

Zillow's Estimates of Single-Family Housing Values

by Daniel R. Hollas, PhD, Ronald C. Rutherford, PhD, and Thomas A. Thomson, PhD

ABSTRACT

This article compares Zillow.com's estimates of home values and the actual sale prices of 2045 single-family residential properties sold in Arlington, Texas, in 2006. Zillow indicates that this market is one where its data has its highest accuracy rating. However, the results show the median difference between the Zillow estimate of market value and the actual sale price is 7.92%, and the average overpricing by Zillow is 11.66%. The results also show that Zillow overestimated the value for 40% of the properties by more than 10%, and only 0.88% of the property prices were underestimated by 10% or more. The study suggests that homeowners' estimates of value may be more accurate than Zillow's estimates.

Over the past several years, home value estimates have been an issue of considerable importance, as evidenced by the numerous news accounts of first rising and then falling home prices. Homeowners, neighbors, tax assessors, appraisers, and real estate agents have been estimating values based on the available market information. In addition, Zillow.com, a real estate Web site launched in February 2006, includes estimated market values for houses.

Zillow provides an estimate of market value for over 46 million homes based on a proprietary formula.¹ In general, it offers free value estimates, or Zestimates, using data from appraisal districts and from multiple listing services (MLSs), depending on availability. Zillow currently also accepts for-sale listings, offers information about buying and selling, and provides links to mortgage providers and real estate professionals. Several of these latest information offerings have been added since the data for this study was collected.

The key issue regarding Zillow's Zestimates is whether they reflect transaction prices. Zillow has been described both as "a useful site" and as "categorically wrong." There have been many instances of praise and many instances of complaints by homeowners using the Web site to estimate the value of their homes. Realtors in general have also been critical of the values produced by Zillow. The objective of this research and article is to examine the relationship between Zillow's Zestimates and actual transaction prices, while also examining how the Zillow model compares to a standard hedonic model of sale prices.

An unanswered question is how Zillow's estimates of value compare to actual sale prices. In 2006, Mullaney stated Zillow had informed *BusinessWeek* that it would be able to obtain a value estimate within $\pm 10\%$ of actual value for 62%

1. For details about Zillow's estimation methods and models, see <http://www.zillow.com/howto/Zestimate.htm>. Zestimate® is a trademark of Zillow, Inc.

of homes.² In *The Wall Street Journal*, Hagerty discussed the findings of a study of 1,000 houses across seven states that examined how Zillow's estimates compare to transaction prices. That study found that Zillow's Zestimates are often close to the sale prices, but there is a large range of mispricing.³ Zillow.com is careful to state that the Zestimates are not appraisals and should be used as one estimate of market value. In fact, they encourage obtaining an appraisal from a professional appraiser. The study presented here examines the relationship between the estimates of value provided by Zillow and the sale prices for a sample of single-family houses in a market where Zillow.com indicates it has the best data and the highest accuracy level in the estimation of home value.

The results indicate that Zillow overestimates value for approximately 80% of the houses in the sample by at least 1%. Fifty-nine percent of the Zillow estimates fall within $\pm 10\%$ of the sale price and only 0.88% of values are underestimated by more than 10%. The average overestimation is 11.66% or \$13,576, with a median of \$9,717 or 7.92%. Zillow's magnitude of overestimation is marginally higher than the value overestimation by recent homebuyers reported in the literature.

Literature Review

In general, this study is interested in how accurate are estimates of housing value. Reliable estimates of value are needed for a number of reasons. For example, tax assessors, appraisers, and real estate agents require reasonably accurate estimates to perform their jobs. One estimate of value that has received ongoing interest in real estate literature is the estimate of value provided by homeowners. Homeowners have been asked to determine the value of their homes in previous studies and in the American Housing Survey conducted by the U.S. Census Bureau.

