Price-Segmented Beliefs and the U.S. Housing Boom

By Margaret M. Jacobson*
April 9, 2026
Federal Reserve Board

Abstract

Using Zillow's ZTRAX housing transaction data, this paper extends Landvoigt et al.'s (2015) analysis of San Diego, CA to several other metropolitan statistical areas and confirms their findings that expected capital gains on housing were higher for relatively lower-priced, rather than higher priced, houses during the U.S. housing boom of the 2000s. Because buyers of lower-priced houses tended to be more sensitive to looser credit conditions than buyers of higher-price houses, this paper documents patterns that are consistent with an interaction of beliefs and credit conditions in a time period where direct evidence on house price beliefs is scarce.

Keywords: housing booms, beliefs, transaction data.
JEL: D14, D91, R21, R31

1 Introduction

Optimistic beliefs and looser credit conditions are the two main drivers for the U.S. housing boom of the 2000s. Because Kaplan et al. (2020) show that (1) looser credit conditions can explain expanded access to homeownership and (2) optimistic beliefs can account for rapid house price growth, the interaction of the two drivers is key for understanding the dynamics of this episode.[1] Kaplan et al. (2020) also show that those who benefited from expanded access to credit were renters who became first-time homeowners purchasing the same sized

*This material reflects the views of the author and not those of the Federal Reserve Board of Governors. This note was previously included as a section of an earlier draft of the author's paper "Beliefs, Aggregate Risk, and the U.S. Housing Boom." Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX) and distributed through Inter-University Consortium for Political and Social Research. More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group. I thank the Bureau of Economic Analysis, specifically Scott Wentland and Ben Bridgeman, for access to and help with ZTRAX. The author thanks Aditya Aladangady, Eirik Brandsaas, and Raven Molloy for helpful comments. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant NO. 2015174787. Contact: Margaret.M.Jacobson@frb.gov

[1] See Jacobson (2024), Johnson (2019), Cox and Ludvigson (2021), and Dong et al. (2022) for studies on the interaction of beliefs and credit conditions.

home they were renting, limited data on beliefs prior to 2007 makes studying the interaction of beliefs and credit conditions challenging.[2]

To investigate how beliefs vary across price-segmented housing submarkets, this paper estimates a statistical model of prices changes developed by Landvoigt et al. (2015) using Zillow (2026) ZTRAX transaction data on repeat housing sales. Transaction-level data is key because it provides sales price data on repeated sales of the same property, which, in turn, allows for the estimation of a common component of expected capital gains and a cross-sectional dispersion component across types of houses segmented by price. Furthermore, expected capital gains can be backed out from this statistical model and thus be used to proxy for beliefs about future housing.

By expanding the analysis of Landvoigt et al. (2015) beyond San Diego, CA to include Phoenix, AZ and Cleveland, OH, this paper adds new evidence to the dearth of data on beliefs in the 2000s. Because the U.S. housing boom of the 2000s varied in timing and magnitude across metropolitan statistical areas (MSAs), as documented by Ferreira and Gyourko (2023), estimating beliefs for multiple MSAs is important for a comprehensive account of the episode.

In all three MSAs studied, expected capital gains were higher in relatively lower-priced, rather than higher priced, houses. This suggests that optimistic beliefs were more prevalent in lower-priced segments of the housing market where potential buyers were most likely to benefit from looser credit conditions that expanded their access to homeownership. This finding aligns with Kindermann et al. (2024) who show that renters tend to have more optimistic beliefs about future house prices than existing homeowners, and Kaplan et al. (2020) who find that renters purchased smaller (and hence less expensive) homes than homeowners in response to expanded access to credit.

Complementing these within-MSA findings, this paper also estimates substantial variation in expected capital gains over time and across the three MSAs studied, which corroborates the empirical evidence of Soo (2018).

