes to non-transacting homes, all the more so because much of what makes a home unique is unobservable to researchers.

Historically, measures of housing wealth have typically been based on homeowners' reported values in surveys or extrapolated from previous sales using changes in a repeat-sales price index. Both of these methods are known to be flawed in distinct ways. For example, studies have found that owner-reported estimates of house values are biased up on average, perhaps

---

<sup>1</sup> This statistic is calculated as the ratio of the average value of the primary residence to average total assets among households between the 45<sup>th</sup> and 55<sup>th</sup> percentiles of the wealth distribution. Calculation provided by Kevin Moore and Peter Hansen of the Federal Reserve Board.

because owners are overly optimistic. Moreover, owners appear to have difficulty identifying market turning points, causing the bias to fluctuate over the housing cycle.<sup>2</sup> Other studies have shown problems with using repeat-sales price indexes, due in part to the fact that the properties that are sold are not representative of those that are not sold. This bias may also be cyclical, as the degree of difference between transacting and non-transacting homes may shift systematically over the housing cycle.<sup>3</sup> Another issue with using repeat-sales indexes to measure housing wealth is that these indexes do not account for either home improvements or depreciation, which can affect the value of a house substantially.

In this paper, we examine the strengths and weaknesses of measuring housing wealth using an Automated Valuation Model (AVM). Specifically, we use the estimates from an AVM constructed by Zillow, a private real estate and analytics firm.<sup>4</sup> AVMs, which can be loosely thought of as a set of algorithms that combine a large amount of information on a home's characteristics, neighborhood features, nearby sales, and homes listed for sale, offer an alternative for valuing individual homes. Although versions of AVMs have been in use for decades, private firms have recently created much more sophisticated and comprehensive AVMs using very large property-level datasets and machine-learning algorithms to impute values of individual housing units to large swaths of residential real estate in the U.S. This combination of big data and machine learning offers the potential for more accurate estimates of housing values than those based on surveys or repeat-sales indexes.

---

<sup>2</sup> See, for example Bucks and Pence (2006); Goodman and Ittner (1992); Henriques (2013); Kiel and Zabel (1999); Kuzmenko and Timmons (2011); and Chan, Dastrup and Ellen (2016).

<sup>3</sup> See, for example, Case, Pollakowski, and Wachter (1997), Gatzlaff and Haurin (1997), Glennon, Kiefer and Mayock (2016) and Dreiman and Pennington-Cross (2004).

<sup>4</sup> Zillow. 2017. "Custom aggregation of Zillow AVM and Transaction Data: 2017-02." Zillow Group, Inc. <http://www.zillow.com/data/>.

Our objective is to use an AVM to measure aggregate household-sector housing wealth intended for *own-use*, which—by definition—excludes rental investments. This measure is conceptually analogous to the housing wealth series reported in the Federal Reserve’s *Financial Accounts of the United States*, which is often used as an input into a variety of economic models and a data source widely cited by policy-makers and the media.<sup>5</sup> Our new estimates also line up conceptually with other common measures of housing wealth. For example, most surveys only report values of owner-occupied property and other housing units intended for owner-occupancy; therefore, rental investment is typically excluded from housing wealth aggregates constructed from these data.<sup>6</sup>

Constructing our measure of aggregate housing wealth is not as simple as adding up the value estimates of all properties in the Zillow data. First, Zillow’s universe, while very large, still does not cover all residential properties in the US. And for the homes that they do cover, Zillow does not always have enough information to provide precise value estimates. Second, Zillow’s universe includes rental properties, which are not part of our housing wealth concept and which cannot be straightforwardly identified and removed from Zillow’s data. To address these two coverage issues, we calculate the average value of properties in the Zillow AVM data, by location and property type, and use property count data from the American Community Survey (ACS)—which is nationally representative and identifies own-use properties separately from rental units—to gross up each segment. That is, we use the ACS to measure the quantity of own-

---

<sup>5</sup> See *Financial Accounts* table B.101, line 14, which reports \$24.5 trillion in aggregate household-sector housing wealth for 2017q4. This measure does not include the value of rental properties, which in the *Financial Accounts* are included in the business sector. Note that we focus on the total value of real estate assets rather than home equity, which would subtract mortgage debt.

<sup>6</sup> One notable exception is the Survey of Consumer Finances where survey respondents report the amount of rental investment owned by a household.

use housing units by property type and geographical location, and we use the Zillow data to measure the average value of homes in each market segment.

We next compare estimates from our new method with aggregates based on other approaches. Aggregates of owners' reported values come from the Survey of Consumer Finances (SCF) and the ACS. We also compare our estimates with the *Financial Accounts*, which is put together using information from surveys, house-price indexes, new residential construction, and residential improvements. Most of the dynamics of the *Financial Accounts* measure after 2005 come from changes in the house-price index. Thus, these two points of comparison (owner-reports and the *Financial Accounts*) encompass different methods that are commonly used to measure housing wealth, helping to illustrate the key measurement issues.

While the three measures are largely in line with one another from 2001 to 2006, they diverge notably from 2006 to 2016, a time period that included an enormous housing bust and a gradual recovery. In particular, the *Financial Accounts* measure shows a very pronounced downturn and recovery over this period, while owner-reports show a significantly milder cycle. The AVM-based measure lies in between the two. We also find that the timing of the cycle differs among the measures. Whereas the AVM and *Financial Accounts* measures date the peak of the cycle in 2006, the owner-reported measure does not peak until 2008.

The comparison to the ACS highlights some of the pitfalls of using survey data to measure housing wealth. In particular, it suggests that survey respondents were either unaware of the market fluctuations in real time, or they believed that their home values were different than those in the surrounding market. To the extent that they did acknowledge changes in the market, it appears that they were late to do so.

The comparison to the *Financial Accounts* measure illustrates the implications of extrapolating aggregate housing wealth using a price index. The much more pronounced swings in wealth in the *Financial Accounts* likely result from applying the house-price dynamics of transacting homes to all homes. As pointed out by Case, Pollakowski, and Wachter (1997), Gatzlaff and Haurin (1997), and Glennon, Kiefer and Mayock (2016), the sample of homes that are sold in a given time period may not accurately represent the universe of all homes, the vast majority of which will not have been sold in that period. For example, the types of homes that tend to sell during periods of market stress could be subject to different supply and demand conditions than other homes, leading to differential price trajectories for the two groups.

Our findings have direct implications for studies that use housing wealth estimates from the *Financial Accounts*, house-price indexes, or self-reported values from surveys. First, our finding that the recent housing cycle may have been more muted than suggested by repeat-sales house price indexes could affect the results of studies that use the *Financial Accounts* (or a repeat-sales index itself) to measure changes in household net worth.<sup>7</sup> At the same time, our finding that the recent downturn-and-recovery cycle in housing was *more* pronounced—and happened somewhat earlier—relative to self-reports from surveys could affect a variety of findings in recent studies that use self-reported values to examine fluctuations in house values or housing wealth over the Great Recession.<sup>8</sup>

---

<sup>7</sup> Recent studies of this type include Carroll, Otsuka and Slacalek (2010), which looks at the response of consumption to wealth changes; Glover, Heathcote, Krueger, and Rios-Rull (2016) who study the impact of the Great Recession on the intergenerational reallocation of household wealth; or Saez and Zucman (2016) who combine aggregate wealth data from the National Accounts with income tax records to impute a distribution of U.S. household wealth and to study its evolution over time.

<sup>8</sup> See, for example, Ferreira, Gyourko and Tracy (2010); Hur (2016); Peterman and Sommer (2016); or Bricker et al. (2011).

The rest of the paper is organized as follows. Section 2 provides a summary of the difficulties in measuring housing wealth. Section 3 describes the basics of Zillow’s AVM methodology. Section 4 describes how we use Zillow’s AVM to produce a nationally representative measure of aggregate housing wealth. Section 5 describes our estimates of housing wealth for United States and how these measures compare to the *Financial Accounts* and aggregated survey data. Section 6 concludes with a discussion of the implications of our paper for other studies and future uses of large-scale AVMs in economic research.

## **II. Measuring Housing Wealth**

As with any other economic statistic, the best measure of aggregate housing wealth would be unbiased, precise, and available at a high-frequency and with a short reporting lag. Ideally, we would observe the current market value of every home at all times, and simply add them up to measure the aggregate. In reality, current market values exist only for homes that have sold recently, which make up a minority of the stock of homes. This raises two problems for measuring aggregate housing wealth. First, one needs a good way to estimate market values for non-transacting homes. Second, one needs to account for new construction, renovations, tear-downs, conversions, and other issues reflecting changes to the aggregate housing stock that may not be reflected in recent sales.

In practice, measurements of housing wealth have typically derived from one or both of two sources: surveys of home owners who are asked to estimate the value of their homes, and repeat-sales price indexes derived from recent market transactions. These approaches offer complementary advantages and disadvantages. Surveys (in principle) represent the entire stock of housing and take into account the details of the home and local market conditions. Moreover, these data allows one to identify own-use properties—the object of this study. However, as

discussed below, self-reported values may be biased and inaccurate if the survey respondents are biased or if they lack information relevant for valuation. Price indexes, on the other hand, are based on actual market transactions, but do not necessarily represent non-transacting homes well, particularly during boom times and periods of market stress. Price indexes also may not reflect changes in the characteristics of the housing stock.