In their study, Kish and Lansing ask homeowners to estimate the market value of their houses and then ask professional appraisers to also value the houses.⁴ Their findings indicate that approximately 37% of the homeowners estimate a value within an interval of $\pm 10\%$ of the value estimated by the professional appraisers. Kain and Quigley also examine this question using the same methods in a different city and find similar results.⁵ Goodman and Ittner compare owners' estimates of value with subsequent sale prices; their results indicate that homeowners overestimate value by approximately 6%.⁶ Kiel and Zabel find that the average owner overestimates the value of his or her house by 5.1% and recent buyers overestimate value by 8.4% compared to the stated sale price.⁷ Follain and Malpezzi's study results suggest that homeowners underestimate their values by 2%.⁸ More recently, Agarwal looks at house value estimates of applicants for home equity loans and finds, on average, a 3.1% overestimate of value and an average pricing error of 13.9%.⁹ These studies are primarily concerned with whether an owner's estimate of value is reliable.

Kiel and Zabel note that if the owner-estimated values in the American Housing Survey (AHS) were accurate, then the AHS would provide reliable data for researchers to study a host of subjects. However, the evidence generally suggests that owners overestimate the values of their homes. Similarly, if Zillow's estimates of values are found to be relatively precise, owners could use the information to more accurately estimate their home values for use in their financial decision making. Based on the results from the previously mentioned *Wall Street Journal* article and the results reported in this article, however, there is reason to doubt that Zillow's market value estimates are any more reliable than owners' estimates of value.

2. Tim Mullaney, "Hot Property: Zeroing in on Zillow," *BusinessWeek*, April 13, 2006, http://www.businessweek.com/the_thread/hotproperty/archives/2006/04/zeroing_in_on_z.html.

3. James R. Hagerty, "How Good Are Zillow's Estimates?" *Wall Street Journal*, February 14, 2007, Eastern edition, sec. D.

4. Leslie Kish and John B. Lansing, "Response Errors in Estimating the Value of Homes," *Journal of the American Statistical Association* 49, no. 267 (1954): 520-538.

5. John F. Kain and John M. Quigley, "Note on Owner's Estimate of Housing Value," *Journal of the American Statistical Association* 67, no. 340 (1972): 803-806.

6. John L. Goodman, Jr., and John B. Ittner, "The Accuracy of Home Owners' Estimates of House Value," *Journal of Housing Economics* 2, no. 4 (December 1992): 339-357.

7. Katherine A. Kiel and Jeffrey E. Zabel, "The Accuracy of Owner-Provided House Values: The 1978-1991 American Housing Survey," *Real Estate Economics* 27, no. 2 (1999): 263-298.

8. James R. Follain and Stephen Malpezzi, "Are Occupants Accurate Appraisers?" *Review of Public Data Use* 9, no. 1 (1981): 47-55.

9. Sumit Agarwal, "The Impact of Homeowners' Housing Wealth Misestimation on Consumption and Saving Decisions," *Real Estate Economics* 35, no. 2 (2007): 135-154.

The following section of this article presents the study methods used. A discussion of the data is presented next and then the empirical results are presented. The final section presents the study conclusions.

Zillow's Zestimates Compared to Sale Prices

Methodology

The current study compares the percentage difference between the Zillow value estimate and the actual sale price of single-family homes. The sample in this study consists of recent sales in a market where Zillow.com states it has the highest accuracy.

The relationship between the sale price of a home (SP_i) and the Zillow Zestimate of value (ZV_i) is examined for the same house. A simple OLS model is used to examine the relationship between Zillow's estimate and sale price:

$$SP_i = \beta ZV_i + \varepsilon_i \quad (1)$$

If Zestimates fully reflect transaction prices, then it is expected that $E[\beta_i | ZV_i, \varepsilon_i] = 1$, and the intercept = 0. If Zillow's estimates are on balance higher than the sale price of the house, then the estimated slope coefficient, β , will be less than one and vice versa.

Examining Differences in Coefficients in Sale Price and Zillow's Value

The relationship between Zillow's valuation of housing characteristics and the market's valuation of the same housing characteristics is examined. The model estimated is

$$\ln SP_i = a_i + h_i\beta + l_i\delta + \varepsilon_i \quad (2)$$

where the dependent variable is the log of the sale price. The regressors include vectors of housing characteristics (h_i) and location (l_i) dummy variables.

$$\ln ZV_i = a_i + h_i b + l_i c + \varepsilon_i \quad (3)$$

In Equation 3, the dependent variable is the log of Zillow's Zestimate. The regressors include the same set of housing characteristics (h_i) and location (l_i)

dummy variables as the sale price model. The most common housing characteristics on the Zillow Web site are number of bedrooms, number of bathrooms, total square feet, lot size, year built, and number of stories. Although Zillow.com lists a number of other housing characteristics, most of the time the information for the additional characteristics is not provided on the Web site. The location of the property is shown along with the street address and zip code. Estimation of the two models shown in Equations 2 and 3 allows examination of whether the Zillow valuation of housing characteristics and the sales-based transaction valuation of housing characteristics are the same.