[2]See Kuchler et al. (2023, Table 2) for documentation of U.S. house price beliefs and their limited availability prior to 2007. Jacobson (2024) navigates limited data on beliefs by constructing an empirical proxy based on the tight correlation of a question on house price beliefs that is available starting in 2007 with a question on selling conditions that goes back to 1992 in the University of Michigan Survey of Consumers. Additionally, Mian and Sufi (2009, 2011, 2022) show that the low-income zip codes were the most sensitive to changes in fundamentals throughout the housing boom. Corbae and Quintin (2015) and Favilukis et al. (2017) show that credit constrained households are more affected by lending standards. On the other hand, Adelino et al. (2016) show that middle-income households increased mortgage origination throughout the boom.

2 Estimates of Expected Capital Gains

Housing market transactions of repeat sales of single family homes for the MSAs of Cleveland, Phoenix, and San Diego are obtained via Zillow (2026) Ztrax assessment and transaction data accessed through the Bureau of Economic Analysis in 2019. Applying the cleaning steps detailed in Appendix A to arm's length, non-foreclosed sales of residential properties made by owner-occupiers results in over 280,000 transactions of properties that are sold at least once from 1998 to 2007.[3] It is important to limit the sample to repeat sales because the estimates of expected capital gains from the statistical model are based on observed sales of the same property.

Expected capital gains of house i at time t in MSA j vary by their current log price $\log p_t^{i,j}$ where the idiosyncratic shocks $e_{t+1}^{i,j}$ have mean zero and are such that the law of large number holds in the cross section of houses. By allowing the return for house i in MSA j to depend on its log price $\log p_t^{i,j}$, one can assess whether or not returns are uniform across all price segments by testing if $\hat{b}_t^j = 0$ as defined in equation (1). Given that Kaplan et al. (2020) find that households who are more sensitive to credit conditions tend to purchase smaller (less expensive) houses, testing if $\hat{b}_t^j = 0$ can provide insights on the interaction of beliefs and credit conditions.

The statistical model can be written as:

$$\log p_{t+1}^{i,j} - \log p_t^{i,j} = a_t^j + b_t^j \log p_t^{i,j} + e_{t+1}^{i,j} \tag{1}$$

Estimating the equation above would restrict the sample to properties that sold in successive years, such as a sale in 2000 and 2001, for example. To incorporate information from longer-dated sales, such as a house that sold in 2000 and again in 2003, for example, the above equation is iterated forward to capture expected capital gains between years $t+k$ and $t+\ell$, which are functions of a_t^j and b_t^j,

$$\log p_{t+\ell}^{i,j} - \log p_{t+k}^{i,j} = a_{t+\ell,t+k}^j + b_{t+\ell,t+k}^j \log p_{t+k}^{i,j} + \epsilon_{t+\ell,t+k}^{i,j} \tag{2}$$

Where $k \in [0, \ell - 1]$. The coefficients $a_{t+\ell,t+k}^j$ and $b_{t+\ell,t+k}^j$ in equation (2) are estimated by two-step GMM with a robust weight matrix. The first-step residuals are used to estimate the weighting matrix for the second step, which ensures that the estimator is robust to heteroskedasticity. Equation (2) is estimated separately for the three MSAs in the sample, $j \in \{\text{Cleveland, Phoenix, San Diego}\}$. The coefficients $a_{t+\ell,t+k}^j$ and $b_{t+\ell,t+k}^j$ are the appropri-

[3] Appendix A also compares estimates of San Diego, CA to those of Landvoigt et al. (2015) who use housing transaction data from Trulia instead of Zillow.

ate weighted sums/products of future a_t^j and b_t^j coefficients between years $t+\ell$ and $t+k$, respectively.