A long literature explores these issues in detail. Regarding the ability of house price indexes to estimate aggregate home values, several papers have shown that the samples of housing units used to form repeat-sales indexes are not representative of a broader set of homes. Case, Pollakowski, and Wachter (1997) find that homes that trade more frequently have systematically higher price appreciation. Gatzlaff and Haurin (1997) find that the selection bias associated with repeat-sales is highly correlated with economic conditions. Doerner and Leventis (2015) find that the rise in distressed sales during the post-2008 housing bust created a downward bias in FHFA price indexes. Korteweg and Sorenson (2016) develop a selection-corrected pricing model that accounts for the endogeneity of transactions, and find that this correction significantly increases the share of homes estimated to be worth less than their outstanding mortgages. Malone and Redfearn (2017) calculate repeat-sales price indexes for various sub-markets and housing unit types in Los Angeles and Sweden and show that in both samples, house price appreciation varies substantially by property type and location.

Surveys do not have the same problems with representativeness as repeat-sales indexes because they can be designed to be nationally representative. However, basing an estimate of aggregate wealth on owners' valuations as reported in survey requires assuming that owners can accurately determine the value of their property. Assessing the validity of this assumption is complicated by the fact that one needs to determine a "correct" value to compare with. Ideally,



an arms-length market transaction price would be the best choice. Goodman and Ittner (1992) compare survey respondents' house valuations to subsequent sales prices (over the next two years) using data from the American Housing Survey (AHS), and find that, on average, owners over-estimate the value of their homes by about 8%.<sup>9</sup> We are not aware of any other studies that compare owner reports with market transaction prices.

Other studies compare owner reports to a value that is determined by price indexes. Because of the biases in repeat-sales indexes discussed above, these analyses should be interpreted cautiously. Even so, they tend to find material differences between owner reports and other value estimates. For example, Chan, Dastrup and Ellen (2016) compare owner reports in the AHS and Health and Retirement Study (HRS) to market value estimates derived from the prior sales price inflated by the change in the Zillow House Value Index in the ZIP code.<sup>10</sup> Over their entire time period (1997 to 2011 in the AHS, 1998 to 2010 in the HRS), the average difference between the owner report and estimated market value is 3 percent (AHS) and 4 percent (HRS). However, this difference varies substantially in the cross section and over time, with owner-reported values not falling nearly as much as market values during the housing crisis. Henriques (2013) compares changes in aggregate reported house value from the SCF to changes in price indexes, and finds that the survey reports rose much more rapidly than price indexes during the pre-2008 boom, and fell slightly less afterwards. Davis and Quintin (2017) compare average owner reports by metropolitan area to metro-level repeat-sales indexes. In contrast to

---

<sup>9</sup> To adjust for the time lag between when the survey was conducted and when the home was sold, the authors inflate the owner valuations using a metropolitan area house price index. This adjustment would not be accurate if the true value of the home did not appreciate in line with the metropolitan area average.

<sup>10</sup> Unlike the Zillow AVM that we use in this paper, the Zillow House Value Index (ZHVI) keeps the number of properties and property characteristics fixed; it is meant to reflect pure price changes over time. Because the ZHVI is based on value estimates for the entire housing stock, it should not suffer from the same selection bias as repeat-sales price indexes.

Henriques (2013), they find that self-assessed house prices rose *less* quickly than price indexes during the boom, and fell less quickly during the bust, which they posit is consistent with a model of optimal filtering of information on house prices. Using Dutch data, van der Crujssen et al (2014) find that 75% of survey respondents overvalue their homes relative to regional trends, with a median overvaluation of 13 percent.<sup>11</sup>

In part due to the complementary advantages and disadvantages of survey data and price indexes, some measures of aggregate housing wealth combine information from both sources. An example is the measure constructed for the *Financial Accounts of the United States*, which combines data from the biennial American Housing Survey (AHS), the CoreLogic house-price index, and a measure of net residential investment from the Bureau of Economic Analysis (BEA). From 1999 to 2005, the aggregate level of owner-occupied housing wealth reported in the *Financial Accounts* was benchmarked to the AHS using a weighted sum of the values reported by survey respondents.<sup>12</sup> Between AHS benchmarks, and exclusively after 2005, the aggregate level was extrapolated by growing the previous level estimate by the CoreLogic price index and adding the BEA's estimate of net investment in residential structures. The *Financial Accounts* discontinued benchmarking to the AHS after 2005 because the housing wealth estimates from the subsequent AHS surveys diverged dramatically with the extrapolation based on the CoreLogic house-price index and net investment. In particular, the AHS results initially suggested that respondents were continuing to report significant gains in the values of their homes, even after the price-index measures began showing a steep downturn. Because it did not

---

<sup>11</sup> Other papers finding overvaluation among survey respondents include Ihlanfeldt and Martinez-Vazquez (1986), who report an average overvaluation of 16%, Kiel and Zabel (1999), who find an average overvaluation of 5%, and Benítez-Silva et al (2016), who report an average overvaluation of 8%.

<sup>12</sup> To account for the overvaluation of survey respondents discussed above, the benchmarks derived from the AHS are adjusted down by 5½ percent.

seem credible that house values were continuing to rise during a time of falling home sales and surging mortgage delinquency and foreclosure, the benchmarking to the AHS was discontinued.

### **III. Automated Valuation Models**

While issues inherent to the house price index and owner-report methodologies have been recognized by researchers for years, no good alternatives existed for constructing estimates of aggregate housing wealth. Recently, however, AVMs have emerged as a promising contender. Although financial institutions have used versions of AVMs for decades to value mortgage portfolios, the models and data have only recently reached the point where they can provide a viable method for measuring aggregate housing wealth. In particular, private firms such as Zillow have recently created sophisticated models based on machine learning and assembled very large property-level datasets to drive AVMs that can impute values of individual housing units for large swaths of residential real estate in the U.S. This combination of extensive, detailed data and machine learning offers the potential for more accurate and more representative estimates of housing values than those based on surveys or price indexes.<sup>13</sup>

Zillow's AVM attempts to assign values to all single-family homes as well as co-op and condominium apartments.<sup>14</sup> While the details of Zillow's AVM are proprietary, Zillow has made public a general description of their methods. Additionally, Zillow has provided us with a confidential set of summary statistics that allow us to construct measures of aggregate housing wealth.<sup>15</sup> As described in the next section, in order to do so, we need to devise methods to address some of the ways the AVM data are, in their raw form, not ideal for our purposes.

---

<sup>13</sup> Zillow released the first home value index built from AVMs in 2005.

<sup>14</sup> Zillow does not attempt to value apartments in rental buildings because such buildings are bought and sold as a single property. Therefore, an AVM approach based on transaction prices for individual homes would not be a valid.

<sup>15</sup> See <https://www.zillow.com/research/zestimate-forecast-methodology/>.

At its most basic level, Zillow's AVM is an algorithm that converts data on housing characteristics and transaction prices into property-level value estimates for most single-family residential properties in the U.S. Although the rest of this section will make clear that this analogy is imperfect, as a rough first pass one can think of the AVM as a hedonic regression relating house values to a rich set of property and neighborhood characteristics. The relationships between these characteristics and house values are estimated from observed sales prices. The estimated model can then be applied to properties that do not have a recent sales price to produce value estimates for the full stock of homes for which sufficient data on characteristics are available or can be imputed.

Zillow's AVM uses a wide variety of data sources. Deeds records and property tax records are the backbone of their data, as these records are nearly universal and, when combined, typically include property characteristics and transaction details such as sales prices, dates, etc. However, deeds and property tax records are not perfect. For example, "non-disclosure" states do not require that sales prices be disclosed in deeds records and property tax records do not capture all of the myriad property characteristics that affect a home's value.<sup>16</sup> To add additional information on property characteristics, Zillow supplements the deeds records with data from Multiple Listing Service (MLS) registries, mortgage servicers, and other sources. For example, Zillow's data includes information about water views, local school quality, and other local amenities that would be very difficult to assemble through other means. In addition, the Zillow

---

<sup>16</sup> The non-disclosure states are Alaska, Idaho, Kansas, Louisiana, Mississippi, Montana, New Mexico, North Dakota, Texas, Utah, and Wyoming. In addition, some counties in Missouri do not require that sales prices be disclosed in deeds records. Even in non-disclosure states, mortgage loan amounts are often disclosed at recorders' offices. See <http://www.zillowgroup.com/news/chronicles-of-data-collection-ii-non-disclosure-states/>.

website allows homeowners to update or correct the characteristics of their property that might be missing or inaccurate in Zillow's database.