Data

For this study, data was obtained from the multiple listing service (MLS) and Zillow.com. The sale price data, housing characteristics, and location data are obtained from the MLS for the city of Arlington, Texas (Tarrant County), for sales during the last six months of 2006. The sale price data was acquired in January 2007. Next, data from Zillow.com was obtained for each of the MLS sales to acquire Zestimates during the last week of January 2007 and the first two weeks of February 2007.

The sample includes 2045 properties that sold and that have a Zestimate.¹⁰ The sample sales include both vacant and occupied single-family homes. The average sale price for the sample is \$138,064, with a \$151,640 average Zestimate. The mean difference is \$13,576. The median sale price is \$124,000, whereas the median Zestimate is \$134,714. The average housing characteristics for the single-family homes are 3.3 bedrooms, 2.2 bathrooms, 1,937 square feet, 1.2 stories, and 1.8 garage spaces. Zip code dummy variables were used to control for location. Table 1 shows a complete set of descriptive statistics. Table 2 shows the distribution of the percentage difference and dollar difference. The median for the percentage difference is 7.92%, and the median for the dollar difference is \$9,717.

Empirical Results

The results discussed are for a city in Tarrant County, a county which Zillow gives its highest accuracy

10. Zillow.com indicates that different markets have different rating in terms of Zestimate accuracy. Arlington is located in Tarrant County; Tarrant County has a rating of four stars, which equates with a "Best Zestimate." For all of Tarrant County, Zillow.com indicates that 99% of the homes are on its Web site, 99% have Zestimates, 60% are within 10% of the sale price, and the median error is 7.5%. For Tarrant County, the certified appraisal values are available August 25, 2006 in this market and all sales are available as soon as they are reported to the appraisal district. It is partly for this reason that Zillow classifies the accuracy of its Zestimates in Tarrant County as a four-star market, the highest level of accuracy. Tarrant County states on its Web site that sales are available from a given date until the date of the extraction of the sales file; data is available free on the Tarrant Appraisal District's Web site at http://www.tad.org/WebPages/free_downloads.cfm and includes sales that have occurred on accounts between January 1, 2003 and the date of extraction. This data file is only for sales obtained by TAD that may be disclosed.

Table 1 Descriptive Statistics for 2045 Sales in Arlington, Texas

Variable	Mean	Standard Deviation	Minimum	Maximum
Sale Price (SP)	\$138,064	\$75,103	\$34,000	\$1,000,000
Zestimate	\$151,640	\$82,643	\$47,974	\$1,285,166
Percent Difference (Zestimate – SP)/SP*100	11.66	15.34	-53.04	92.04
Difference (Zestimate – SP)	\$13,576	\$23,075	\$-126,701	\$470,166
Square Feet	1,937	711	551	8,220
Age	22.15	13.61	1.0	78.0
Number of Bedrooms	3.33	0.60	1.0	6.0
Number of Bathrooms	2.18	0.57	1.0	5.0
Pool (yes = 1)	0.15	0.36	0.0	1.0
Vacant (yes = 1)	0.37	0.48	0.0	1.0
Garage Spaces	1.84	0.57	0.0	3.0
Number of Stories	1.24	0.43	1.0	2.0
Zip_Code1	0.126	0.332	0.0	1.0
Zip_Code2	0.111	0.314	0.0	1.0
Zip_Code3	0.035	0.184	0.0	1.0
Zip_Code4	0.071	0.257	0.0	1.0
Zip_Code5	0.021	0.144	0.0	1.0
Zip_Code6	0.065	0.247	0.0	1.0
Zip_Code7	0.057	0.231	0.0	1.0
Zip_Code8	0.075	0.264	0.0	1.0
Zip_Code9	0.029	0.169	0.0	1.0
Zip_Code10	0.126	0.332	0.0	1.0
Zip_Code11	0.176	0.381	0.0	1.0
Zip_Code12	0.108	0.311	0.0	1.0