The slopes $b_{t+\ell,t+k}^j$ are the coefficients of interest and tests for cross-sectional dispersion in expected capital gains across housing price segments for any given pair of years $t+\ell$ and $t+k$ for each MSA j. If $\hat{b}_{t+\ell,t+k}^j = 0$ then there is no cross-sectional dispersion and all houses in that MSA have the same expected capital gains regardless of their price. If $\hat{b}_{t+\ell,t+k}^j > 0$, then expected capital gains are relatively higher for more expensive houses in MSA j, which are those houses that have a higher sales price at $t+\ell$. If $\hat{b}_{t+\ell,t+k}^j < 0$, then the opposite is true and relatively cheaper houses in MSA j have a higher expected capital gain. The intercepts $a_{t+\ell,t+k}^j$ is the average expected capital gain for repeat sales in an MSA between years $t+\ell$ and $t+k$.[4]

Figure (1) shows the GMM estimates of the coefficients in equation (2) as the markers and the 95% confidence intervals as the shaded bands for the MSAs of Cleveland, OH; Phoenix, AZ; and San Diego, CA. Estimates shown are those where $\ell = 1$ and $k = 0$ such that the expected capital gains from one year t to the next year $t+1$ are shown.

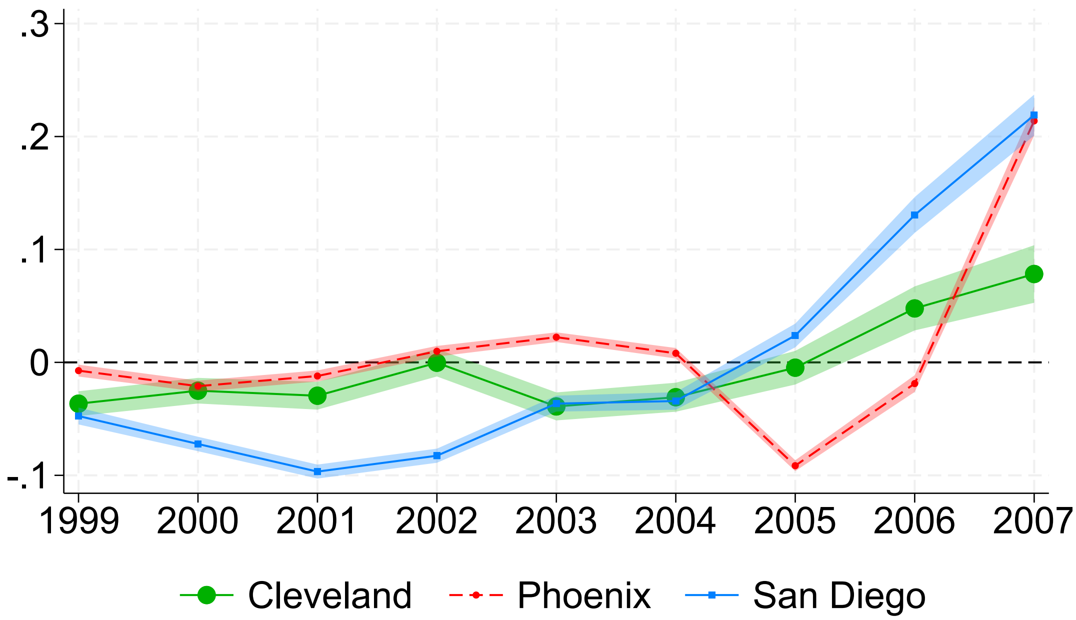
The top panel (1a) shows that expected capital gains were relatively higher for lower-priced houses for all three MSAs shown until about 2005, as shown by the estimates of the cross-sectional dispersion of expected capital gains across house price segments $\hat{b}_{t,t+1}^j$. These estimates show that a 1 percent less expensive house (relative to the average house) was expected to have at most a 0.1 percentage point higher expected capital gain at time t in MSA j between 1998 and 2007.