One reason the Zillow AVM is not literally a hedonic regression is that Zillow's AVM is not actually a single model, but a very large number of distinct models that work together to produce value estimates. Each model uses different data in different ways; the resulting value estimates are then combined to form the AVM's single best guess of the property's value. Different models may have different rules for selecting the homes used as a reference for a given property. For example, some models might require complete deeds records, while other models may be able to produce a (noisier) estimate without deeds data at all. The AVM might use the more spartan models only in the case that complete deeds records are not available. However, the AVM will not assign a value to a property if it determines that the available data are insufficient or if the AVM ascribes too much uncertainty about the estimated value.<sup>17</sup> The properties excluded for this reason tend to be in less populated areas where transactions are sparse and data quality is poor.

Zillow updates and re-estimates their models daily to onboard new data as it becomes available. The high frequency is useful in validating the AVM because as house sales occur, they can be compared against very recent AVM estimates. Zillow's daily model run allows them to continually assess their model errors for bias and update their model calibration.

#### **IV. Using Zillow's AVM to Measure Aggregate Household Housing Wealth**

---

<sup>17</sup> Indeed, the reticence of Zillow's AVM to assign estimates for such properties lowers their data coverage and partially motivates our use of ACS property counts to "gross up" to a national aggregate. From our point of view, any unbiased estimate of property's value is worth including in the calculation of aggregate wealth. However, Zillow's objective is to report individual property estimates as accurately as possible, which motivates their decision to exclude highly uncertain estimates.

As stated above, the objective of this paper is to measure the aggregate value of household housing intended for own-use. This approach is consistent with concepts from the *Financial Accounts*, and is also comparable to aggregates constructed from most survey data.

A key issue in using Zillow’s AVM universe as a basis for estimating aggregate U.S. housing wealth is Zillow’s coverage of own-use housing units. Specifically, not all residential housing units are valued (due to accuracy or data coverage limitations described in the previous section). All else equal, excluding own-use units with missing AVM values from our measure would result in an underestimate of aggregate housing wealth intended for own-use. Conversely, Zillow’s universe also includes some rental properties, which we do not want to include in a measure of aggregate own-use housing wealth. These properties cannot be straightforwardly identified in the Zillow data and—if included—would inflate our estimate of aggregate household housing wealth intended for own-use.

Table 1: Property Counts in 2015  
(millions of units)

	ACS		Zillow
	Total	Own-use	Total
Single-family	90.6	70.6	79.4
Multifamily	35.6	5.4	9.8
Total	126.2	75.9	89.3

Table 1 illustrates how the Zillow data compares with the universe of own-use properties in 2015. According to the (nationally representative) ACS, there were about 90 million single-family homes in the US in 2015, about 70 million of which were for own-use and 20 million of

which were for rental use.<sup>18</sup> Zillow's AVM is able to provide value estimates for about 80 million single-family homes in 2015. Thus, Zillow's overall coverage of the single-family market, at about 90 percent, is quite high. However, because owner-occupied homes are not identified as such in the Zillow data, we cannot know exactly how many single-family homes meant for own use are missed by Zillow or how many single-family rental homes are included.

These coverage issues are more severe for multifamily housing units. Based on the ACS, there were about 36 million multifamily homes in 2015, of which about 5 million were for own use. Zillow's sample for 2015 includes about 10 million multifamily housing units, only a small fraction of the total number of multifamily units in the ACS. As mentioned above, Zillow only attempts to value apartments that are in co-op or condo buildings, and excludes apartments in rental buildings (i.e., buildings in which a single property-tax parcel contains multiple units). This exclusion is a helpful feature for our purposes, as the majority of "missing" units in Zillow's multifamily sample are ones we do not want to value anyway. However, the Zillow multifamily sample, at about 10 million units, is still about 5 million units larger than the ACS own-use count for multifamily units, which suggests that the Zillow multifamily sample includes many rental units that we would like to exclude. In addition, it likely also misses some own-use properties that we would like to include.

Turning to geographical coverage, Figure 1 shows that, at the state level, there is a strong positive correlation between the number of properties covered by Zillow in 2015 and the number of own-use properties in the ACS. Zillow's coverage of single-family units is a little lower for

---

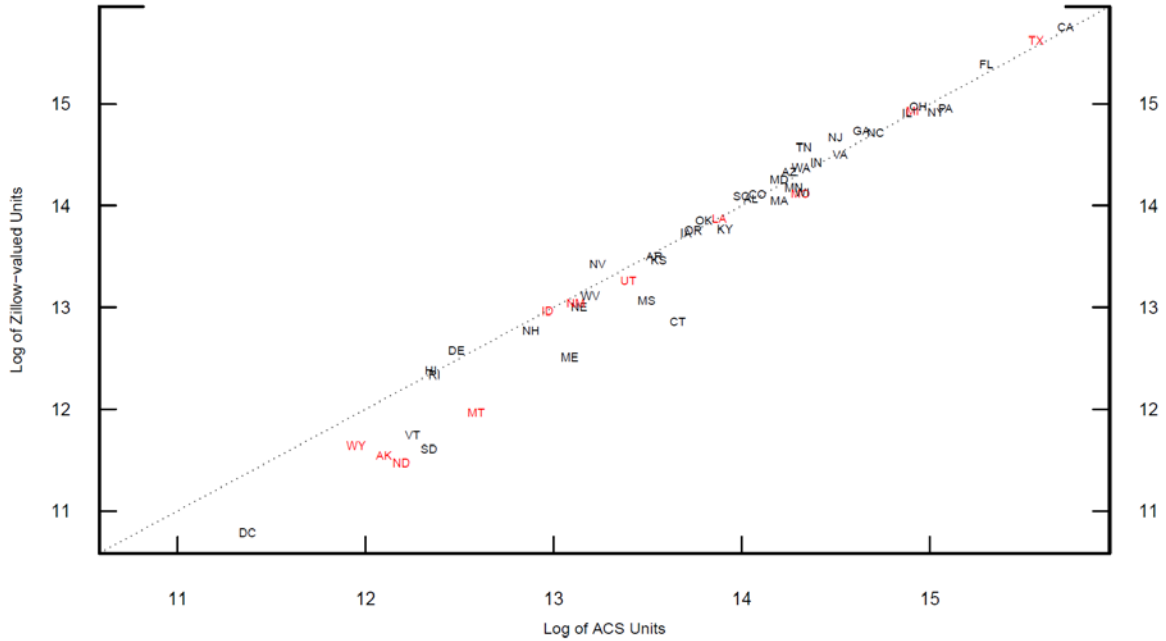
<sup>18</sup> We define properties in the ACS as "own-use" if they are owner-occupied or vacant and likely intended for own-use. The latter category includes units that are for sale and a fraction of all other vacant units that are not for rent. This fraction is determined by the ratio of owner-occupied to renter-occupied units by state and property type. See below for details.

units in smaller and/or less densely populated states, which is not surprising because housing markets are likely less thick in these areas and Zillow's coverage is partially determined by the availability of comparable property sales. For multifamily properties (lower panel) there is more variation in coverage across states, but Zillow tends to have more properties than the estimated number of own-use properties in the ACS. Somewhat surprisingly, Zillow is able to estimate values for a large fraction of own-use properties in non-disclosure states (depicted in red). Zillow uses other data sources, such as mortgage records and sales prices from MLS registries, to estimate property value in those states.



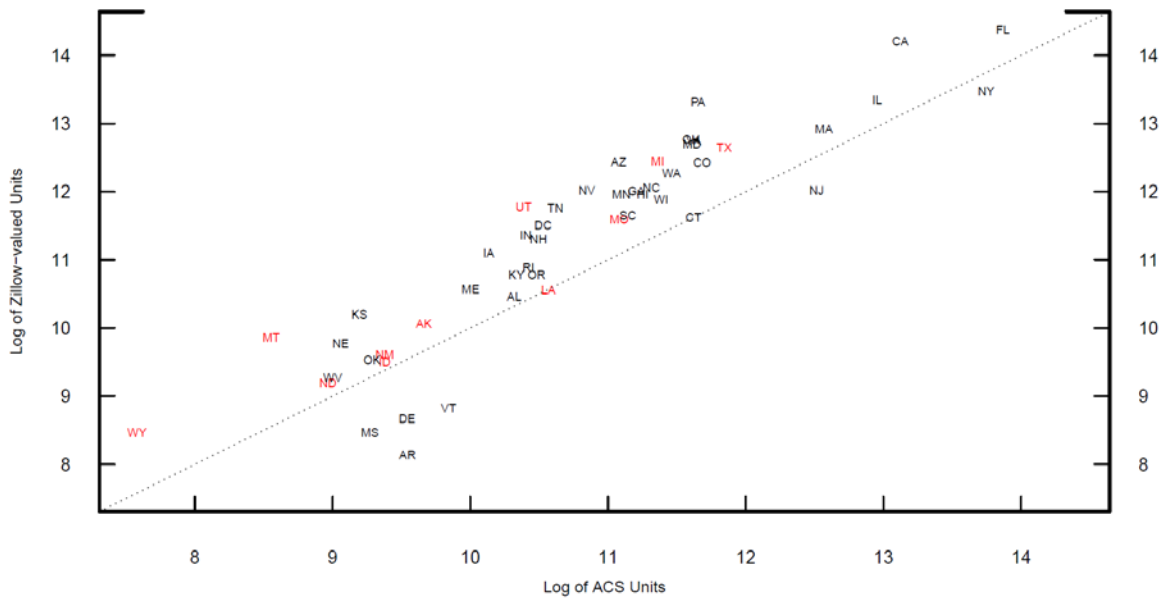
**Figure 1: Ratio of Zillow to ACS Property Counts in 2015 by State**

**Single Family, ACS Own-use v. Zillow**



Red denotes non-disclosure states.