Table 2 Distribution for Percentage Difference and Actual Difference between Zillow's Zestimate and Sale Price

Percentile	(Zestimate – SP)/SP *100	Zestimate – Sale Price
Smallest value	-53.04%	\$-126,701
1%	-9.46	-21,107
5%	-4.18	-6,352
10%	-1.96	-2,815
25%	2.57	3,328
50% (median)	7.92	9,717
75%	16.06	18,961
90%	31.05	32,712
95%	43.95	43,238
99%	67.55	89,105
Largest value	92.04	470,166

rating. There is some consistency across the descriptive statistics, the descriptive statistics from Zillow, and the descriptive statistics from *The Wall Street Journal*. The market and sample studied provide the best case for Zillow. The likelihood is that in more a volatile market with a lower accuracy rating Zillow would misprice at a higher rate and larger amount. The findings show that Zillow estimates of value overprice value by 10% compared to the actual sale price. Based on results

in the prior literature, homeowners appear to more accurately price their own houses compared to an automated valuation system such as Zillow's.

Zestimates Compared to Sale Prices

In this study, Zestimates are compared to the actual sale prices to examine the accuracy of Zillow's value estimates. As indicated in the data section, the mean difference between sale price and value estimate is

\$13,576 and the percent difference is 11.66%, with a standard deviation of 0.1534.

Equation 1 is used with and without a constant term to test Zillow's accuracy relative to the sale price. Both models reject the hypothesis that Zestimates are equal to sale prices; Table 3 reports the results. The estimated coefficient on the Zestimate without the constant is 0.90 and is significant at the 1% level. The coefficient is tested to determine if it is equal to 1; in the results, the F -test is 106.65. When the constant is included, the Zestimate coefficient is 0.8739 and the constant is 5,538. The constant is not statistically different from zero, while the F -test on the coefficient is 14.90. The 0.90 coefficient from the model without a constant indicates that Zillow overvalues houses in the sample by approximately 10%, and the coefficient in the model with the constant indicates an overvaluation of approximately 12.6%. These estimated valuations are consistent with the 11.66% percent difference noted in the descriptive statistics.

Zillow's Value of Housing Characteristics Compared to Market Value

Equation 2 and Equation 3 are used next to test the null hypothesis that Zillow accurately estimates the parameters of the hedonic model compared to the market model, that is $\alpha_i = a_i$, $\beta_i = b_i$, and $\delta_i = c_i$. The models are estimated using the Zellner seemingly unrelated regression.¹¹ The parameter estimates along with test statistics are reported in Table 4.

All variables, except *Age squared* and *Number of Bedrooms*, are statistically significant in each of the regression models. Based on Chi-square tests for equality of coefficients across models, the coefficients *Number of Bedrooms*, *Vacant*, *Garage Spaces*, and *Number of Stories* are statistically different across the models. In the sale price model, the coefficient on *Vacant* is -0.064, while it is only -0.026 in the Zestimate model. It may be more difficult for Zillow to determine the occupancy status of a house, and thus it fails to price vacancy as the market does.¹² The coefficient on *Garage Spaces* for the sale price model is 0.096 versus

0.069 for the Zestimate model; this is surprising since Zillow should be able to obtain accurate counts and estimates of this variable. Also *Number of Stories* is priced higher for Zillow than for the sale price model. The results suggest that Zillow's valuations of these housing characteristics are different from market value, possibly due to inaccurate data. The Zestimate model does have a lower root mean square error (RMSE) and a higher R -squared.