Panel (1a) also shows that lower-priced houses were more likely to have higher expected capital gains in San Diego, CA than in Phoenix, AZ or Cleveland, OH. Because the housing supply elasticity is lower in San Diego than in either Phoenix or Cleveland according to Saiz (2010), increased demand for less expensive houses was more likely to result in relatively higher price increases, and thus capital gains.[5] Conversely, the estimates of dispersion for Phoenix, AZ are the closest to zero in the period shown, which suggests little dispersion in expected capital gains across house price segments. The relatively uniform capital gains can be explained via Phoenix having a larger share of speculative investors than San Diego or Cleveland. As shown by Gao et al. (2020), these investors tended to be insensitive to credit conditions.[6]

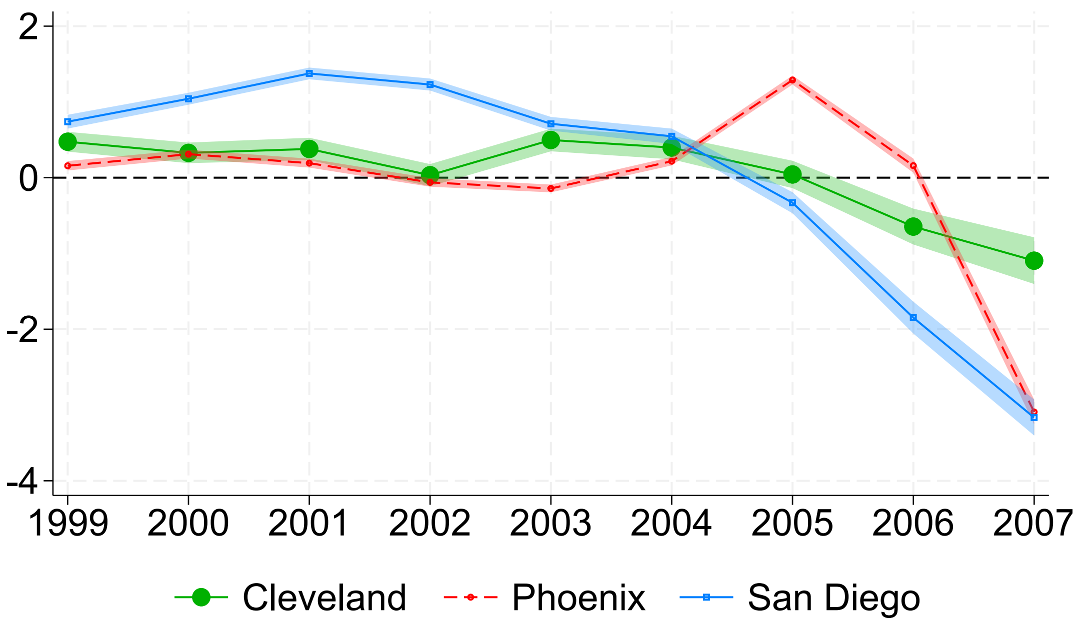
[4]If $t = 2000$ and $\ell = 3$, then for $k = 0$ $a_{2003,2000}$ is the average expected capital gains between 2000 and 2003. For $k = 1$ and $k = 2$, $a_{2003,2001}$ and $a_{2003,2002}$ are the expected capital gains between years 2001 to 2003 and 2002 to 2003, respectively.

[5]For discussions of the limitations of the estimates of Saiz (2010) see Aastveit et al. (2023), Oh et al. (2025), and Louie et al. (2025).

[6]Other studies on the role of investors include Graham (2024), Chinco and Mayer (2016), Mian and Sufi (2022), Bayer et al. (2021), Albanesi et al. (2022). Because non-owner occupiers like housing investors are



(a) Cross-sectional Dispersion of Expected Capital Gains ($\hat{b}^j_{t,t+1}$), percentage points



(b) Average Expected Capital Gains ($\hat{a}^j_{t,t+1}$), percentage points

Figure 1: Estimated average and cross-sectional dispersion of expected capital gains, with 95% confidence bands, for repeat sales of single family homes in Cleveland, OH; Phoenix, AZ; and San Diego, CA from the equation $\log p^{i,j}_{t+\ell} - \log p^{i,j}_{t+k} = a^j_{t+\ell,t+k} + b^j_{t+\ell,t+k} \log p^{i,j}_{t+k} + \epsilon^{i,j}_{t+\ell,t+k}$ where $k \in [0, \ell-1]$. The data consist of 48,968 repeat sales in Cleveland, OH; 148,842 repeat sales in Phoenix, AZ; and 84,076 repeat sales in San Diego, CA via ZTRAX (Zillow, 2019).