**Multifamily, ACS Own-Use v. Zillow**



Red denotes non-disclosure states.

The nature of Zillow’s coverage of properties implies that constructing a measure of aggregate housing wealth intended for own-use is not as simple as adding up the AVM values for all properties in the Zillow universe. A simple sum of Zillow’s AVM values would miss some own-use homes. At the same time, some rental units—which our definition of housing wealth excludes—would be included, artificially inflating our measure. Moreover, the number of properties valued by Zillow changes over time due to the availability of the data needed to estimate the AVM. We would not want an estimate of aggregate housing wealth to fluctuate over time because housing units move in and out of the sample.

To address these issues, we propose a method that combines Zillow AVM average values by property type and location and with ACS property counts into a national total. Specifically, we define market segments by county  $c$  and structure type  $s$  (i.e., single-family or multi-family), and estimate the value of own-use housing at time  $t$  as

$$\hat{V}(t) = \sum_{s,c} N^A(s, c, t) \bar{V}^Z(s, c, t) F(s, c, t),$$

where  $\bar{V}^Z(s, c, t)$  is the average AVM value for residential properties in county  $c$  and of structure type  $s$  at time  $t$ ,  $N^A(s, c, t)$  is an estimate of the number of properties intended for own use from the ACS, and  $F(s, c, t)$  is a factor used to adjust the AVM value estimate for any known and quantifiable bias.

To construct the counts of property intended for own use from the ACS, we split all housing units reported in the survey into three mutually exclusive and exhaustive categories: units that are unambiguously for own use (owner-occupied plus vacant-for-sale), units that are unambiguously for rental use (renter-occupied plus vacant-for-rent), and units that are vacant but are not for sale or for rent. The total number of properties intended for own use is then the sum

of units intended for own use plus a share of vacant properties, where that share is estimated from the fraction of unambiguously own-use properties.<sup>19</sup> That is:

$$\begin{aligned}
 N^A(s, c, t) &= N^A(s, c, t; \text{own use}) + N^A(s, c, t; \text{vacant}) * \phi(s, c, t) \\
 \phi(s, c, t) &= \frac{N^A(s, c, t; \text{own use})}{N^A(s, c, t; \text{own use}) + N^A(s, c, t; \text{rental use})}
 \end{aligned}$$

Although Zillow’s sample is quite comprehensive, they are not able to estimate values for properties in every county in the US. We do not want to assume that the aggregate value of property is zero in counties for which we do not have an average Zillow AVM. Instead, for counties that do not have an average AVM estimate, we assume that the average value is equal to the average value in the state.

Because our approach uses Zillow data to calculate average valuations for a given property type in a given county, it will yield an unbiased measure of housing wealth under four assumptions. First, we assume that Zillow’s AVM methodology is not inherently biased at this level of analysis—i.e., that the Zillow value-weighted average estimate for a given county and property type is an unbiased estimate of the true value-weighted average (or that we can construct a factor to adjust for any bias, such as  $F(s, c, t)$  in our notation above). Second, we assume that the lack of complete coverage in the Zillow data does not bias the valuations—i.e., that the average value of properties with and without an AVM are the same, conditional on county and property type. Third, we assume that the inclusion of rental properties in Zillow’s sample does not bias our measure away from our intended measure of own-use properties—i.e., that the average value of rental and own-use properties is the same, conditional on county and

---

<sup>19</sup> We assume that the share of vacant properties that are intended for personal use is the same as the share of occupied properties for personal use. Otherwise, our discussion here presumes that the ACS property counts are unbiased.

property type. Fourth, we assume that imputing valuations for counties with no Zillow observations does not introduce a bias—i.e., that the average value of homes in counties where no property has an AVM estimate is the same as the average value of all properties in the state. In what follows, we will investigate the validity of each of these assumptions in turn.

One final issue that bears mentioning is that we can only use the methodology described above for the 2007-2016 time period. The version of the Zillow AVM that we are currently using is not available for earlier years. However, a different version of the Zillow AVM is available back to 2000. Although this version is based on a somewhat different set of property and neighborhood characteristics and has a somewhat different geographic footprint, the estimate of aggregate house value that we obtain using this AVM is very close to the estimate that we obtain using our preferred AVM in the years for which the two overlap (2007 to 2015). Consequently, in the analysis below we combine the results derived from both AVMs in order to derive the longest possible time series. In the future Zillow plans to expand the sample period of our preferred AVM back in time, in which case we will no longer need to rely on the older version of the AVM.

#### **IV.A. Is the value-weighted AVM an unbiased estimate of true average value?**

Our method assumes that Zillow's AVM provides an unbiased estimate of average value by county and property type. We cannot test this assumption directly, both because we lack the AVM micro data and because model errors are only available for properties that transact. Nonetheless, Zillow has provided us with data on county-level error distributions that allow us to estimate the average value-weighted AVM error for transacting properties over time. We can therefore test explicitly whether this assumption holds approximately true.

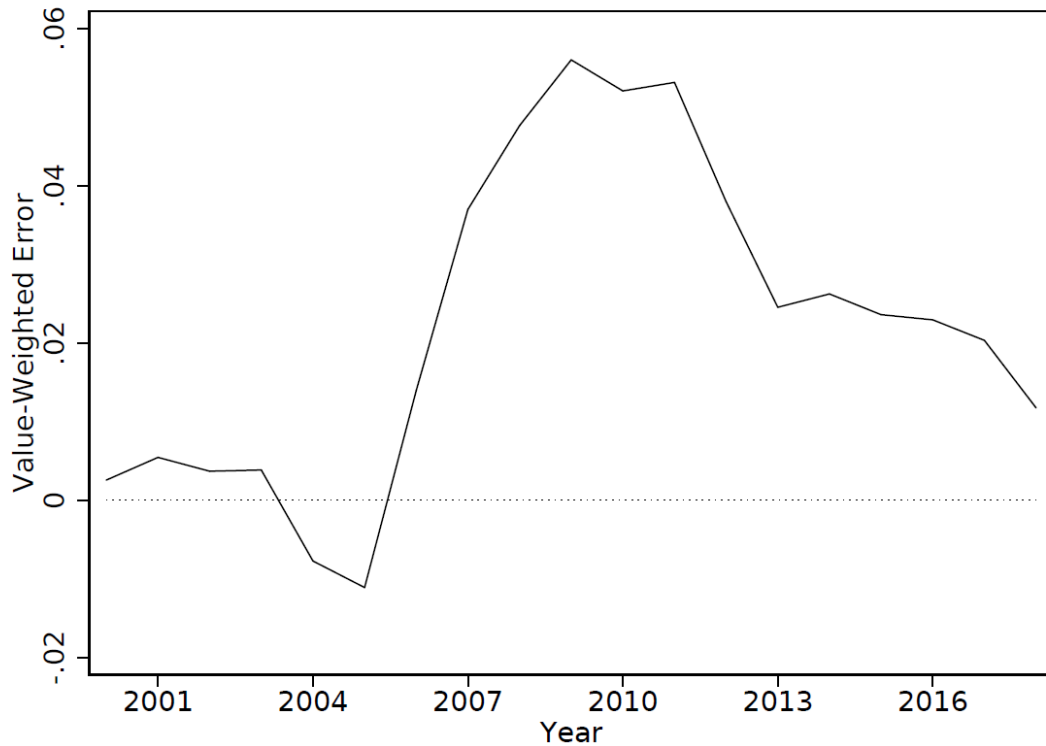
The detailed description of our approach to calculating the average weighted AVM error is described in the Appendix. In brief, for each recently transacted property, the error is computed as the difference between the AVM estimate at the end of the month prior to a transaction and the transaction price, expressed as a share of the transaction price. A positive error corresponds to Zillow’s AVM overvaluing a property relative to the transaction price. Because our goal is to estimate aggregate housing wealth and higher-value properties have a larger influence on aggregate wealth, we calculate the value-weighted average error across properties.

As Figure 2 illustrates, the average value-weighted error was near zero in the early 2000s. The average dipped below zero in 2004 and 2005 as market prices were rising so briskly that the AVM did not fully keep up. During the housing crisis, the error turned positive and then widened considerably because market prices fell much more quickly than estimated by the AVM. After peaking at approximately 5 percent during 2009 to 2011, the average error has declined to about 2 percent in recent years. The size of this average error is large enough that we think it is important to account for in our estimates of aggregate house value. Therefore, we adjust for this bias at the county level by multiplying our “raw” measure of aggregate housing wealth by the inverse of one plus the estimated value-weighted average error in that county and year.<sup>20</sup>

---

<sup>20</sup> See the Appendix for details. We use the combined average error rates for single-family and multifamily units because the magnitudes of the errors in these two sectors are fairly similar (see Figure 8 in the Appendix). Additionally, because the errors for the multifamily sector are frequently based on a small number of transactions, we think the average across all property types provides a more reliable estimate of the true bias.