The study also examined the coefficients and differences between coefficients for the occupied sample only, but the results are not included because they are essentially the same as for the model with vacant houses and because the Zillow estimates are provided for both vacant and occupied houses. The coefficients and differences in coefficients across models are essentially the same when only occupied homes are included in the sample. The only exception is *Number of Bathrooms*, which is not statistically different across the models. It would have been preferable to use Zillow's actual model to make the comparisons; however, Zillow strongly emphasizes that its model is proprietary.¹³

Summary and Conclusions

This article examines how accurate Zillow's Zestimates are compared to actual sale prices. Zillow.com rates the accuracy of its estimates by city, county, or MSA. Thus far, independent academic verification of Zillow's descriptive statistics has not been available. The Arlington, Texas, sample used in this study was selected based on the availability of housing sales and Zillow's indication that the Zestimates for this location are at its highest accuracy level because it has the best data for this area. In addition, this was a stable market during the sample period, with house price growth of approximately 1% per year in 2006 and early 2007. This study also provides direct evidence on the relationship between sale prices and Zillow's Zestimates, following prior research on the relationship between prices and other estimates of housing values, most recently research by Agarwal,

11. Seemingly unrelated regression is a regression model developed by Arnold Zellner. It allows for analyzing a system of multiple equations with cross-equation parameter restrictions and correlated error terms; see Arnold Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias," *Journal of the American Statistical Association* 57 (1962): 348–368.

12. For example, see articles that include a vacancy variable, Fred A. Forgey, Ronald C. Rutherford, and Thomas M. Springer, "Search and Liquidity in Single-Family Housing," *Real Estate Economics* 24, no. 3 (1996): 273–292; and Thomas M. Springer, "Single-Family Housing Transactions: Seller Motivations, Price, and Marketing Time," *Journal of Real Estate Finance and Economics* 13, no. 3 (1996): 237–254. The results suggest that approximately 4% of the overpricing may be a result of the failure to model occupancy. In the sample, when the vacant properties are excluded and the regression models for Equation 1 are estimated, the coefficient is 0.91 and remains significant at 1%. This is similar to the coefficient of 0.90 with the vacant properties included. The coefficients on the two regression models are similar to the models with vacancy included with slight variation. This suggests that vacancy does not account for the overpricing.

13. See <http://www.zillow.com/wikipages/What-is-a-Zestimate/>.

Table 3 Zillow's Zestimates Compared to Sale Price

Dependent Variable	Coef.	t-Test	F-Test	F-Test	F-Model	R-Squared	RMSE
Sale Price (SP)		Coef. = 0	Coef. = 1	Coef. = 0			
Equation 1 without a constant							
Zestimate	0.902	96.17**	106.65**		9058.09	0.9826	20760
Equation 1 with a constant							
Zestimate	0.874	26.76**	14.90**				
Constant	5538.146	1.18		1.40	716.13	0.9248	20595

The two models are: $SP = Zestimate$ without a constant and $SP = Zestimate$ with a constant, see Equation 1. The F -statistic on the $Zestimate$ coefficient tests whether the coefficient is equal to 1. The F -statistic on the constant tests whether the constant is equal to 0.

**Significant at 1%, *Significant at 5%

Table 4 Regression Estimates

Dependent Variable	Log (Sale Price)	t-Test	Log (Zestimate)	t-Test	Test Across-Equation Equality of Coefficients (chi-square)
Constant	11.0886	299.06**	11.1764	379.67**	379.67**
Square Feet/100	0.03617	35.99**	0.03747	46.70**	2.98
Age/10	-0.0520	-4.79**	-0.0555	-6.41**	0.19
(Age/10) ²	0.0009	0.45	0.0023	1.45	0.90
Number of Bedrooms	0.0147	1.74	-0.0008	0.11	5.94*
Number of Bathrooms	0.0800	6.75**	0.0648	6.84**	2.94
Pool (yes=1)	0.1119	10.17**	0.1096	12.48**	0.08
Vacant (yes=1)	-0.0642	-8.40**	-0.0257	-4.21**	45.21**
Garage Spaces	0.0960	12.34**	0.0688	11.07**	21.75**
Number of Stories	-0.1084	-10.12**	-0.0703	-8.22**	22.52**
Zip Code Dummies	yes		yes		Across model coefficient tests for the zip code dummy variables are not statistically significant at 1%, with only one significant at 5%.
Observations	2,045		2,045		
RMSE	0.1625		0.1297		
R ²	0.8421		0.8833		
Model chi-square	10,904		15,481		
F-test	273.48		448.24		
AIC	-1585.16		-2507.28		
BIC	-1468.07		-2389.20		
LL (mode)	814.07		1274.64		

Two equations are estimated using Zellner's seemingly unrelated regression model. The equality of the coefficients across the two equations is also tested.