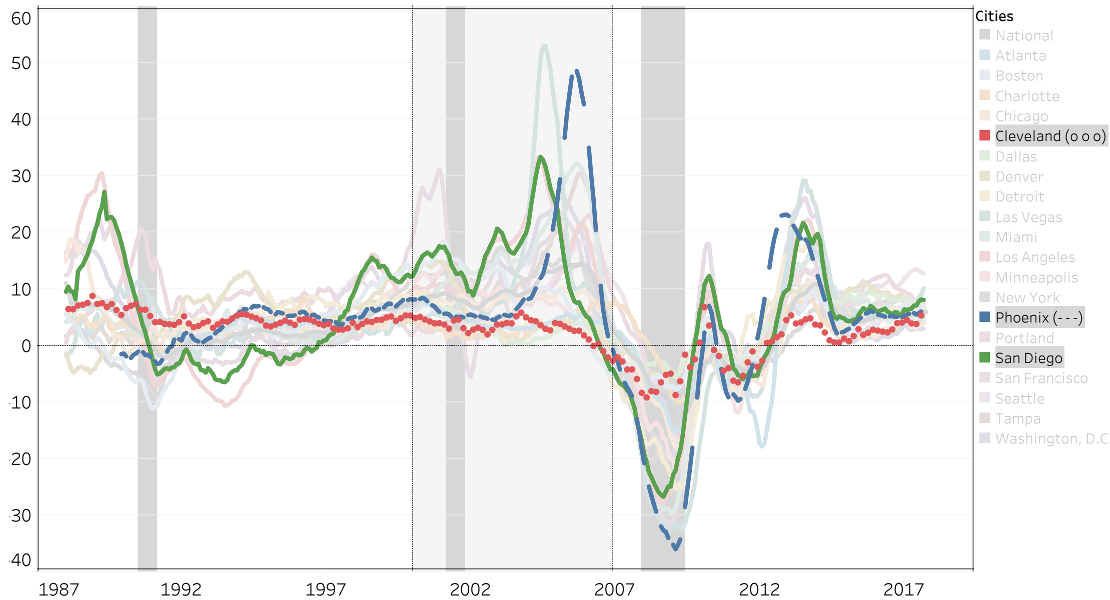


Figure 2: 12-month percentage change in house prices for select metropolitan statistical areas, percentage points. Source: Corelogic/Case-Shiller accessed via Bloomberg.

By construction, the estimated average expected capital gains, $\hat{a}^j_{t+\ell,t+k}$ shown in the bottom panel (1b), closely track the 12-month percentage change in house prices shown in figure (2) for each MSA shown. This alignment helps validate that the estimated data generating process in equation (2) correctly reflects key attributes about regional house prices—specifically the early peak for San Diego, the later arriving rapid surge for Phoenix, and virtually no boom for Cleveland. Notably, expected capital gains in Phoenix and Cleveland were similar and remained below those of San Diego until 2003, which likely reflects lower buyer incomes in these regions, as noted by Ferreira and Gyourko (2023). However, after 2003, expected capital gains in Phoenix and Cleveland diverge sharply, which is likely due to the influx of housing investors in the former, but not the latter.

It is also worth noting that figure (1) shows substantial time variation in estimates of both average and cross-sectional expected capital gains for each MSA. By the start of the bust in 2007, the positive value $\hat{b}^j_{t,t+1}$ for all MSAs indicates that relatively more expensive houses were holding their value than less expensive houses, while the negative value of $\hat{a}^j_{t,t+1}$ indicates that average expected capital gains were negative. This time variation is important for understanding the evolution of beliefs throughout various stages of a boom-bust cycle.

excluded from these estimates shown, and investors tended to be more optimistic than owner occupiers, these estimates shown are likely a lower bound on expected capital gains for Phoenix, AZ.