**Figure 2: Average Value-Weighted Errors**



Given the evidence of a modest bias in the AVM, a natural question that arises is whether using the average value from an AVM is any more accurate than using the average value from another method, such as owner reports or extrapolating using a house price index. A detailed comparison of value estimates based on these different methods is beyond the scope of this paper. Nevertheless, results reported in other studies suggest that an AVM is probably at least as good at estimating “true” market value as other approaches. For example, Glennon, Kiefer and Mayock (2016) extrapolate prior sales prices with a repeat-sales price index, and find that the (unweighted) average difference between this extrapolated value and the subsequent transaction price in 2010 was 26 percent in Los Angeles county, 41 percent Maricopa county, 81 percent in Miami-Dade county, and 113 percent in Cook county (their analysis only covers these four counties). By contrast, the average error of the Zillow AVM in those four counties ranged

between 6 and 17 percent in 2010.<sup>21</sup> Thus, the AVM appears to be more accurate than an approach to valuation based on a house price index, at least during the housing downturn.<sup>22</sup>

Goodman and Ittner (1992) compare owner reported values in 1985 to the sales prices of homes that sold within the subsequent two years.<sup>23</sup> We cannot make a direct comparison since Zillow does not have AVM estimates or subsequent transaction prices for this time period. Nevertheless, we compare these estimates with estimates based on a Zillow AVM from 2000 to 2003, as the rate of increase of aggregate house prices was roughly similar in these two periods. The average (unweighted) error of the Zillow AVM during this period was 7 percent, fairly similar to the average error of 8 percent reported in Goodman and Ittner (1992). The unweighted Zillow average error is noticeably larger than the value-weighted average shown in Figure 2 because the errors tend to be larger for lower-value properties.

#### **IV. B. Is the average value of homes that do not have an AVM equal to the average value of homes with an AVM?**

The second assumption—that the types of homes that Zillow values in a given county are not materially different from the types of homes it cannot value—is not directly testable in our data. Homes might not have an AVM estimate for a variety of reasons, including a lack of information about property characteristics or a lack of sales of comparable homes. We suspect that this

---

<sup>21</sup> To be comparable to the Glennon, Kiefer and Mayock paper, the Zillow average errors reported here are not value-weighted.

<sup>22</sup> Glennon, Kiefer and Mayock (2016) report smaller average errors for 2005, ranging from 3 to 7 percent. At this time we do not have Zillow AVM error estimates for these four counties in 2005. However, consistent with Glennon, Kiefer, and Mayock's finding, the aggregate results in Figure 2 suggest that Zillow's average weighted error in these counties was also likely lower in 2005 than in 2010.

<sup>23</sup> They adjust the owner reports by metropolitan-area house price index to adjust for the time lag between when the owner was asked to estimate value and when the home sold.

assumption is relatively benign for single-family properties, where the excellent coverage of Zillow's AVM means that only a small fraction of properties do not have an AVM estimate.<sup>24</sup>

The representativeness assumption is harder to justify in the case of multifamily properties where, as shown in Table 1 and Figure 1, the Zillow coverage differs materially from the set of housing units that we would like to value. That said, the multifamily properties that Zillow does value are much more likely to be representative of own-use housing units than the unvalued properties because they are based on market transactions of condos and co-ops and Zillow only attempts to estimate values for tax parcels with single units. Therefore, we believe it to be reasonable to extrapolate the average values of these multifamily units to the values of all condos and co-ops, whereas it would be implausible to assume that these average values apply to the typical rental apartment.

To investigate this second identifying assumption further, we examined property-level data from the 2014 ACS that were merged by address with a property-specific AVM estimate from another data provider<sup>25</sup>. While not directly testing the Zillow AVM we use in this paper, this sample allows us to compare the owner-reported values of homes that do have an AVM estimate with owner-reported values of homes that do not. Specifically, for each property type (single- or

---

<sup>24</sup> A simple example makes this point clearly. Suppose that  $p \in (0,1)$  represents the share of single-family homes valued by Zillow. On the extreme assumption that the unvalued homes (whose share is  $(1 - p)$ ) have an average value of 0, our method would lead to an overstatement of the true average value by a factor of  $p^{-1}$ , which declines to 0 quickly as  $p$  approaches 1. For the coverage rates (values of  $p$ ) we see in the data, this extreme selection would result in an overestimate on the order of 15-20%. If the unvalued homes are even three-fourths as valuable as the valued homes, then the overestimate falls to around 4% when the model coverage is 85%.

<sup>25</sup> The Census Bureau purchased AVM estimates from another data provider, so we were able to conduct this analysis using the confidential microdata available at the Census Bureau.



multifamily) we regress the natural logarithm of the owner-reported value on an indicator for whether the AVM estimate is missing, controlling for county fixed effects.<sup>26</sup>

Reassuringly, as Table 2 shows, the coefficient on the missing AVM indicator is near zero for single-family homes; i.e., owner-reported values of single-family homes with a missing AVM are only about 1% different, on average, than homes within the same county that do have an AVM estimate. However, when the regression is done without fixed effects, we find that single-family homes with a missing AVM have a 19 percent lower average value than homes that have an AVM estimate, and in a regression with state-level (rather than county-level) fixed effects the coefficient is 10 percent. These results illustrate the non-random geographical distribution of the missing AVM values and thus the importance of aggregating property-level AVM estimates at granular level. For multifamily housing units, the value difference between those with an AVM and those without an AVM is a little larger—units without an AVM are about 7 percent lower in value than other units in the same county that do have an AVM estimate.

Table 2: Differences in Owner-Reported Values by AVM Missing Status

	(1)	(2)	(3)
AVM Missing Indicator	County FE	State FE	No Controls
Single-Family (N=1,416,264)	0.012*** (0.0014)	-0.10*** (0.0014)	-0.19*** (0.0014)
Multi-Family (N=63,838)	-0.068*** (0.0063)	-0.055*** (0.0068)	0.029*** (0.0072)

Standard errors in parentheses. Data are trimmed by excluding values less than \$10,000 or more than \$4 million.  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

<sup>26</sup> Since AVMs are more likely to be missing in sparsely populated areas with fewer transactions, failure to control for geography would conflate geographical heterogeneity with the effect of having an AVM valuation within the same market segment.

Overall, the results in Table 2 suggest that properties without an AVM estimate have a similar value as other properties within the same county, implying that our second identifying assumption roughly holds provided that we aggregate from the county level up. Of course, this conclusion supposes that the types of housing units for which Zillow cannot estimate an AVM are similar to the types of housing units that did not have an AVM in the internal Census data.

#### **IV. C. Is the average value of rental units the same as the average value of own-use units?**

Our third identifying assumption is that the rental properties included in the Zillow data have the same average values as own-use units within the same county and property type. The public deeds records do not allow one to easily identify which homes are held for own use and which homes are intended as rental units.<sup>27</sup> Consequently, the Zillow AVM averages will include the values of some rental units, and will not represent own-use properties if rental units have systematically lower or higher values.

In principle, this rental bias could go either way. On the one hand, rental units are likely to be smaller and of lower quality than owner-occupied units, dragging the average AVM estimate down. On the other hand, rental units may be in more desirable locations, and hence be located on more valuable land.

We evaluate this assumption using the same merged ACS/AVM property-level data from 2014 that was used in the previous section. In this case, we regress the natural logarithm of the AVM on an indicator for whether the ACS identifies the property as a rental unit and fixed effects for various geographies. Table 3 shows that on average, single-family rental units are 34

---

<sup>27</sup> In future work, we might be able to infer ownership of the properties by matching the address of the property to the address of the owner in the tax assessors' records.

percent lower value than owner-occupied units within the same county (column 1). For multifamily units, this differential is a little smaller, but multifamily rental units are still 17 percent lower in value than owner-occupied units within the same county.<sup>28</sup>

Table 3: Differences in AVM Value by Rental Status

Rental Indicator	(1) County FE	(2) State FE	(3) No Controls
Single-Family (N=1,147,726)	-0.34*** (0.0013)	-0.39*** (0.0016)	-0.37*** (0.0018)
Multifamily (N=65,998)	-0.17*** (0.0046)	-0.21*** (0.0052)	-0.25*** (0.0064)

Standard errors in parentheses. Data are trimmed by excluding values below \$10,000 and above \$4 million.  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Because of the large value differential between rental units and owner-occupied units, our third assumption appears not to hold. We estimate the effect that this could have on aggregate housing wealth value based on the average value differential between owner-occupied and rental units in the merged ACS/AVM sample and the share of rental units that we estimate to be included in Zillow’s AVM data; the Appendix describes our approach in detail. This analysis suggests that adjusting for the unintended inclusion of rental units would increase our estimate of aggregate housing wealth in 2014 by about 6 percent. Since we cannot tell how this effect changes over time (because we do not have the merged ACS/AVM data for years prior to 2014), we do not adjust our time series of aggregate wealth to account for the bias imparted by the inclusion of rental units.

<sup>28</sup> It is worth keeping in mind that most of the multifamily rental units with an AVM estimate are likely condominium units that are rented out. These results would likely be quite different if the AVM were available for rental units in buildings where the entire building has a single owner.

