Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Measuring Aggregate Housing Wealth: New Insights from Machine Learning

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

2018-064

Please cite this paper as:
Gallin, Joshua H., Raven Molloy, Eric Nielsen, Paul Smith, and Kamila Sommer (2018). "Measuring Aggregate Housing Wealth: New Insights from Machine Learning," Finance and Economics Discussion Series 2018-064. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2018.064r1.

Measuring Aggregate Housing Wealth: New Insights from Machine Learning

Joshua Gallin, Raven Molloy, Eric Nielsen, Paul Smith, Kamila Sommer*

July 26, 2019

Abstract

We construct a new measure of aggregate housing wealth for the U.S. based on (1) home-value estimates derived from machine learning algorithms applied to detailed information on property characteristics and recent transaction prices, and (2) Census housing unit counts. According to our new measure, the timing and amplitude of the recent house-price cycle differs materially but plausibly from commonly-used measures, which are based on survey data or repeat-sales price indexes. Thus, our methodology generates estimates that should be of considerable value to researchers and policymakers interested in the dynamics of aggregate housing wealth.

JEL Codes: C82, E21, R31. Keywords: Residential real estate, Consumer economics and finance, Data collection and estimation, Flow of funds.

---

*Please do not cite without the permission of the authors. Earlier versions of this paper were circulated under the title "Measuring Aggregate Housing Wealth: New Insights from Automated Valuation Models."

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.

1 Introduction

Owner-occupied housing is a major component of households' balance sheets.[1] As a result, changes in aggregate housing wealth can affect aggregate consumption and savings and, by extension, macroeconomic outcomes such as economic growth and business cycles. However, housing wealth is quite difficult to measure (as we will discuss below), which has made it difficult for researchers to reliably observe its dynamics. In an effort to improve the measurement of housing wealth, this paper introduces a new method to make use of local property value estimates that are derived from machine learning algorithms applied to detailed data on property sales and characteristics from public records and other sources. We combine these property value estimates with housing unit counts from the Census to derive new estimates of aggregate U.S. housing wealth from 2001 to 2018. Our estimates show considerably more responsiveness to changing market conditions than survey measures and somewhat less volatility than repeat-sales measures, highlighting some of the key biases plaguing these commonly used estimates of housing wealth. Thus, our methodology generates estimates that should be of considerable value to researchers and policymakers interested in the dynamics of housing wealth and the role that it plays in economic outcomes.

The difficulty in measuring aggregate housing wealth stems from the inherent difficulty in measuring individual property values. Transaction prices, which are the best measure of a home's value, are relatively infrequent for a given property, with years or even decades between sales. Consequently, commonly used measures of individual house values have typically been based on homeowners' reports from surveys or extrapolated from previous sales using changes in a repeat-sales price index. Research has found both of these methods to be flawed in distinct ways. For example, studies have found that owner-reported estimates of house values are biased up on average, perhaps because owners are overly optimistic. Moreover, owners appear to have difficulty identifying market turning points, causing the bias to fluctuate over the housing cycle.[2] 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 always representative of those that are not sold. This bias may also

[1] According to the 2018q4 Distributional Financial Accounts, owner-occupied real estate accounted for 53 percent of the assets of the bottom half of the wealth distribution, and 32 percent of the assets of those in the 50^{th}-90^{th} percentile of wealth. See https://www.federalreserve.gov/releases/efa/efa-distributional-financial-accounts.htm.

[2] See, for example Goodman and Ittner (1992); Kiel and Zabel (1999); Bucks and Pence (2006); Kuzmenko and Timmins (2011); Henriques (2013); Chan et al. (2016).

be cyclical, as the degree of difference between transacting and non-transacting homes may shift systematically over the housing cycle.[3] By extension, aggregate housing values constructed directly from survey data or by extrapolating from a given base using a repeat-sales house price index will also be affected by these same biases.

The method of measuring housing wealth that we develop in this paper uses an automated valuation model (AVM), which can be loosely thought of as an algorithm that combines information on a home's characteristics, neighborhood features, nearby sales, and homes listed for sale to produce an estimate of the home's current market value. Although versions of AVMs have been in use for decades, private tech companies have recently created much more sophisticated and comprehensive AVMs by combining very large and detailed property-level datasets with 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 techniques offers the potential for more accurate estimates of housing values – especially during market turning points – than those based on surveys or repeat-sales indexes.

The estimates of aggregate housing wealth that we construct are based on an AVM created by Zillow, a private real estate and data analytics firm that provides estimated home values for over 100 million properties in the U.S. Constructing our measure of aggregate housing wealth is not as simple as adding up the value estimates of all properties in the Zillow data. Zillow's AVM coverage, while extensive, is not universal. Moreover, Zillow's estimates include some rental properties that we do not want to include in our measure of aggregate housing wealth and that we cannot easily identify. We address these issues by combining the AVM data with data from the Census Bureau's American Community Survey (ACS). Specifically, we calculate the quantity of owner-occupied housing units by property type and county from the ACS and multiply these quantities by the average value of homes in each market segment (county and property type) as determined by Zillow's AVM. In Sections 4 and 5, we validate our method by carefully investigating the properties of the AVM estimates and the representativeness of the sample on which these estimates are based. Where we can, we test for potential bias in our new measure.

Our method yields new high-frequency (monthly) estimates of aggregate owner-occupied

[3]See, for example and Case et al. (1997); Gatzlaff and Haurin (1997); Dreiman and Pennington-Cross (2004); Glennon et al. (2018).

housing wealth from 2001 to 2018, thereby offering a fresh look at the dynamics over the recent housing cycle. We find that from 2001 to 2006, the AVM estimates are largely in line with estimates based on owners' reported values in surveys such as the annual ACS, the biennial American Housing Survey (AHS), or the triennial Survey of Consumer Finances (SCF). By contrast, our measure diverges notably from survey measures from 2006 to 2012, a time period that included an enormous housing bust and a gradual recovery. In particular, the AVM-based measure turns down earlier and falls by much more than a measure based on owner reports. This result is consistent with prior research suggesting 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 owners did acknowledge changes in the market in their survey responses, it appears that they were late to do so.

The AVM-based measure also differs from the measure of housing wealth reported in the Federal Reserve's *Financial Accounts of the United States*, which is largely driven by changes in a repeat-sales house price index from 2005 onward. Specifically, while the contraction in wealth and subsequent recovery in the AVM measure is more pronounced than it is in the survey measures, the cycle is less pronounced in the AVM measure than in the Financial Accounts. We interpret this result as illustrating the possibility that repeat-sales indexes overstate the effect of market changes on aggregate housing wealth because they inaccurately extrapolate the house price dynamics of transacting homes to all homes.

We view one of the contributions of this paper as showing how data that are collected in the private domain for other purposes can be combined with survey data to produce an aggregate time series for use in national statistics. Researchers are currently engaged in applications that attempt to make use of such data to measure a variety of aggregate outcomes including retail spending, services consumption, employment, and business formation.[4] Consistent with these studies, one lesson from our analysis is that privately generated data may still need to be augmented with other data sources in order to construct nationally representative statistics.

Another contribution is showing how machine learning techniques (as used in Zillow's AVM) can be used to improve estimates of aggregate housing wealth.[5] Perhaps most importantly, our

[4]For examples, see Aladangady et al. (2019); Batch et al. (2019); Cajner et al. (2019); Gindelsky et al. (2019); Glaeser et al. (2019).

[5]Machine learning has been used in a variety of applications from prediction to causal estimation. Notable recent

findings suggest that studies that use self-reported values from surveys like the SCF to examine fluctuations in house values or housing wealth could be understating quite severely the cyclical changes in aggregate wealth. Since housing wealth is such a big part of total household wealth, any bias in owner-reported home values in turn affects measurement of cyclical dynamics of household net worth. Our results also suggest that economists and policy-makers using the *Financial Accounts* or repeat-sales price indexes to measure aggregate wealth might overstate the size of the housing wealth cycle. Especially because housing wealth measurement has the potential to affect results in a wide range of studies, the discussion and findings in this paper should be of interest to researchers and policy makers interested in quantifying the economic effects of cyclical fluctuations in wealth.

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. Sections 5, 6 and 6 discuss possible biases in our method as well as corrections we develop to address these biases. Section 7 describes our estimates of housing wealth for the United States and how these measures compare to the Financial Accounts and aggregated survey data. Section 8 concludes with a discussion of the implications of our paper for other studies and future uses of large-scale AVMs in economic research.

2 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. For example, in the 2017 American Housing Survey, only 10 percent of owner-occupied properties had transacted in the previous two years; for more than one third, the last transaction was in 2000 or earlier.

As noted above, measurements of housing wealth have typically derived from surveys of homeowners who are asked to estimate the value of their homes, or repeat-sales price indexes

examples are summarized in Kleinberg et al. (2015); Athey (2018); Athey and Luca (2019).

derived from recent market transactions. A long literature explores the issues associated with each approach. Regarding the implications of using house price indexes to estimate changes in aggregate home values, several papers have found that the samples of housing units used to form repeat-sales indexes are not representative of a broader set of homes. Homes that trade more frequently tend to have systematically higher house price appreciation (Korteweg and Sorensen, 2016; Case et al., 1997). Moreover, the selection bias associated with transacting more frequently is correlated with economic conditions and the housing cycle (Gatzlaff and Haurin, 1997; Malone and Redfearn, 2009), complicating the inference of cyclical changes in housing wealth from a repeat-sales house price index. Likely because of these biases and selection issues, Glennon et al. (2018) find very large differences between value estimates based on repeat-sales indexes and transaction prices during the housing crisis.