**Significant at 1%, * Significant at 5%

and Kiel and Zabel. There does not appear to be any other research that has examined this issue.

The results show that Zillow estimates overvalue property by 10% compared to the sale price. Zillow also overestimates values for approximately 80% of the houses in this sample by at least 1%; 0.88% of home values are underestimated by more than 10%, again compared to the sale price. If this were an automated appraisal district model used to estimate values for tax purpose, this would cause considerable concern. One unanswered question is why do Zestimates overestimate the value of most houses in a market, even where it has its best database and best accuracy rating?

This study provides some evidence on what variables may have been left out of the Zillow model and

may be affecting its accuracy. There is no doubt that appraisal districts track occupancy and use it in their estimations; Zillow appears to leave this out of its model even though this information is accessible through both appraisal district data and some MLS data.

The evidence indicates that Zillow overprices houses by approximately 10% based on a simple difference in means and also an OLS model. In the sample, 40% of the homes were overpriced by Zillow by more than 10%, with only 0.88% underpriced by more than 10%. The study also finds that 59% of the Zestimates fall within $\pm 10\%$ of sale price. This is similar to Mullaney's¹⁴ statement that Zillow indicates that 62% of homes are expected to be accurate within 10%. The coefficients on a Zillow model compared to the

14. Mullaney, "Hot Property."

coefficients on a sale price model indicate that Zillow prices some housing characteristics differently than the market. Specifically, vacant properties are overvalued. It appears that Zillow does not track the occupancy of a property, yet vacancy is known to affect value. The other variables that are priced differently—*Garage Spaces* and *Number of Stories*—may be the result of measurement error or perhaps failure to include these variables in the Zillow’s algorithms.

The use of Zillow.com to provide an estimate of housing value does not appear to be as accurate as owners’ estimates of value that several studies have documented. For example, Kiel and Zabel find that, on average, homeowners overestimate the value of their homes by 5.1% and new owners overvalue their homes by approximately 8.4%.

Given these results, homeowners can make good use of Zillow’s Zestimates, but in most cases Zillow should not be relied on to provide an accurate estimate of value. Zillow’s house value estimate and related information may be useful as a quick way to learn about the relative value of houses in a neighborhood, but it appears less precise than a homeowner’s own estimate of home value. Homeowner motivation and factors other than housing characteristics may make it difficult to obtain accurate prices. It appears, based on results from prior research and the current results, that homeowners are able to determine the value of

their own homes better than an automated system such as Zillow.

Daniel R. Hollas, PhD, received his PhD in economics from the University of Illinois at Urbana-Champaign. He is a professor of economics in the Department of Economics at the University of Texas at San Antonio.

His research interests include industry studies, particularly in the areas of real estate, and electricity and gas distribution. **Contact: Daniel.Hollas@utsa.edu**

Ronald C. Rutherford, PhD, received his PhD in real estate from the University of Georgia. He holds the Elmo J. Burke, Jr., Endowed Chair and is a professor of finance and real estate in the Department of Finance at the University of Texas at San Antonio. His research interests include real estate, housing analysis, corporate survival and IPOs, lease analysis, agency issues, and restructuring. **Contact: Ronald.Rutherford@utsa.edu**

Thomas A. Thomson, PhD, received his PhD in business administration from the University of Michigan, Ann Arbor. He is a professor of finance and real estate, and the director of the Real Estate Finance and Development Program in the College of Business at the University of Texas at San Antonio. His research interests include mortgage default, foreclosure, the role of appraisal in default, and other real estate issues. **Contact: Thomas.Thomson@utsa.edu**

Web Connections

Internet resources suggested by the Lum Library

American Housing Survey, U.S. Census Bureau
<http://www.census.gov/hhes/www/housing/ahs/ahs.html>

HomeGain.com
<http://www.homegain.com/>

Housing Topics, U.S. Census Bureau
<http://www.census.gov/hhes/www/>

Zillow.com
<http://www.zillow.com/>

ZipRealty.com
<http://www.ziprealty.com/>