#### **IV. D. Is the average value of homes in counties without any AVM estimates the same as the average value in the state?**

Zillow currently does not provide average values for properties in every county. For counties that do not have an average AVM estimate, we impute the missing value with the average value in the state. At first blush, this approach might appear problematic because these counties are more likely to have thin housing markets and be in rural areas, suggesting they should have a lower average value than other counties in the same state. Indeed, Table 2 shows that properties without an AVM tend to have lower value than properties with an AVM in the same state. However, the bias imparted by this assumption on the measure of aggregate housing wealth is almost surely quite small, as the counties with a Zillow AVM average cover the vast majority of housing units. Specifically, for single-family units in 2015, Zillow provides us with average values for 445 counties, which represent 96 percent of own-use housing units. For multifamily units in 2015, we have average Zillow values for 429 counties, accounting for 97 percent of own-use units. Statistics are similar for earlier years. We have explored the use of alternate assumptions that base the average value for “missing” counties on characteristics of the county. But the effects on aggregate value are trivial, so for simplicity we maintain the assumption that counties with a missing AVM have the same average as the state-wide average.

The results of our analysis of our four maintained assumptions can be summarized as follows. First, the Zillow AVM does appear to reflect market transaction prices with a modest degree of error, and this error has fluctuated over time—the value-weighted error averaged about 2 percent in recent years, but was close to 6 percent during the collapse of the housing market. We attempt to correct for this known bias by calculating county-level adjustment factors for each year of the sample. Second, in a separate sample that matches owner reports to AVM valuations

in 2014, we find little difference between owner-reported house values for homes with and without an AVM valuation. Third, in the same sample, we find evidence of a significant difference in average values between rental homes and homes for own use. We do not have sufficient data to calculate adjustment factors for this difference over our time series, but the results indicate that Zillow's inclusion of rental properties lowered the calculated aggregate value of housing wealth by about 6 percent in 2014. Finally, we find little effect on our calculations from missing counties in the Zillow data.

## V. Results

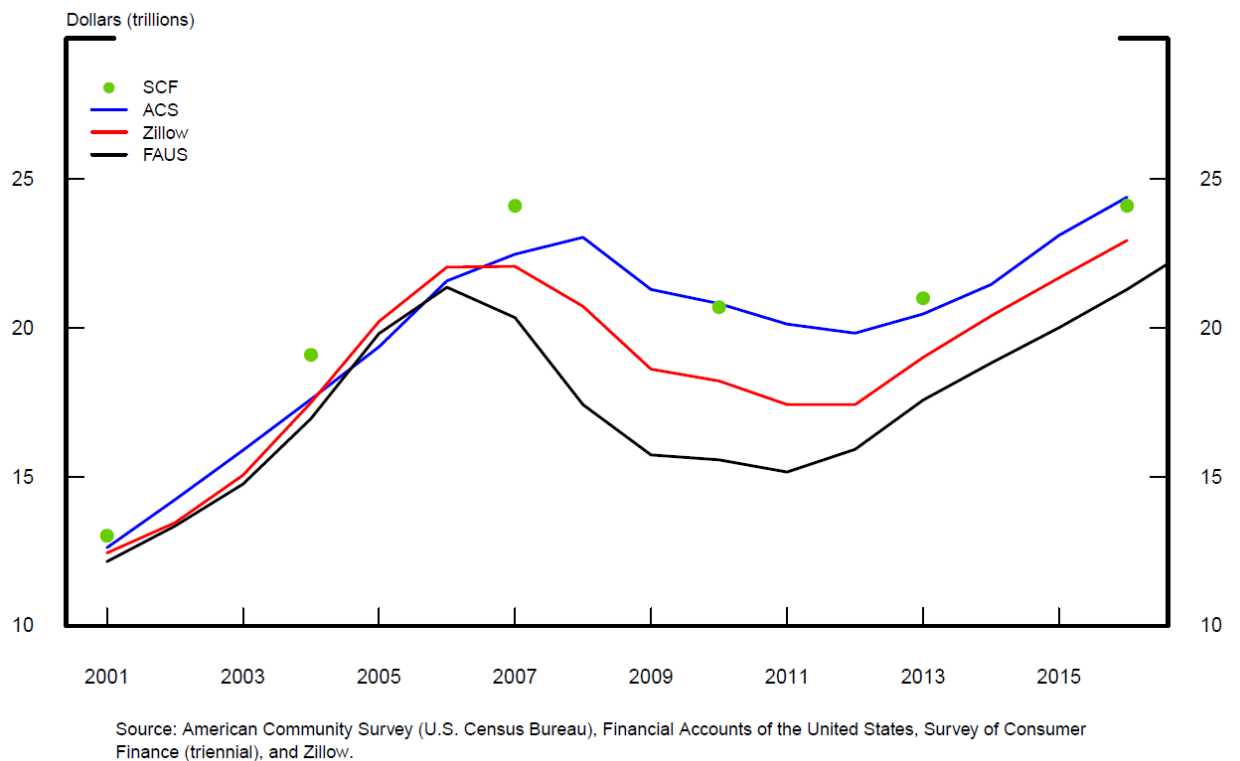
Figure 3 shows the aggregate own-use housing wealth measure based on Zillow's AVM (red line), compared to similar measures from the *Financial Accounts* (black line), from an aggregation of the ACS (blue line) and from the Survey of Consumer Finances (green dots).<sup>29</sup> The AVM, ACS, and *Financial Accounts* methods are quite close to each other from 2001 to 2006; all three rose by about 75 percent from 2001 to 2006. After 2006 the three measures diverge notably because the measures differ on the timing of the market peak and the severity of the subsequent bust. Whereas the AVM and *Financial Accounts* date the peak of the cycle in 2006, the ACS measure shows continued robust growth until 2008. Regarding the magnitude of the bust, the ACS measure fell by 14 percent from peak to trough, the *Financial Accounts* fell by 29 percent from peak to trough, and the AVM measure splits the difference between these two, falling by 21 percent from peak to trough. The measures from the ACS, AVM, and *Financial*

---

<sup>29</sup> To create a measure of aggregate house value from the ACS, we use the same procedure as with the AVM-based estimate except that we use average owner values (of owner-occupied property) by county and property type from the ACS instead of average AVM estimates. We calculate the average ACS estimates from the property level, public use data, which only provide a county identifier for large counties. For counties not identified in the ACS, we use the average value of properties that are identified as being in the same state but that are not in an identified county. Our SCF measure uses only the reported value of the primary residence. Income producing and seasonal properties are stripped from the aggregate to better align the measure with our wealth concept.

*Accounts* are in closer agreement in the dating of the market trough, with the measure from the *Financial Accounts* clearly turning up after 2011 and the AVM and ACS measures clearly turning up after 2012. Interestingly, the growth rates of all three of these measures are fairly similar from 2013 to 2016, all having risen by about 20 percent during this period. Overall, housing wealth estimates from the SCF estimates lie fairly close to those from the ACS (with the possible exception of 2004 and 2007), though it is difficult to assess the exact timing of the housing cycle using the SCF because the survey is triennial.

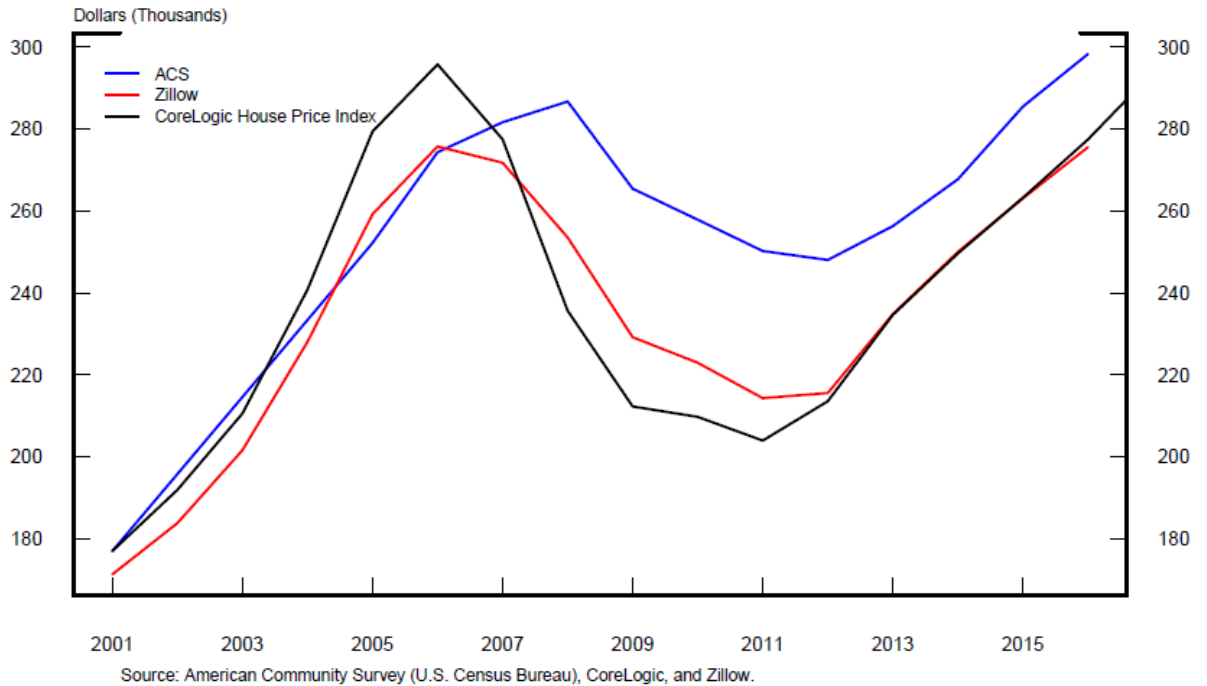
**Figure 3: Aggregate Own-Use Housing Wealth**



While we do not know which of these measures of aggregate housing wealth is the most accurate, one factor in favor of the AVM measure is that it is based on frequently updated data and models that are regularly tested against actual market transaction prices and improved based

on their performance. As the relative scarcity of existing studies illustrates, comparisons to market prices for survey data and repeat-sales indexes are infrequent and usually limited to particular geographic areas, so it is difficult to assess the reliability of those estimates.