A further complication with using house price indexes to measure aggregate housing wealth is the fact that extrapolation using a price index requires starting from a nationally-representative base at some point in time – the index on its own cannot speak to the level of housing wealth. Such a base – which is frequently constructed from available survey data – is in itself quite difficult to accurately measure, for reasons we will discuss below. Moreover, extrapolating forward from any base using a price index introduces the problem that price indexes are designed to abstract from changes in the quantity and quality of the housing stock.[6] But changes in the aggregate value of housing should include changes in quantity and quality, so attempts to extrapolate wealth using a price index must somehow account for these factors using other data sources.[7] This issue may not be so important for extrapolating housing wealth over a few quarters since the housing stock changes slowly over time. But the longer the time period, the more likely the lack of information on quality will matter.

Measurement of housing wealth directly from surveys addresses the problems associated with the use of house price indexes, while introducing new issues related to the accuracy of homeowner reporting. On the one hand, nationally representative surveys contain value estimates representing

[6]Specifically, a repeat-sales index assumes that the quality of a housing unit is constant between transaction pairs, no matter how much time has elapsed between the two sales dates.

[7]In the *Financial Accounts* after 2005, quantity and quality adjustments come only in the form of estimates of net fixed investment from the Bureau of Economic Analysis. That is, the *Financial Accounts* employs a perpetual inventory approach in which changes in housing wealth come from capital gains, estimated using a repeat-sales index, and net fixed investment, which includes estimates of the value of additions, renovations, and construction of new units.

all owner-occupied homes. The sum of these survey responses thus yields a straightforward estimate of aggregate value for every survey year. Since these reported values in principle reflect the changing characteristics of the housing stock over time, the resulting aggregate estimates should account for changes in quality. On the other hand, this approach to measuring aggregate housing wealth will only work if owner valuations are unbiased. A long line of research finds evidence of systematic bias in owner-reported house values, although the precise magnitude of this bias is difficult to assess.

Some studies assess the bias of owner valuations by comparing owner-occupant estimates directly with subsequent sale prices. Goodman and Ittner (1992) compare survey respondents' house valuations to subsequent sales prices (over the next two years) using data from the 1985 and 1987 waves of the AHS and find that, on average, owners over estimate the value of their homes by about 8 percent. More recently, Molloy and Nielsen (2018) compare owner estimates in the 2014 ACS with sale prices in 2016 and find that the average owner overestimates the value of their home by 6 percent. This general approach to assessing bias in self-reports is not foolproof, however. Typically, a house price index must be used to extrapolate the owner-reported value forward in time to the sale date.[8] While focusing on transactions within two years of the survey date mitigates potential bias from the use of price indexes (discussed above), the resulting transaction/survey pairs are few in number and may not be representative of the full survey sample. For example, in anticipation of a future move, homeowners who are only a few years from selling may be better informed about their local housing market, suggesting a smaller bias than what would be typical for respondents who do not intend to sell.

Other studies attempt to assess bias in owner reports by comparing them to a home value that is extrapolated from a previous sale price, also using a house price index. The degree of owner overvaluation estimated in this manner can be quite large, ranging between 3 and 16 percent (Ihlanfeldt and Martinez-Vazquez, 1986; Kiel and Zabel, 1999; Benitez-Silva et al., 2015; Chan et al., 2016; van der Cruijsen et al., 2018). Moreover, the overvaluation appears to increase during market downturns when owner valuations do not fall as much as price indexes (Henriques, 2013;

[8] Goodman and Ittner (1992) inflate the owner valuations using a metropolitan area house price index while Molloy and Nielsen (2018) adjust for the time difference between the owner valuation and the sale using a county-level Zillow Home Value Index. Neither adjustment is accurate if the true value of the home does not appreciate in line with the associated price index.

Chan et al., 2016; Davis and Quintin, 2017). In these studies, there is usually a fairly long gap between the survey date and the previous sale date, so any bias from extrapolation using a price index would be greatly amplified.

In summary, prior literature has documented the potential for considerable biases in existing measures of housing wealth, although the estimated sizes of these biases range widely.

3 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 have existed for constructing estimates of aggregate housing wealth. Recently, however, AVMs created by private firms 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, in our view, they can plausibly provide a viable method for measuring aggregate housing wealth. In particular, companies such as Zillow have assembled very large property-level datasets containing tens of millions of records and combined them with sophisticated machine learning models to 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.

Zillow's AVM attempts to assign values to all single-family homes, as well as co-op and condominium apartments.[9] As a first pass, one can think of the AVM as resembling a hedonic regression relating house values to a rich set of property and neighborhood characteristics, estimated from comprehensive data on observed sale prices in the local area around each home. The estimated model is then applied to properties that do not have a recent sale price to produce value estimates for the full stock of homes for which sufficient data on characteristics are available or can be imputed.[10]

In practice, Zillow's AVM is not a single hedonic model, but a very large number of distinct models that work together to produce a value estimate for each property. The individual sub-

[9]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.

[10]For very general background information on Zillow's AVM, see https://www.zillow.com/zestimate/.

models are estimated using standard machine learning techniques that use the underlying data in different ways. The resulting estimates from the sub-models are then combined into a single final estimate based on out-of-sample predictive performance and data quality filters. While the exact details of Zillow's estimation are proprietary, their AVM is an ensemble model, the type of model that has been shown to outperform other common machine learning approaches in predicting house prices (Mullainathan and Spiess, 2017).[11] The AVM will not assign a value to a property in the event of too much uncertainty about the estimated value, e.g., from missing data or unusual property characteristics. The properties excluded for these reasons tend to be in less populated areas where transactions are sparse and data quality is poor.

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 both property characteristics and transaction details such as sales prices and dates. 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 always capture the myriad property characteristics that affect a home's value.[12] 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 website invites homeowners to update or correct the characteristics of their property that might be missing or inaccurate in Zillow's database.

Zillow updates and re-estimates their models daily to onboard new data as it becomes available. These daily runs allow Zillow to continually assess their model errors for bias and update their algorithms to maximize the accuracy of the prediction at any point in time.

The available (albeit limited) information to date suggests that AVMs are at least somewhat

[11] Mullainathan and Spiess (2017) also provide a very helpful summary of machine learning techniques. They argue that ensemble models, such as Zillow's, tend to perform very well in virtually all prediction exercises. In general, the machine learning techniques employed by Zillow follow best practices as outlined in Mullainathan and Spiess (2017) and Athey (2018).

[12] 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 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/.

better than other methods of housing valuation during normal times, and can be considerably better during market downturns. Glennon et al. (2018) evaluate repeat sales indexes by extrapolating prior sales prices with a repeat-sales index and compare these valuations to sales prices. Across the four counties that they examine, they report average errors ranging from 3 to 7 percent in 2005, and from 26 to 113 percent in 2010.[13] By comparison, using data provided by Zillow that we will discuss below, for those same four counties we calculate average AVM errors ranging from -7 to 2 percent in 2005 and from 9 to 19 percent in 2010. Molloy and Nielsen (2018) analyze a sample of homes with a different AVM and owner valuations in 2014 that subsequently sold in 2016. Although the average errors were about 6 percent using either valuation method, the distribution of AVM errors was centered very close to zero, whereas the distribution of owner valuation errors was centered around 2 percent (i.e. the valuation was 2 percent higher than the sales price).[14] On the whole, it seems that AVMs have the potential to materially improve estimates of aggregate housing wealth during market downturns, and may be an improvement over other existing methods even during normal times.

4 Using Zillow's AVM to Measure Aggregate Own-Use Housing Wealth

We use Zillow's AVM to construct a measure of the aggregate housing wealth held by households for their own use; i.e., excluding rental units.[15] Our measure is thus directly comparable to the most commonly used measures of aggregate housing wealth, which also exclude rental property. One such measure is the widely cited series from the Federal Reserve's *Financial Accounts of the United States*, which is a hybrid between survey data (through 2005) and a repeat-sales house price index after 2005.[16] Another common method to measure aggregate housing wealth is to aggregate estimates directly from surveys; since most surveys only ask owner occupants to report house

[13]A positive error means the valuation was higher than the sales price.

[14]The AVM accuracy tends to improve over time with better data and model improvements. See https://www.zillow.com/zestimate/#acc for up-to-date information about the accuracy of Zillow's AVM.

[15]We focus on the total value of real estate assets rather than home equity, which would subtract mortgage debt.

[16]The series we construct in this paper is directly comparable to the *Financial Accounts* series FL155035013, the component of total housing wealth that excludes vacant land and mobile homes. Over the period 2001-present, FL155035013 represents about 93 percent to 95 percent of total household housing wealth reported in Table B.101.h of the *Financial Accounts*.

values, these surveys do not provide data on the value of rental units.[17]

4.1 Representativeness of Zillow's Data

A key consideration in using Zillow to construct an aggregate time series is that Zillow's coverage may not be broad enough to be nationally representative. Another consideration is that Zillow's universe includes some rental properties (held by businesses or households), which we do not want to include in our measure of aggregate own-use housing wealth, and which cannot be straightforwardly identified in the Zillow data.

Table 1: Property Counts in 2017 (millions)

	ACS		Zillow
	Total	Own-use	Total
Single-Family	92.9	76.4	78.7
Multi-Family	35.9	5.8	7.7
Total	128.8	82.1	86.3

Table 1 illustrates how the Zillow data compare with the universe of own-use properties in 2017 as measured in the nationally-representative ACS. According to the ACS, there were about 93 million single-family homes in the U.S. in 2017, about 76 million of which were for households' own use and 17 million of which were for rental use.[18] Zillow's AVM is able to provide value estimates for about 79 million single-family homes in 2017 (including both own-use and rental). Thus, Zillow's overall coverage of the single-family market, at about 85 percent, is fairly 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.

Turning to the multifamily market, there were about 36 million multifamily homes in 2017 ACS, but only about 6 million were for own use. Zillow's sample for 2017 includes about 8 million multifamily housing units, including both own-use and rental properties. Zillow's overall coverage

[17] One notable exception is the Survey of Consumer Finances, in which survey respondents report the value of rental property owned by a household.

[18] Consistent with the *Financial Accounts* definition, we define properties in the ACS as "own-use" if they are (1) owner-occupied or vacant and (2) 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.

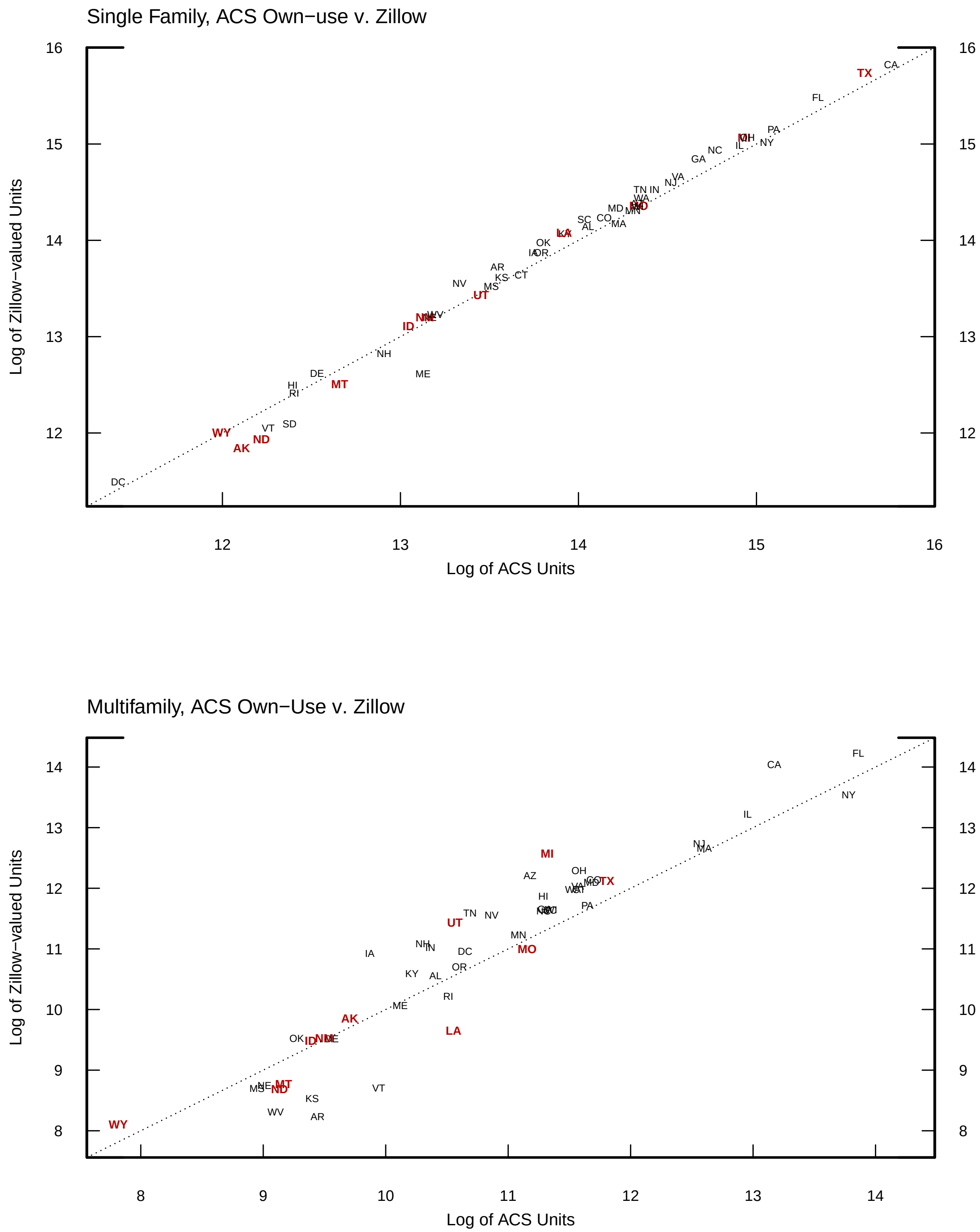
rate (about 20 percent) is much lower for multifamily properties by construction, because Zillow does not attempt to value apartments in rental buildings (defined as buildings in which a single property-tax parcel contains multiple units); rather, their focus is on multifamily housing units that are sold individually, like condos and co-ops. The exclusion of rental buildings from Zillow's valuation universe is a helpful feature for our purpose, as we do not include these units in our wealth measure.[19] For the units in multifamily buildings that we do want to include (i.e., condos and co-ops), the available evidence suggests that the Zillow's coverage is actually quite high. In particular, using the Census Bureau's 2012 Rental Housing Finance Survey, we estimate that roughly 25 percent of rental multifamily housing units were condos. If we apply this share to the 2017 ACS data, the estimated total number of condo and co-op units would be about 8 million, very close to the total number valued by Zillow (see Table 1).[20] However, the Zillow multifamily sample of 8 million is still larger than the ACS count of own-use multifamily homes, which is about 6 million units. This difference indicates that the Zillow multifamily sample still includes a significant number of rental units that we would like to exclude. In addition, it likely also misses some own-use properties that we would like to include.

Finally, in constructing an aggregate measure, it is important to assess whether Zillow's coverage varies systematically across different parts of the U.S. Toward that end, Figure 1 shows Zillow's coverage at the state level. Specifically, it shows the relationship between the number of properties covered by Zillow in 2017 in a given state and the number of own-use properties in the ACS for that state. The closer the state is to the dashed 45-degree line, the better the alignment of housing counts between Zillow and the ACS. Zillow's coverage of single-family own-use units (upper panel) is a little lower for smaller and/or less densely populated states (such as Alaska and North Dakota), but the relatively close proximity of most states to the 45-degree line indicates that Zillow's universe is generally quite representative of the ACS own-use universe. Zillow's coverage of multifamily units (lower panel) varies more across states, but is still generally high.

[19]These units are generally held by households only indirectly, through their holdings of corporate and non-corporate equity. As such, we do not classify them as direct household holdings of own-use housing wealth.

[20]We cannot estimate Zillow's overall coverage of condo and co-op units directly from the ACS data because the ACS does not include information that would allow us to reliably identify which of the units in multifamily buildings are condos or co-ops.

Figure 1: Ratio of Zillow to ACS Property Counts in 2017 by State



Bolded red denotes non−disclosure states.

4.2 Our New Method

Although Zillow's coverage appears high enough and broad enough to be nationally representative, the number of units in Zillow's data can change discontinuously over time as new data sources are integrated into their estimation framework. For example, the number of housing units in Zillow's sample jumped by 3.5 percent from May 2009 to June 2009 as their coverage expanded. While we want to take advantage of the potential for the addition of new units to improve the quality of the Zillow estimates, we do not want our measure of aggregate housing wealth to jump discontinuously because Zillow improved their sample by adding units that already existed. Moreover, we would not want our measure to be affected by the inclusion of rental properties, which are not part of our definition of own-use housing wealth.

To get around these issues, we divide the aggregate U. S. housing market into segments and calculate the wealth in each segment by multiplying the average value from Zillow's AVM by the number of own-use housing units derived from the nationally representative ACS. We define market segments as a combination of property type (single-family or multi-family) and county. Thus, for each county c and property type p, we estimate the value of own-use housing at time t as:

$$\hat{V}(p,c,t) = N^{ACS}(p,c,t)\bar{V}^{Z}(p,c,t),$$

where $N^{ACS}(p,c,t)$ is an estimate of the number of properties intended for own use from the ACS and $\bar{V}^{Z}(p,c,t)$ is the average AVM value for residential properties.

To construct the counts of properties intended for own use from the ACS in each county, 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. Consistent with the *Financial Accounts* concepts, we define the total number of properties intended for own use as the sum of units intended for own use plus a share (ϕ) of vacant properties that are not for sale or rent:

$$N^{ACS}(p,c,t) = N^{ACS}(p,c,t \mid \text{own use}) + N^{ACS}(p,c,t \mid \text{vacant})\phi(p,c,t).$$

We assume that the share of vacant properties that are intended for personal use, ϕ, is the same as the share of occupied properties for own use so that:

$$\phi(p,c,t) = \frac{N^{ACS}(p,c,t \mid \text{own use})}{N^{ACS}(p,c,t \mid \text{own use}) + N^{ACS}(p,c,t \mid \text{rental use})}.$$

If we had accurate housing counts of all own-use units (N^{ACS}) and unbiased average home values ($\bar{V}^Z$) for each property type and county, then the above method would yield unbiased estimates of aggregate housing wealth in each county. These county-level estimates could then be straightforwardly summed to produce an estimate of the national aggregate. However, this theoretical framework suffers from two sets of practical limitations. The first set of issues relates to potential bias in the average home value. Most obviously, any bias in the property-level AVM estimates could result in biased county-level averages. Another source of bias comes from the fact that Zillow is unable to produce reliable estimates for some properties in their sample, and the average value of these omitted properties might differ from the average value of included properties. Finally, Zillow's inclusion of some rental properties in their sample will tend to bias down our estimate of average value to the extent that rental properties have lower average values that own-use units. In Section 5, we examine each of these potential biases in more detail. The second set of practical limitations results from issues related to geographic coverage. Neither the ACS nor Zillow provide complete coverage of all counties in the U. S. – both are missing data for some counties, often in rural areas. We discuss these coverage issues in Section 6.

5 Testing for Bias in County-Level Average Home Values

This section examines three conditions that are required for the county-level average AVM estimates ($\bar{V}^Z$) to be unbiased estimates of the county-level own-use average values.