**Figure 4: Average Value of Single-Family Homes**



Comparing the AVM-based measure to the *Financial Accounts*, which moves largely in response to a repeat-sales price index, illustrates the effects of using price indexes to extrapolate housing wealth. In this case, the timing of the swings in wealth are similar, which is not terribly surprising given that both measures are based on market prices. However, the larger swings in wealth in the *Financial Accounts* measure highlight the possibility that the repeat-sales price index methodology may not accurately reflect changes in the values of non-transacting homes, as discussed above. The higher volatility of a repeat-sales index can be seen more clearly in Figure 4, which shows the CoreLogic house price index (rebased to equal the average house value from

the ACS in 2001) and the average value of single-family homes in the ACS and Zillow.<sup>30</sup>

Compared with the Zillow average value, which measure the value of all homes, the CoreLogic index rises more during the housing boom, falls more during the contraction, and rises more during the recovery.

Comparing the AVM-based measure to the ACS-based measure illustrates some of the difficulties with using survey data to measure housing wealth. Specifically, changes in the ACS-based measure lag the AVM-based measure, and the magnitudes of changes in wealth are noticeably smaller. Indeed, the ACS and SCF measures indicate that survey respondents thought that the values of their homes, on average, were still rising even as major financial institutions were being toppled by the housing crisis and the entire financial system was being pushed to the brink. This suggests that survey respondents were either unaware of the market fluctuations in real time, or they believed—perhaps correctly, perhaps not—that their home values were different from what sales in the surrounding market would imply. To the extent that they did acknowledge changes in the market, it appears that they were late to catch on. By contrast, the AVM-based measures lines up fairly well with the measures based on owner reports during the housing boom. This result echoes our finding above that the average errors of the AVM during the 2000-2003 period were similar in magnitude to the average errors of owner reports documented by Goodman and Ittner (1992) for the 1980s. Thus, it appears that the factors that cause homeowner perceptions of value to differ from market value vary depending on housing market conditions.

## **VI. Conclusions**

---

<sup>30</sup> The CoreLogic house price index, which is used in the construction of the *Financial Accounts*, does not include units in multifamily structures.



We have shown how an AVM using public records data can be used to calculate aggregate housing wealth, and how doing so sheds new light on the strengths and weaknesses of existing measures. To do so, we develop a methodology that adjusts the raw estimates to be nationally-representative and accounts for the average bias of the particular AVM that we use. The resulting measure of housing wealth looks similar to measures based on owner's reported values during the early to mid-2000s, but differs substantially thereafter. In particular, it peaks two years earlier than a measure based on owner-reported values and falls by much more during the housing contraction. These results hint that homeowners may be slow to catch on to market turning points, and may not realize the full extent of a downturn in value. By contrast, the peak-to-trough decline in AVM-based value is not as large as that derived from a repeat-sales price index. More generally, the average value based on an AVM is less volatile than a repeat-sales price index, suggesting that the market prices of homes that transact may not be representative of changes in value of homes that do not transact.

Thus far, our analysis has focused on a measure of "own-use" housing wealth because that is the most comparable measure to household-owned housing wealth in the *Financial Accounts*. In the future, we plan to use a similar methodology to calculate a measure of aggregate household housing wealth that includes rental property. Such a measure might be more useful than own-use housing wealth for understanding household consumption decisions, since households probably take rental property assets into account when making consumption decisions.

Measures of housing wealth are a key ingredient for a wide range of empirical and simulation-based studies of subjects such as consumer finance, economic growth, business cycles, mortgage lending, wealth inequality, economic mobility, business formation, investment

in education, geographic mobility, and tax policy. Our results suggest that these literatures might need to be re-examined to see if their conclusions are robust to different ways of measuring housing wealth.

## References

Abraham, J.M., and W.S. Schauman. (1991). "New Evidence on Home Prices from Freddie Mac Repeat Sales," AREUEA Journal 19(3), 333-352.

Anenberg, Elliot, Joseph Nichols, and Lindsay Relihan. "Decomposing the error in self-reported home values." working paper. Washington: Board of Governors of the Federal Reserve (2012).

Bricker, Jesse, Brian Bucks, Arthur B. Kennickell, Traci L. Mach, and Kevin B. Moore (2011). "Surveying the aftermath of the storm: changes in family finances from 2007 to 2009," Finance and Economics Discussion Series 2011-17. Board of Governors of the Federal Reserve System (U.S.).

Bucks, Brian, and Karen M. Pence. "Do homeowners know their house values and mortgage terms?" Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, 2006.

Carroll, Christopher D., Misuzu Otsuka, and Jiri Slacalek. "How large are housing and financial wealth effects? A new approach." Journal of Money, Credit and Banking 43.1 (2011): 55-79.

Case, Bradford, Henry O. Pollakowski, and Susan M. Wachter. "Frequency of transaction and house price modeling." The Journal of Real Estate Finance and Economics 14.1-2 (1997): 173-187.

Clapp, John M., Hyon-Jung Kim, and Alan E. Gelfand. "Predicting spatial patterns of house prices using LPR and Bayesian smoothing." Real Estate Economics 30.4 (2002): 505-532.

Davis and Quintin (2016). "On the Nature of Self-Assessed House Prices," Manuscript on Marginal Q.

Dreiman, Michelle H., and Anthony Pennington-Cross. "Alternative methods of increasing the precision of weighted repeat sales house prices indices." The Journal of Real Estate Finance and Economics 28.4 (2004): 299-317.

Favilukis, Jack, Sydney C. Ludvigson, and Stijn Van Nieuwerburgh. "The macroeconomic effects of housing wealth, housing finance, and limited risk sharing in general equilibrium." Journal of Political Economy 125.1 (2017): 000-000.

Ferreira, Fernando, Joseph Gyourko, and Joseph Tracy. "Housing busts and household mobility." Journal of Urban Economics 68.1 (2010): 34-45.

Gatzlaff, Dean H., and Donald R. Haurin. "Sample selection bias and repeat-sales index estimates." The Journal of Real Estate Finance and Economics 14.1-2 (1997): 33-50.

Genesove, David, and Christopher Mayer. "Loss aversion and seller behavior: evidence from the housing market." The Quarterly Journal of Economics 116.4 (2001): 1233-1260.

Glennon, Dennis and Kiefer, Hua and Mayock, Tom. "Measurement error in residential property valuation: an application of forecast combination." (2016).

Glover, Andrew, et al. "Intergenerational redistribution in the great recession." No. w16924. National Bureau of Economic Research, 2011.

Goodman, John L., and John B. Ittner. "The accuracy of home owners' estimates of house value." *Journal of Housing Economics* 2.4 (1992): 339-357.

Henriques, Alice M. "Are homeowners in denial about their house values? Comparing owner perceptions with transaction-based indexes." *Finance and Economics Discussion Series* 2013-79 (2013). Board of Governors of the Federal Reserve System (U.S.).

Holmquist, Elizabeth, and Susan McIntosh. "U.S. Net Wealth in the Financial Accounts of the United States." *FEDS Note* October 8, 2015. Board of Governors of the Federal Reserve System (U.S.).

Hur, Sewon. 2016. "The lost generation of the Great Recession." working paper. Pittsburgh: University of Pittsburgh.

Iacoviello, Matteo, and Stefano Neri. "Housing market spillovers: evidence from an estimated DSGE model." *American Economic Journal: Macroeconomics* 2.2 (2010): 125-164.

Kiel, Katherine A., and Jeffrey E. Zabel. "The accuracy of owner-provided house values: The 1978–1991 American Housing Survey." *Real Estate Economics* 27.2 (1999): 263-298.

Kuzmenko, Tatyana and Christopher Timmons. 2011. "Persistence in housing wealth perceptions: evidence from the census data." working paper. Durham: Duke University.

Peterman, William B. and Kamila Sommer (2014). "How well did social security mitigate the effects of the Great Recession?" *Finance and Economics Discussion Series* 2014-13. Board of Governors of the Federal Reserve System (U.S.).