5.1 Property-Level Bias

Our aggregate wealth estimates will be biased if the property-level AVM estimates are biased. On several levels, assessing the bias for an AVM is considerably more straightforward than it is for surveys. As property sales occur, Zillow compares the sales price against the AVM estimate made at the end of the month prior to the sale to generate a distribution of out-of-sample model errors.[21]

[21] Similar to surveys, AVM estimates can only be compared to sales prices of properties that transact. One potential criticisms of any such error analysis is that homes that transact may not be representative of all homes in the market.

The near-contemporaneousness of the AVM measurement and the home's sale eliminates the need for a house price index to compare the AVM valuation with the actual sales price. Moreover, since Zillow has AVM estimates for nearly all sales in the deeds records, there are many more AVM-sales pairs than the number of owner valuation-sales pairs available in surveys.

To assess the importance of any AVM bias on our aggregate measure of housing wealth, Zillow has provided us with average errors by county, property type, and house value from 2001 to 2017.[22] Because our ultimate goal is to estimate aggregate housing wealth and because higher-value properties have a larger influence on aggregate wealth, we calculate value-weighted average errors by county and year based on the error distributions provided by Zillow; see the Appendix for details. For most counties, these value-weighted errors are fairly close to zero, suggesting accurate AVM predictions on average.[23] However, in some (especially smaller and sparser) counties, the average value-weighted error can be notably different from zero. Moreover, the size of the bias in a typical country appears to fluctuate over time. Figure 2 shows the evolution of the model errors over time, averaging across counties, where the error is defined as the AVM estimate minus the sale price, as a percent of the sale price. This value-weighted average was about -0.5 percent in the early 2000s, then dipped to about -2 percent in 2004 and 2005 as market prices were rising so briskly that the AVM did not fully keep up. In 2006, the average error turned positive and subsequently widened to about 5 percent as market prices fell more quickly than estimated by the AVM. As the market stabilized, the error rate declined to about 2 percent, and remained in the 2-3 percent range through 2017.

To reduce the bias imparted by these average errors, we bias-adjust the county average property values by multiplying the average AVM estimate for each county and property type by the inverse of one plus the value-weighted average error in that county and year. This bias adjustment, described in the Appendix, has only a muted effect on the level and trajectory of housing wealth.[24]

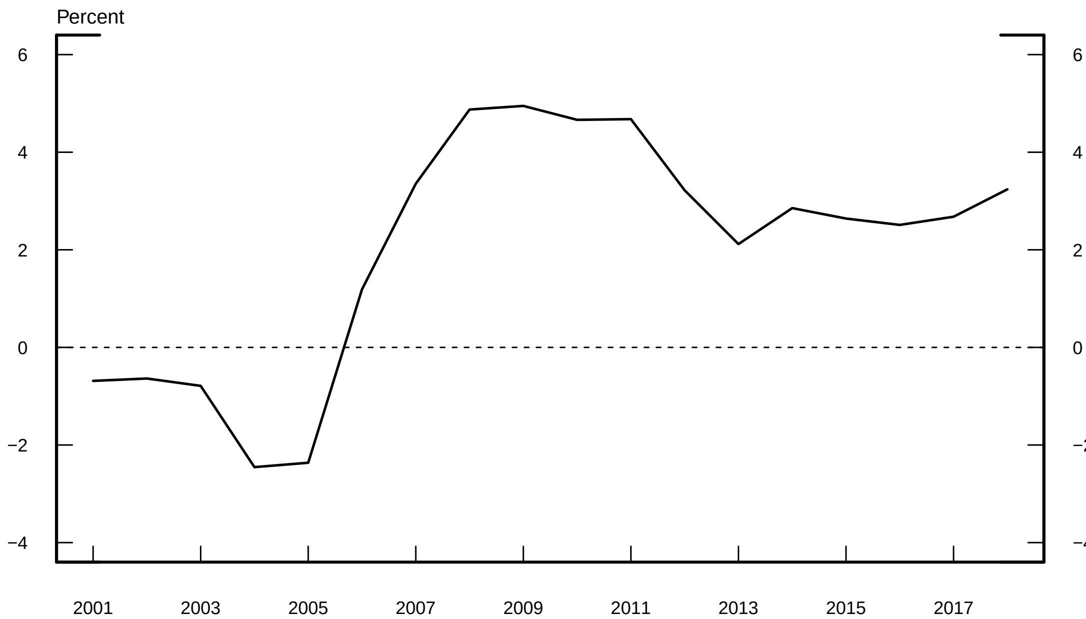
This issue here differs from the one affecting the representativeness of repeat-sales price indexes for valuing aggregate housing wealth. In that case, the issue is whether price changes for homes that transact are similar to prices for homes that do not transact. Here, the issue is whether model errors for homes that transact can be used to assess the accuracy of the model for homes that do not transact.

[22] The average value data that we obtain from Zillow are "raw" estimates, in that they are not adjusted for any model bias. The value estimates that Zillow publishes on its website are adjusted for model bias.

[23] However, the AVM accuracy tends to be lower in the tails of the price distribution. Especially for very inexpensive homes, the AVM is more likely to have a positive bias (i.e. the AVM estimate is higher than the actual sale price), while for very expensive homes, the reverse tends to be true.

[24] Consistent with Figure 2, this is because larger errors tend to be concentrated in smaller, less-densely populated

Figure 2: Value-Weighted Average County-level AVM Errors (U. S. Average)



Source: Zillow and Federal Reserve calculations.

5.2 Bias from Missing AVM Estimates

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. One reason we rely on ACS property counts as a benchmark is to overcome this challenge. But by applying average AVM values to ACS property counts, we are implicitly assuming that the average value of properties with and without an AVM are the same, conditional on county and property type. As such, a key consideration is whether properties with and without an AVM estimate are sufficiently homogeneous in characteristics to be similarly valued on average. We suspect that this assumption is relatively benign for the purpose of calculating aggregate housing wealth, since Zillow's coverage of own-use housing units tends to be fairly high.

To investigate this issue further, we examine property-level data from the 2014 ACS that were merged by address with a property-specific AVM estimate from another data provider.[25]

counties with fewer properties. Moreover, the distribution of average value-weighted errors across counties at any given period is roughly symmetric, so that the counties with outsized positive biases are roughly counterbalanced by counties with outsized negative biases.

[25] 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.

Although this merged dataset does not allow us to directly evaluate Zillow's AVM, it does allow us to compare the owner-reported values of homes that have estimates from this alternative AVM to owner-reported values for homes that do not have estimates from the same AVM. To the extent that the alternative AVM is representative of Zillow's AVM, this exercise can help us understand the value differences between the properties with a Zillow AVM estimate and those without. To this end, for each property type (single- or multifamily), we regress the natural logarithm of the owner-reported value on an indicator for whether the AVM estimate is missing, controlling for geographic location using fixed effects.

As shown in Table 2, the coefficient on the missing AVM indicator for single-family homes is about 1.2 percent, meaning that, on average, homes that do not have an AVM estimate are valued by their owners about 1.2 percent higher than homes with an AVM in the same county – a relatively small difference. However, when the regression excludes location fixed effects, we find that single-family homes with a missing AVM have a 19 percent lower average owner-reported value than homes that have an AVM estimate, and in a regression with state-level (rather than county-level) fixed effects the homes with a missing AVM are valued about 10 percent less. These results illustrate the non-random geographic distribution of the missing AVM values and thus the importance of aggregating property-level AVM estimates at granular (county) level for single-family homes.

For multifamily housing units (i.e., co-ops and condos), properties without an AVM are estimated to be 6.8 percent lower in value than those with an AVM. However, unlike with single-family homes, the within-county value difference is actually larger than (and opposite in sign to) the difference unconditional on geography. These results suggest that our county-up aggregation might overstate the aggregate value by a small amount for multifamily homes. However, because multifamily homes are such a small share of the aggregate housing stock – in the ACS, they account for only 7 percent of all own-use units in 2017 (Table 1) – the degree of overstatement for the aggregate housing stock will be very small.[26]

[26]Conservatively assuming that Zillow's coverage of condo and co-op units is 80 percent, the 7 percent difference between valued and unvalued properties would translate to an overall bias for multifamily of about 1.4 percent. Since multifamily units account only about 7 percent of all own-use units in the U. S., the total contribution of this bias to the aggregate wealth level would be negligible at 1.4 percent (0.2*0.07=0.014).

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

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

Note: 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.

5.3 Bias from the Inclusion of Rental Units in Zillow's Universe

A key reason we rely on ACS property counts is that we would like to focus on own-use units, and, as indicated in Table 1, Zillow's data includes some rental units. By applying the average Zillow values to the ACS property counts, we are implicitly assuming that the Zillow averages are not biased by the inclusion of rental properties. Ideally, we would remove the rental units from the Zillow data, but 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.[27] Consequently, the Zillow AVM averages will include the values of some rental units, and will not accurately 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, thereby boosting the average AVM estimate.

We evaluate this issue using the same merged ACS/AVM property-level data from 2014 that was described above. In this case, we regress the natural logarithm of the AVM for each property on an indicator for whether the ACS identifies the property as a rental unit, conditional on geographic location. Table 2 shows that on average, single-family rental units are valued 34 percent lower than owner-occupied units within the same county (column 1). For multifamily units, this differential is a little smaller, with rental units valued about 17 percent lower than owner-occupied multifamily

[27] One 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. While this is currently beyond our capability, such inference may become feasible in the future.

units within the same county.[28] The average value differences conditional on state (column 2) or unconditional on geography (column 3) are modestly larger for both single family and multifamily homes.

Table 3: Differences in AVM Value by Rental Status

Rental Indicator	(1) County FE	(2) State FE	(3) No Controls
Single-Family	-0.34***	-0.39***	-0.37***
(N=1,417,726)	(0.0013)	(0.0016)	(0.0018)
Multi-Family	-0.17***	-0.21***	-0.25***
(N=65,998)	(0.0046)	(0.0052)	(0.0064)

Note: 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.

The large difference of the non-Zillow AVM between rental units and owner-occupied suggests that Zillow's inclusion of some rental properties could pull down the average AVM value that we apply to the ACS property counts. 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 reported in Table 2 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 the values for the unintended inclusion of rental units would increase our estimate of aggregate housing wealth in 2014 by about 6 percent. It seems quite plausible that the value differential between owner-occupied and rental units could have changed materially over time, but we only have the ACS/AVM merged data for 2014. Because we want to evaluate the fluctuations in the AVM-based measure of housing wealth over the housing cycle and we have no idea how this bias may have evolved, and because the result comes from a non-Zillow AVM, we do not adjust our time series of aggregate wealth based on the estimated bias for 2014. We leave this issue for future study as methods for identifying and separating rental properties in deeds records may become available in the future.

[28] 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.

6 AVM and ACS Coverage Issues

6.1 Availability of Average AVM values by County

Due to data limitations, Zillow does not provide average AVM values for all counties, property types, and time periods. For the missing market segments, we impute the average value as the average value in the state for that property type. A potential problem with this approach is that these counties are more likely to have thin housing markets (e.g., in rural areas), and they could have a lower average value than other counties in the same state, as suggested by Table 2. However, the bias imparted by this assumption on aggregate housing wealth is likely quite small, as the counties without a Zillow AVM average cover only a small fraction of housing units – less than one percent of all ACS housing units in 2017.[29]

6.2 Availability of ACS Counts by County

We compute the number of own-use housing units by county and property type using the public-use microdata from the ACS. We use these data rather than published counts because it is the only way to calculate the number of vacant properties that are intended for own-use. The drawback of this approach is that the ACS only identifies 480 counties in the public-use data, covering about 60 percent of all occupied units.

To calculate the aggregate housing wealth in the counties that are not observed in the ACS, we need to know the number of own-use units in these counties and the average value. We estimate the number of housing units as the difference between the number of own-use units in the state and the sum of the county-level unit counts in that state.[30] Thus, one can think of our approach as aggregating all of the unobserved counties into a single "rest of the state" market. We impute the average values for these residual markets using the average AVM of counties that are included in these residual markets. Since the average AVM values are missing for so few counties, we think this approach should yield a fairly close approximation to the average value in the "rest of state" segment. A modified but closely related version of our aggregation method that employs

[29]We explored the use of alternate assumptions that estimate the average value for "missing" counties based 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.

[30]States are identified for all housing units in the public use data.

county-level property counts from published ACS tables and accounts for about 80 percent of all housing units yields estimates that are comparable to our baseline method.[31]

Two additional issues related to the ACS coverage bear mentioning. First, the full implementation of survey did not begin until 2005. To estimate housing unit counts for 2001 to 2004, we estimate the number of housing units by market segment in the 2000 Census and assume a constant growth rate between 2000 and 2005. Second, in order to obtain timely estimates of aggregate housing wealth, we must estimate the number of housing units after 2017 (the last available year of the ACS). Since housing counts are generally very slow moving, we assume that growth in 2018 was equal to the average growth rate from 2015 to 2017.

7 New Estimates of Aggregate Own-Use Housing Wealth

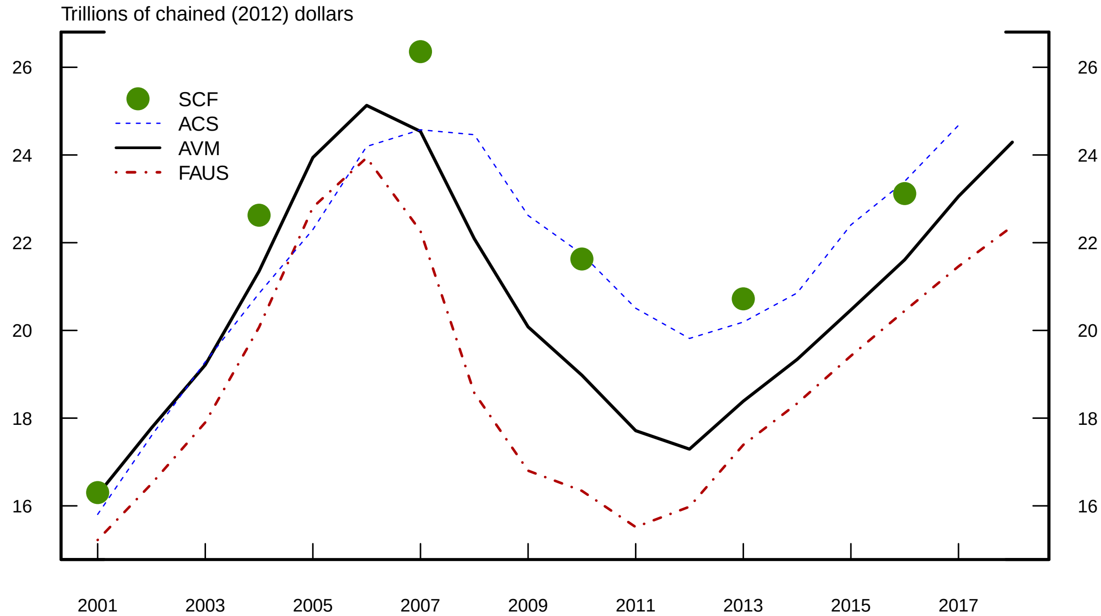
Figure 3 offers a fresh look at the evolution of aggregate own-use housing wealth since 2001 by plotting our new measure constructed from Zillow's AVM along with measures from the *Financial Accounts* and from owner-reported valuations in surveys (i.e., ACS and SCF). The *Financial Accounts* is a useful point of comparison because it is often used in macroeconomic models and moves largely with a house-price index after 2005.[32] From 2001 to 2005, the *Financial Accounts* is benchmarked to a weighted sum of owner valuations reported in the American Housing Survey. The ACS provides our baseline comparison with owner-reported valuations because it is available annually. To create a measure of aggregate house value from the ACS, we multiply average owner-reported values from the ACS by county and property type by the total own-use housing unit counts that were used to construct the AVM-based wealth estimate.[33] Thus, the ACS measure accounts for the aggregate value of vacant own-use housing wealth for which owner-reported value estimates are not directly available. For completeness, the SCF triennial estimates provide another measure of

[31] The published ACS tables cover more counties than the public-use microdata, accounting for about 80 percent of housing units. However, the published tables are not detailed enough to allow us to estimate the number of own-use vacant units by county. In an alternative approach, we used the number of owner-occupied units by county from the tables and estimated the number of own-use vacant units in each county from the state-wide ratio of own-use vacant units to owner-occupied units. This alternative estimate is quite similar to our baseline estimate from 2001 to 2004, and is about 3 percent higher than the baseline from 2005 to 2017.

[32] For observations after 2005, the aggregate level is extrapolated using the CoreLogic repeat-sales price index (to estimate capital gains on existing properties) and an estimate of net investment in residential structures (based on estimates of the value of new construction, renovation, and depreciation) from the Bureau of Economic Analysis.

[33] In accordance with our discussion in Section 6, we calculate the average ACS estimates from the property level, public use data, which only provide a county identifier for larger 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.

Figure 3: Alternative Measures of Aggregate Own-Use Housing Wealth



Source: American Community Survey (U.S. Census Bureau), Financial Accounts of the United States, Survey of Consumer Finance (triennial), and Zillow.
Note: Nominal values have been adjusted for inflation using the chain price index for personal consumption expenditures published by the Bureau of Economic Analysis.

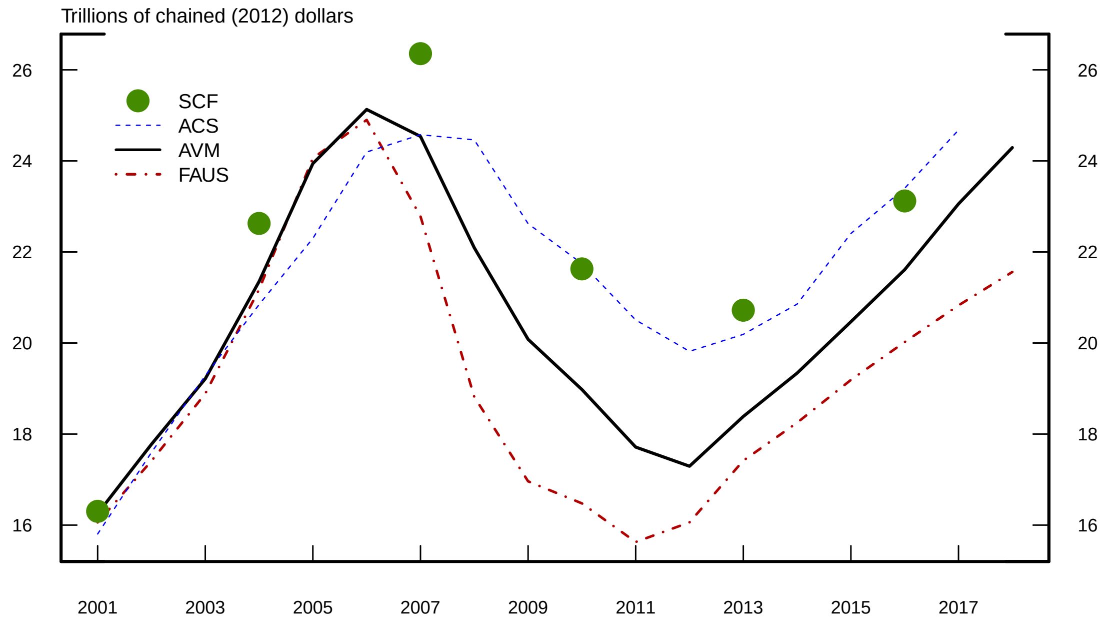
value-weighted aggregate wealth constructed from owner-reported house values.[34] All the wealth series are reported in 2012-constant dollars.