Saez, Emmanuel, and Gabriel Zucman. "Wealth inequality in the United States since 1913: evidence from capitalized income tax data." *The Quarterly Journal of Economics* 131.2 (2016): 519-578.

## **Appendix**

This appendix discusses some aspects of our methodology in greater detail. In particular, we describe here our methods for adjusting the Zillow average values for model bias and calculating the effects of rental bias.

### Rental Bias:

A key assumption in our paper is that the Zillow average property values accurately estimate the relevant own-use average values. Since Zillow’s model does not distinguish between rental properties and ow-use properties, large differences in average value by ownership status will tend to bias our results. As discussed in Section IV. C, it appears that rental homes are on average substantially less valuable than owner occupied properties. Estimates of the aggregate value of own-use housing will be biased downwards by the inclusion of rental properties in Zillow’s averages.

We therefore implement the following procedure to estimate the effect of the inclusion of rental properties in Zillow’s data. Our procedure makes use of a property-level match between a different (though broadly similar) AVM and ACS microdata. This matched data allows us to estimate the average AVM values separately for owner-occupied and rental-occupied properties as indicated in the ACS. (Such a calculation is not possible using the Zillow data because we do not have property-level estimates). In greater detail, our procedure consists of the following steps:

1. Match the 2014 ACS to the 2014 alternative AVM estimates at the property level.
2. For each geography  $g$  and property type  $s$ , calculate the ratio of AVM values for rental properties to owner-occupied properties:

$$\delta_{g,s} = \frac{\bar{V}^{CL}(s, g, rental)}{\bar{V}^Z(s, g, ownocc)}$$

3. For each county  $c$ , property type  $s$ , and time period  $t$ , let  $\beta_{c,s,t}$  be the estimated share of properties in Zillow’s data that are owner-occupied. We calculate these shares differently for single-family and multifamily properties. The single-family splits are calculated using the same owner-occupied/rental splits from the ACS that we use to calculate the single-family aggregate values. We do not use the ACS splits for multifamily because we do not think that the full universe of multifamily properties covered in the ACS is likely to be representative of the set of multifamily properties in Zillow’s average value estimates. Instead, we use the Census Bureau’s 2012 Rental Housing Finance Survey (RHFS) to estimate the number of non-condo rental units in 2+ buildings at 20,799,737. The 2012

ACS has 25,354,734 occupied rental units, suggesting that roughly 82% of the ACS universe is non-condos (and therefore likely to not be in Zillow’s deeds-based property records). This 82% estimate does not account for vacant units, however. The Census Housing Vacancy Survey reports a 9.3% vacancy rate on all 2+ multifamily units in 2012. We therefore estimate the total number of multifamily rentals included in the Zillow average by adjusting down the relevant ACS multifamily rental totals by a factor of  $1 - 0.907 * 0.820 = 0.256$ . The effect of this adjustment factor of 0.256 is to modestly decrease the multifamily inflation factors.

4. Let  $g(c)$  denote the geography  $g$  from the 2014 ACS/AVM matched data corresponding to county  $c$ . For example, if  $g$  are states then  $g(c)$  is the state containing  $c$ . Using  $\beta_{c,s,t}$  and  $\delta_{g(c),s}$ , the property type/county/time period adjustment factor is constructed as

$$\lambda_{c,s,t} = \frac{1}{\beta_{c,s,t} + (1 + \beta_{c,s,t})\delta_{g(c),s}}$$

Using the  $\lambda_{c,s,t}$  to adjust the average values from Zillow results in an aggregate estimate that is roughly 6 percent higher than what we report.

#### Model Bias Adjustment:

Zillow’s AVM estimates are biased, particularly for properties in the bottom 30% of the value distribution. To account for this bias, Zillow adjusts their “raw” estimates using the median observed error within a given geography/property type/time.<sup>31</sup> Zillow reports to us both the median-adjusted and “raw” average value estimates. Since we are concerned with average bias, we use the raw Zillow numbers and adjust for average bias using a procedure described below. Ideally, we would adjust the Zillow averages to account for the expected value-weighted model error. We cannot directly observe the value-weighted average errors for two reasons. First, the errors are only calculable against observed transactions. Second, we do not have the full property-by-property error distributions from Zillow. Rather, Zillow reports to us the average

---

<sup>31</sup> Please refer to <https://www.zillow.com/research/zhvi-methodology-6032/> for more details about Zillow’s bias adjustment procedure. It should be noted also that bias may permit greater accuracy in a mean squared error sense. Our objectives are different than Zillow’s, as we care about bias much more than variance. Adjustments which are sensible for us may not be sensible for Zillow.

percent error by transaction decile, along with each decile's upper and lower bounds, by quarter and county, separately for single-family and multifamily properties from 2007Q3 on and for all property types combined before 2007Q3 (evaluated using the earlier-vintage Z5 model).<sup>32</sup> We therefore construct our value-weighted adjustments using the following procedure:

1. Let  $V_{c,s,t,i}$  be the value defining the upper limit of decile  $i$  for county  $c$ , property type  $s$ , and year  $t$ .<sup>33</sup> Define the value share of decile  $i$  as  $w_{c,s,t,i} = \frac{(V_{c,s,t,i} + V_{c,s,t,i-1})/2}{\sum_j (V_{c,s,t,j} + V_{c,s,t,j-1})/2}$ , where  $V_{c,s,t,0}$  is set to 0 and  $V_{c,s,t,10}$  is set equal to  $1.5V_{c,s,t,9}$ . Setting the upper bound on the value distribution at 50% above the 90<sup>th</sup> percentile is a kludge to avoid giving too much weight to the very top of the value distribution, which typically has a very long tail. Our results are not sensitive to this particular choice; we obtain quantitatively similar results using either  $1.2V_{c,s,t,9}$  or  $2.0V_{c,s,t,9}$  as the upper bound.
2. Estimate the value-weighted average error as  $E_{c,s,t} = (\sum_i w_{c,s,t,i} APE_{c,s,t,i})$ , where  $APE_{c,s,t,i}$  is the average percent error in value decile  $i$ . Some of the error distributions are based on very few transactions and are therefore likely estimated with considerable error. In response, we set  $E_{c,s,t}$  to missing if the number of transactions in  $(c, s, t)$  is less than 20. For counties not covered by Zillow's newest model (Z6) or set to missing due to a paucity of transactions, use the average value-weighted error for counties in the same state which do have error distribution data.<sup>34</sup> Let  $\tilde{E}_{c,s,t}$  denote the average errors including these imputations.

---

<sup>32</sup> The county-level average errors prior to 2007 are based on the older version of Zillow's AVM and, unlike county-level average property values, are not available for many counties.

<sup>33</sup> The errors-by-percentile data come to us at a quarterly frequency. The problem that some geographies have too few transactions to accurately estimate a value-weighted adjustment is more extreme at the quarterly frequency. Therefore, we aggregate to a yearly level by summing transactions within a geography separately across each decile bucket. This procedure is not exactly right because the upper and lower bounds of the decile buckets change from quarter to quarter. However, these bounds in practice change very little because they are defined relative to the full distribution of Zillow AVM estimates, rather than relative to the distribution of observed transactions.

<sup>34</sup> Averaging is not quite right, because the decile upper and lower bounds are different across different counties within the same state.

3. Define the adjustment factor  $\gamma_{c,s,t} = \frac{\tilde{E}_{c,s,t}}{APE_{c,s,t}}$ , where  $APE_{c,s,t}$  is the unadjusted average error regardless of the number of transactions (so the floor of 20 is dropped). If the county is not covered by Zillow, set  $APE_{c,s,t}$  equal to the statewide average error prior to computing  $\gamma_{c,s,t}$ . Note that  $(\gamma_{c,s,t}APE_{c,s,t})$  is the estimated value-weighted average error for  $(c, s, t)$ . The only reason to go from  $\tilde{E}_{c,s,t}$  to  $\gamma_{c,s,t}$  and back to  $\gamma_{c,s,t}APE_{c,s,t}$  instead of using  $\tilde{E}_{c,s,t}$  directly is to make use of the unweighted average errors of counties with fewer than 20 observations.
4. Define the combined (sf and mf) value weighted error  $\hat{E}_{c,t}$  as the weighted sum of  $(\gamma_{c,s,t}APE_{c,s,t})$  for sf and mf properties, where the weights are each property type's share of the total observed transactions in county  $c$  and year  $t$ .
5. Finally, adjust the average errors by  $\left(\frac{1}{1+\hat{E}_{c,t}}\right)$ . We combine the mf and sf errors into one county-level series for a number of reasons. First, the share of counties with sufficiently many multifamily transactions to accurately estimate value-weighted errors is quite small. Second, as Figure 8 (below) shows, the value-weighted errors do not appear to be very different for multifamily and single-family properties. Finally, we do not have a breakdown of errors by property type before 2007Q3. As much as possible, we wish to apply the same bias adjustment method everywhere in our sample.
6. For quarters prior to 2007Q3, follow steps (1)-(5) above for the Z5 data, but without breaking out errors separately by property type.



**Figure 8: Average Value-Weighted Errors for Various Models**