Figure 3 shows that the AVM, owner-reported measures from the ACS and SCF, and *Financial Accounts* measures track each other quite closely from 2001 to 2006. The AVM and ACS measures are nearly on top of one another from 2001 to 2004, with the AVM measure rising slightly faster than the ACS from 2004 to 2006. Although the *Financial Accounts* measure lies a little below the other two, this difference owes to the fact that the AHS estimates in the *Financial Accounts* between 2001 and 2005 are adjusted downward by 5.5 percent to reflect the upward bias in owner valuations reported in Goodman and Ittner (1992) and Kiel and Zabel (1999). Figure 4 shows that removing this adjustment causes the *Financial Accounts* to be almost exactly equal to the AVM-based measure from 2001 to 2006.[35] The alignment of our new AVM-based measure with the survey-based measures from ACS, AHS, and SCF in the 2001-2005 period is consistent with the

[34] The SCF estimates include the value of second and third homes that are not used as rental property.

[35] Removing this adjustment makes the *Financial Accounts* more comparable to the ACS, which is not adjusted for bias in owner valuations. Moreover, the 5.5% adjustment is probably too large because it is based on a simple average of owner valuation errors, and Molloy and Nielsen (2018) find that errors in owner valuations tend to be smaller for higher value properties, causing the value-weighted average to be about half as large as the simple average.

Figure 4: Alternative Measures of Aggregate Own-Use Housing Wealth, No Optimism Adjustment



Source: American Community Survey (U.S. Census Bureau), Financial Accounts of the United States, Survey of Consumer Finance (triennial), and Zillow.
Note: Nominal values have been adjusted for inflation using the chain price index for personal consumption expenditures published by the Bureau of Economic Analysis.

idea owner-reported values are quite close to market-based valuations during a period of rising house prices, as found by Molloy and Nielsen (2018).

After 2005, the three measures diverge substantially. First, the measures differ on the timing of the market turning points. Most notably, the AVM and the *Financial Accounts* (which are both informed by market prices over this period) estimate a clear peak in 2006, while the ACS measure starts to decline noticeably only much later in 2009.[36] The timing of the trough also differs across these measures, with the *Financial Accounts* turning up in 2011 and the AVM and ACS turning up in 2012.

Second, the three measures disagree on the severity of the housing cycle. Between 2006 and 2011, the *Financial Accounts* measure drops by 35 percent, whereas the ACS measure declines by only 15 percent. Our new AVM measure is more similar to the *Financial Accounts* over this period, falling 30 percent from peak to trough. Thus, each measure provides a different assessment of the amount of housing wealth lost by households during the crisis. Our new AVM measure indicates that households lost about $7.4 trillion over this period, while the *Financial Accounts* suggests losses

[36] The differences in the timing of the peak between owner-reported and market-based measures is the main reason why the *Financial Accounts* methodology switches from a survey-based measure to a house price index in 2005 .

of \$8.4 trillion and the ACS suggests only \$3.7 trillion. The growth rates of these three measures were fairly similar during the recovery. On net, whereas the AVM and *Financial Accounts* measures were still 8 percent and 10 percent below their 2006 values in 2017, the ACS measure was 2 percent *above* its 2006 value.

The differences between the AVM-based measure and the ACS-based measure are consistent with many of the concerns about owner valuations that have been raised in the literature.[37] Specifically, our results suggest that survey respondents were either unaware of the market fluctuations in real time, or they believed – correctly or not – that their home values were different from what sales in the surrounding market would imply. In particular, the ACS measure indicates that survey respondents thought that their homes were maintaining their values even as the housing market and the financial system were experiencing severe strain. To the extent that ACS respondents did eventually acknowledge a declining market, it appears that they did not take the decline fully on board.

The differences between the AVM and the *Financial Accounts* post-2006 are also consistent with some of the concerns about repeat-sales indexes raised in prior studies.[38] While the AVM measure and the *Financial Accounts* show much more responsiveness to changing market conditions than surveys, the AVM measure shows somewhat less cyclicality than the *Financial Accounts* measure. Since the *Financial Accounts* measure is constructed using a repeat-sales price index during and after the Great Recession, its movements are based on the price changes experienced by transacting homes. If non-transacting homes experienced different price dynamics, as might be the case during periods of market turmoil when "motivated" homeowners experiencing financial strain make up an elevated share of transactions, then applying repeat-sales price index movements to non-transacting homes could overstate fluctuations in the aggregate value.

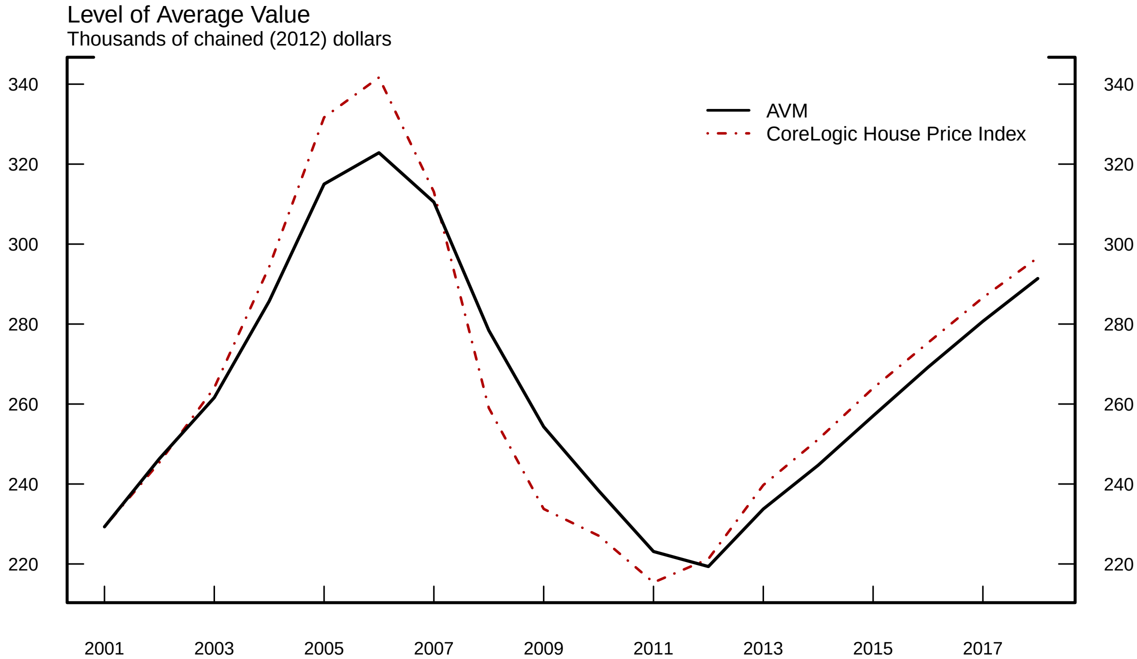
To shed more light on this issue, the top and bottom panels of Figure 5 compare the levels and annual growth rates of the AVM average home values to the CoreLogic repeat-sales price index.[39] The index rises more than the average AVM values during the housing boom and falls more during

[37] Housing wealth estimates from the SCF estimates (not shown) lie fairly close to those from the ACS, illustrating that the movements in the ACS are representative of owner valuations more generally.

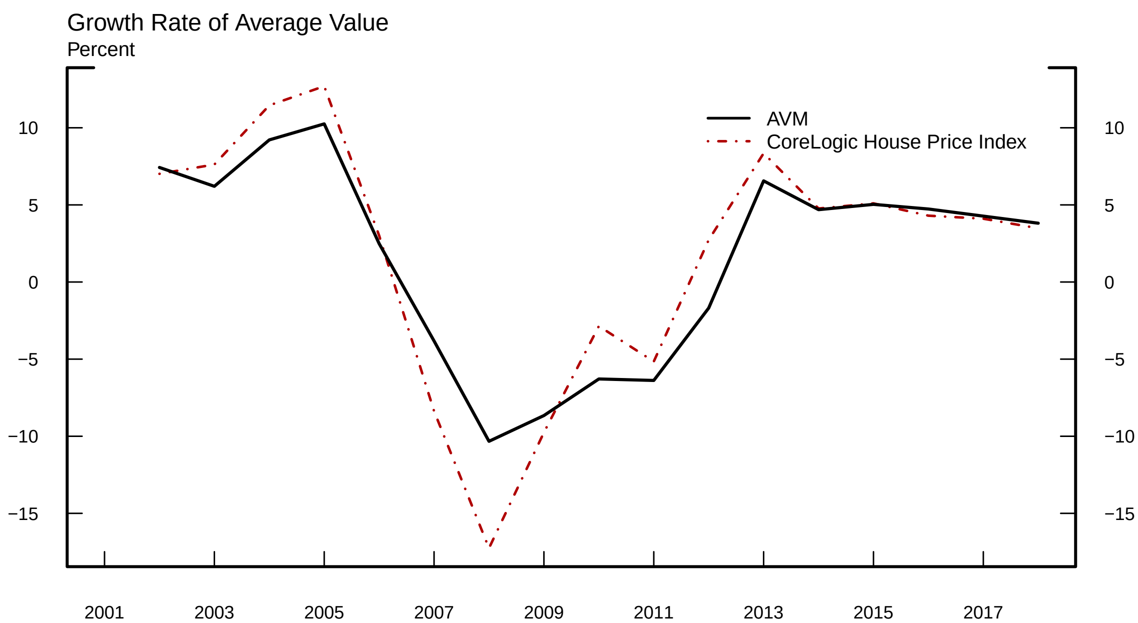
[38] The net investment component of the *Financial Accounts* measure increases over time but does not contribute much cyclicality to aggregate wealth.

[39] Source: CoreLogic, Inc., Private-Label Loan, Home Equity Servicing, and HPI data. This index corresponds to series FI075035243 in the *Financial Accounts*.

Figure 5: Level and Growth Rate of Average Value of Own-Use Housing Units



Source: CoreLogic and Zillow.
Note: Nominal values have been adjusted for inflation using the chain price index for personal consumption expenditures published by the Bureau of Economic Analysis.



Source: CoreLogic and Zillow.
Note: Nominal values have been adjusted for inflation using the chain price index for personal consumption expenditures published by the Bureau of Economic Analysis.

the contraction, suggesting that that the selection biases in the repeat-sales index might indeed be present.

Finally, Figures 3 and 4 plot annual data in order to allow us to make consistent comparisons across all three data series, as the ACS provides only annual averages. Two additional advantages of the AVM-based measure over owner-reports are that they can be computed at a higher frequency and they are more timely. We are not aware of any surveys that provide owner-based valuations at a higher frequency than annually, likely because surveys are quite costly and time-intensive to conduct. By contrast, we can compute quarterly or monthly estimates of the AVM-based measure quite easily because the AVM is re-estimated and updated with new data every day.[40]

8 Conclusion

In this paper, we develop a detailed, high-frequency, and timely measure of aggregate housing wealth using county-level average home values from Zillow's AVM, which is based on machine-learning techniques and detailed information on recent transaction prices and property characteristics. To create our measure, we combine the average home values from Zillow's AVM with county-level property counts from Census, while adjusting for the average county-level bias of the particular AVM that we use. Using this method, we present new estimates of aggregate U. S. housing wealth from 2001 to 2018. Our work demonstrates how data that are collected from a private source for an entirely different purpose can be used to create a nationally representative aggregate time series when combined with other data. Policymakers around the world have been increasingly interested in the use of such data in order to improve aggregate statistics and reduce the cost of production. While our application provides an example of how this can work, it is worth bearing in mind that in almost every application that uses private-firm data to construct aggregate statistics there will be a role for surveys or other data sources to provide a way to aggregate appropriately.

Our AVM-based estimates show considerably more responsiveness to changing market conditions than survey-based measures and somewhat less volatility than repeat-sales measures. The finding that owner valuations (as reported in surveys) appear to underestimate the amplitude of the cycle could affect the findings of studies that use owner valuations, such as those that investigate

[40] Although we only have annual housing stock estimates from the ACS, we linearly interpolate these annual estimates. We believe this interpolation to be fairly accurate because the quantity of housing changes fairly slowly over time.

changes in the distribution of household wealth over the Great Recession (see, for example, Bricker et al. (2011); Hur (2018), or Peterman and Sommer (ming)). Moreover, the finding that the *Financial Accounts* series appears to amplify the recent housing boom and bust is also worth bearing in mind, as various macroeconomic frameworks use this series (or rely on repeat-sales indexes in other ways) to estimate the marginal propensity to consume (e.g., Carroll et al. 2011) and other parameters (Iacoviello and Neri, 2010; Saez and Zucman, 2016; Favilukis et al., 2017; Glover et al., 2019). However, our results also suggest that the bias from repeat-sales indexes is not as large as the bias in owner-valuations, and so during housing busts and recoveries the use of house price indexes to extrapolate housing wealth might not be a bad approximation.

One simple extension of our analysis would be to calculate a measure of aggregate household housing wealth that includes rental property. Such a measure could prove more useful than own-use housing wealth for understanding household consumption decisions. Additionally, since our measure is constructed from county-level estimates, it can naturally be decomposed across states, metropolitan areas, or counties. This type of analysis would add to the growing literature on variation in income and other economic outcomes across locations (Mian and Sufi, 2014; Chetty et al., 2014; Chetty and Hendren, 2018a,b).

Finally, while our work has focused on aggregating AVM estimates to create an aggregate time series, further research should consider merging property-level AVM estimates with nationally-representative survey data. This merged data might improve our ability to assess the aggregate distribution of housing wealth, a topic that has received much attention of late (Carroll et al., 2014; Piketty et al., 2018; Batty et al., 2019). Merged AVM/survey data could also be used to study the connection between housing wealth and other economic outcomes at the household level. We view such analysis as an important avenue for future work.

References

Aladangady, A., Aron-Dine, S., Dunn, W., Feiveson, L., Lengermann, P., and Sahm, C. (2019). From Transactions Data to Economic Statistics: Constructing Real-Time, High-Frequency, Geographic Measures of Consumer Spending. In *Big Data for 21st Century Economic Statistics*. National Bureau of Economic Research, Inc.

Athey, S. (2018). The Impact of Machine Learning on Economics. In *The Economics of Artificial Intelligence: An Agenda*, pages 507–547. University of Chicago Press.

Athey, S. and Luca, M. (2019). Economists (and Economics) in Tech Companies. *Journal of Economic Perspectives*, 33(1):209–30.

Batch, A., Chen, J. C., Dunn, A., Driessen, A., and Hood, K. K. (2019). Off to the Races: A Comparison of Machine Learning and Alternative Data for Predicting Economic Indicators. In *Big Data for 21st Century Economic Statistics*, NBER Chapters. National Bureau of Economic Research, Inc.

Batty, M. M., Bricker, J., Briggs, J. S., Holmquist, E. B., McIntosh, S. H., Moore, K. B., Nielsen, E., Reber, S., Shatto, M., Sommer, K., Sweeney, T., and Henriques, A. (2019). Introducing the Distributional Financial Accounts of the United States. Finance and Economics Discussion Series 2019-017, Board of Governors of the Federal Reserve System (US).

Benitez-Silva, H., Eren, S., Heiland, F., and Jimenez-Martin, S. (2015). How Well Do Individuals Predict the Selling Prices of Their Homes? *Journal of Housing Economics*, 29(C):12–25.

Bricker, J., Bucks, B. K., Kennickell, A. B., Mach, T. L., and Moore, K. B. (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 (US).

Bucks, B. K. and Pence, K. M. (2006). Do Homeowners Know Their House Values and Mortgage Terms? Technical report.

Cajner, T., Crane, L. D., Decker, R. A., Hamins-Puertolas, A., and Kurz, C. (2019). Improving the Accuracy of Economic Measurement with Multiple Data Sources: The Case of Payroll Employment Data. In *Big Data for 21st Century Economic Statistics*, NBER Chapters. National Bureau of Economic Research, Inc.

Carroll, C. D., Otsuka, M., and Slacalek, J. (2011). How Large Are Housing and Financial Wealth Effects? A New Approach. *Journal of Money, Credit and Banking*, 43(1):55–79.

Carroll, C. D., Slacalek, J., and Tokuoka, K. (2014). The Distribution of Wealth and the MPC: Implications of New European Data. *The American Economic Review*, 104(5):107–111.

Case, B., Pollakowski, H. O., and Wachter, S. M. (1997). Frequency of Transaction and House Price Modeling. *The Journal of Real Estate Finance and Economics*, 14(1):173–187.

Chan, S., Dastrup, S., and Ellen, I. (2016). So Homeowners Mark to Market? A Comparison of Self-Reported and Estimated Market Home Values During the Housing Boom and Bust. *Real Estate Economics*, 44(3):627–657.

Chetty, R. and Hendren, N. (2018a). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *Quarterly Journal of Economics*, 113(3).

Chetty, R. and Hendren, N. (2018b). The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Davis, M. A. and Quintin, E. (2017). On the Nature of Self-Assessed House Prices. *Real Estate Economics*, 45:628–649.

Dreiman, M. H. and Pennington-Cross, A. (2004). Alternative Methods of Increasing the Precision of Weighted Repeat Sales House Prices Indices. *The Journal of Real Estate Finance and Economics*, 28(4):299–317.

Favilukis, J., Ludvigson, S. C., and Van Nieuwerburgh, S. (2017). The Macroeconomic Effects of Housing Wealth, Housing Finance, and Limited Risk Sharing in General Equilibrium. *Journal of Political Economy*, 125(1):140 – 223.

Gatzlaff, D. H. and Haurin, D. R. (1997). Sample Selection Bias and Repeat-Sales Index Estimates. *The Journal of Real Estate Finance and Economics*, 14(1):33–50.

Gindelsky, M., Moulton, J., and Wentland, S. A. (2019). Valuing Housing Services in the Era of Big Data: A User Cost Approach Leveraging Zillow Microdata. In *Big Data for 21st Century Economic Statistics*, NBER Chapters. National Bureau of Economic Research, Inc.

Glaeser, E. L., Kim, H., and Luca, M. (2019). Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity. In *Big Data for 21st Century Economic Statistics*, NBER Chapters. National Bureau of Economic Research, Inc.

Glennon, D., Kiefer, H., and Mayock, T. (2018). Measurement Error in Residential Property Valuation: An Application of Forecast Combination. *Journal of Housing Economics*, 41(C):1–29.

Glover, A., Heathcote, J., Krueger, D., and Rios-Rull, J.-V. (2019). Intergenerational Redistribution in the Great Recession. Working Paper 16924, University of Pennsylvania.

Goodman, J. L. and Ittner, J. B. (1992). The Accuracy of Home Owners' Estimates of House Value. *Journal of Housing Economics*, 2(4):339 – 357.

Henriques, A. M. (2013). Are Homeowners in Denial About Their House Values? Comparing Owner Perceptions with Transaction-Based Indexes. Finance and Economics Discussion Series 2013-79, Board of Governors of the Federal Reserve System (US).

Hur, S. (2018). The Lost Generation of the Great Recession. *Review of Economic Dynamics*, 30:179–202.

Iacoviello, M. and Neri, S. (2010). Housing Market Spillovers: Evidence from an Estimated DSGE Model. *American Economic Journal: Macroeconomics*, 2(2):125–64.

Ihlanfeldt, K. and Martinez-Vazquez, J. (1986). Alternative Value Estimates of Owner-Occupied Housing: Evidence on Sample Selection Bias and Systematic Errors. *Journal of Urban Economics*, 20(3):356–369.

Kiel, K. A. and Zabel, J. E. (1999). The Accuracy of Owner-Provided House Values: The 1978-1991 American Housing Survey. *Real Estate Economics*, 27(2):263–298.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z. (2015). Prediction Policy Problems. *American Economic Review*, 105(5):491–95.

Korteweg, A. and Sorensen, M. (2016). Estimating Loan-to-Value Distributions. *Real Estate Economics*, 44(1):41–86.

Kuzmenko, T. and Timmins, C. (2011). Persistence in Housing Wealth Perceptions: Evidence from the Census Data.

Malone, T. and Redfearn, C. L. (2009). To measure globally, aggregate locally: Urban land, submarkets, and biased values of the housing stock. mimeo.

Mian, A. and Sufi, A. (2014). What Explains the 2007-2009 Drop in Employment? *Econometrica*, 82(6):2197–2223.

Molloy, R. and Nielsen, E. (2018). How Can We Measure the Value of a Home? Comparing Model-Based Estimates with Owner-Occupant Estimates. *FEDS Notes*.

Mullainathan, S. and Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2):87–106.

Peterman, W. and Sommer, K. (forthcoming). How Well Did Social Security Mitigate the Effects of the Great Recession? *International Economic Review*.

Piketty, T., Saez, E., and Zucman, G. (2018). Distributional National Accounts: Methods and Estimates for the United States. *The Quarterly Journal of Economics*, 133(2):553–609.

Saez, E. and Zucman, G. (2016). Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. *The Quarterly Journal of Economics*, 131(2):519–578.

van der Cruijsen, C., Jansen, D.-J., and van Rooij, M. (2018). The Rose-Tinted Spectacles of Homeowners. *Journal of Consumer Affairs*, 52(1):61–87.

9 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.

9.1 Model Bias Adjustment

As discussed in the text, Zillow's AVM estimates are biased, in the sense that they have a time-varying non-zero average error rate when compared to subsequent transactions. This is particularly true for properties in the bottom 30 percent 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.[41] Zillow reports to us both the median-adjusted and "raw" average value estimates. Since our goal is to calculate an aggregate, we are concerned with average bias, rather than median, and we use the raw Zillow numbers to 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 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. We therefore construct our value-weighted adjustments using the following procedure:

1. Let $V^u(p,c,t,i)$ be the value defining the upper limit of decile i for property type p, county c, and year t.[42] We define the value share of decile i as $w(p,c,t,i) = \frac{(V^u(p,c,t,i)+V^u(p,c,t,i-1))/2}{\sum_j (V^u(p,c,t,j)+V^u(p,c,t,j-1))/2}$, where $V^u(p,c,t,0)$ is set to 0 and $V^u(p,c,t,10)$ is set equal to $1.5V^u(p,c,t,9)$. The upper bound on the value distribution (50 percent above the 90$^{\text{th}}$ percentile) is an arbitrary limit intended 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; e.g., we obtain quantitatively similar results using either $1.2V^u(p,c,t,9)$ or $2.0V^u(p,c,t,9)$ as the upper bound.

2. We estimate the value-weighted average error as $E(p,c,t) = (\sum_i w(p,c,t,i)APE(p,c,t,i))$, where $APE(p,c,t,i)$ 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 $E(p,c,t)$ to missing if the number of transactions in (p,c,t) is less than 20. We also drop (p,c,t) observations if one or more of the decile average errors is missing. For counties not covered by Zillow's model or set to missing due to one of the above two exclusion criteria, we use the average value-weighted error for counties in the same state

[41] 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. Since our focus is on constructing an aggregate measure, 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.

[42] 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.

which do have error distribution data. Let $\tilde{E}(p,c,t)$ denote the average errors including these imputations.

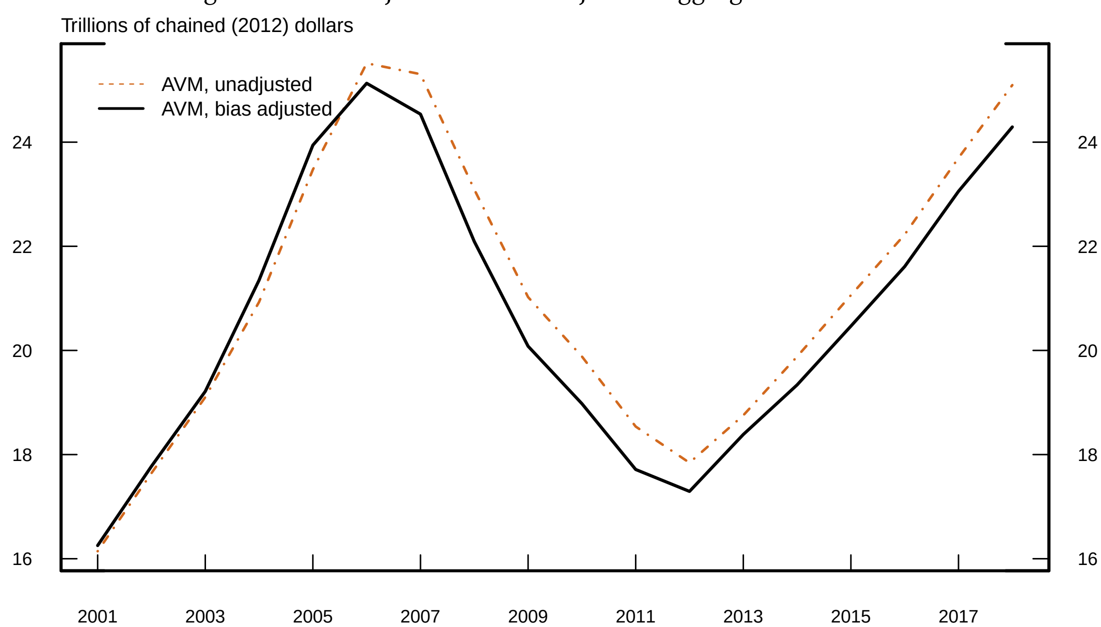
3. We define the adjustment factor $\gamma(p,c,t) = \frac{\tilde{E}(p,c,t)}{APE(p,c,t)}$, where $APE(p,c,t)$ is the unadjusted average error regardless of the number of transactions (i.e., the floor of 20 is dropped). If the county is not covered by Zillow, we set $APE(p,c,t)$ equal to the statewide average error prior to computing $\gamma(p,c,t)$. Note that $(\gamma(p,c,t)APE(p,c,t))$ is the estimated value-weighted average error for (p,c,t). The only reason to go from $\tilde{E}(p,c,t)$ to $\gamma(p,c,t)$ and back to $(\gamma(p,c,t)APE(p,c,t))$ instead of using $\tilde{E}(p,c,t)$ directly is to make use of the unweighted average errors of counties with fewer than 20 observations.

4. We define the combined (sf and mf) value weighted error $\hat{E}(c,t)$ as the weighted sum of $(\gamma(p,c,t)APE(p,c,t))$ for single-family and multifamily properties, where the weights are each property type's share of the total observed transactions in county c and year t.

5. We adjust the average errors by $(1+\hat{E}(c,t))^{-1}$. We combine the multifamily and single-family errors into one county-level series for two reasons. First, the share of counties with sufficiently many multifamily transactions to accurately estimate value-weighted errors is quite small. Second, the value-weighted errors do not appear to be very different for multifamily and single-family properties.

Figure 6 below shows that the bias adjustment changes the aggregate series very little from 2001-2005. After 2005, the effect of the bias adjustment is to slightly lower the aggregate series.

Figure 6: Bias-Adjusted and Unadjusted Aggregate Wealth Series



Source: American Community Survey (U.S. Census Bureau) and Zillow.
Note: Nominal values have been adjusted for inflation using the chain price index for personal consumption expenditures published by the Bureau of Economic Analysis.

9.2 Rental Bias

As noted in the main text, since Zillow's data do not distinguish between rental properties and own-use properties, and we are interested in own-use properties, large differences in average value by ownership status will tend to bias our results. As discussed above, it appears that rental homes are on average substantially less valuable than owner-occupied properties. As a result, estimates of the aggregate value of own-use housing will be biased downwards by the inclusion of rental properties in Zillow's averages.

We implement the following procedure to estimate the effect of the inclusion of rental properties on our aggregate measure. 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 particular, our procedure consists of the following steps:

1. We match the 2014 ACS to the 2014 alternative AVM estimates at the property level.
2. For each geography g and property type p, we calculate the ratio of AVM values for rental properties to owner-occupied properties:
$$\delta(p,g) = \frac{\bar{V}(p,g,\text{rental})}{\bar{V}(p,g,\text{owner})}.$$
3. For each county c, property type p, and time period t, let $\beta(p,c,t)$ 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 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 percent of the ACS universe is non-condos (and therefore likely to not be in Zillow's deeds-based property records). This 82 percent estimate does not account for vacant units, however. The Census Housing Vacancy Survey reports a 9.3 percent 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 $\beta(p,c,t)$ and $\delta(p,g(c))$, the property type/county/time period adjustment factor is constructed as
$$\lambda(p,c,t) = \frac{1}{\beta(p,c,t) + (1-\beta(p,c,t))\delta(p,g(c))}.$$

   Using the $\lambda(p,c,t)$ to adjust the average values from Zillow results in an alternate aggregate estimate that is roughly 6 percent higher in 2014 than what we report in our baseline results.