3 Conclusion

The boom in national U.S. house prices is largely attributed to optimistic beliefs and looser credit conditions. Because data on beliefs prior to 2007 is limited, studying the interaction of beliefs and credit conditions is challenging. Estimating the statistical model of Landvoigt et al. (2015) on transaction-level housing data provides evidence of time-varying beliefs that are disperse across price segments of housing markets. For the MSAs of Cleveland, OH; Phoenix, AZ; and San Diego, CA, relatively lower prices houses had higher expected capital gains than their higher priced counterparts. Because buyers of lower priced houses tended to be more sensitive to looser credit conditions and these houses had higher expected capital gains, this paper documents pattern that support an interaction of credit conditions and optimistic beliefs about future housing.

A Appendix: Data

This section gives detailed cleaning steps of the Zillow (2026) ZTRAX transaction and assessment data to obtain a sample similar to that Landvoigt et al. (2015) for the MSAs of Cleveland, OH; Phoenix, AZ; and San Diego, CA. The Zillow (2026) ZTRAX data is a panel of housing transactions and the cleaning steps can be grouped by those related to deeds, characteristics, and outliers.

First, to obtain housing market transactions, deeds (`documenttype`) that are not arm's length transfers of homes are dropped as are other non-arm's length deeds such as those indicating a partial sale of a house. Following Landvoigt et al. (2015), I keep only grant deeds (`GRDE`), condo deeds (`CDDE`), corporate deeds (`CPDE`), and individual deeds (`IDDE`). Because the list of deeds denoting arm's length transactions is more exhaustive for Cleveland, OH and Phoenix, AZ, I delete entries for `documenttype` that are conservator's deed (`CVDE`), deed in lieu of foreclosure (`DELU`), gift deed (`GFDE`), intrafamily transfer (`INTR`), partnership deed (`PTDE`), personal representative's deed (`PRDE`), sheriff's deed (`SHDE`), trustee's deed (`TRFC`). Deeds that are kept have values for `documenttype` that include administrator's deed (`ADDE`), agreement of sale (`AGSL`), bargain and sale deed (`BSDE`), condominium deed (`CPDE`), court order/action (`COCA`), corporation deed (`CPDE`), correction deed (`CRDE`), deed (`DEED`), fiduciary deed (`FDDE`), guardian's deed (`GDDE`), grant deed (`GRDE`), individual deed (`IDDE`), joint tenancy deed (`JTDE`), land contract (`LDCT`), other (`OTHR`), quitclaim deed (`QCDE`), re-recorded deed (`RRDE`), tax deed (`TXDE`), and warranty deed (`WRDE`).

Second, deeds are dropped based on buyer or house characteristics. Deeds without a latitude or longitude are dropped as are deeds that transfer multiple parcels as identified by the APN number. Second homes and trailers are dropped as are foreclosed properties. Buyers that are not a couple or single person are dropped eliminating buyers that are a corporation or partnership (`CO`,`PT`), a trust (`FT`,`IT`,`LV`,`RL`,`RT`,`TE`), or a beneficiary (`BF`).[7]

Single family homes are denoted by the `propertylanduse` variable from the assessment data[8] and observations that are kept include single family residences (`RR101`), condominiums (`RR106`)[9], cooperatives (`RR107`), row houses (`RR108`), planned unit developments(`RR109`), bungalows (`RR113`), zero lot lines (`RR114`), manufactured, modular and prefabricated homes (`RR115`), patio homes (`RR116`), garden homes (`RR119`), landominiums (`RR120`), and inferred single family homes (`RR999`). Dropped observations have a `propertylanduse` variable from the assessment data equal to rural residences including farms/productive land (`RR102`), mobile homes (`RR103`), residential common areas (`RR110`), time shares (`RR111`), seasonal, cabin, vacation residences (`RR112`), residential parking garages (`RR117`), and other improvements (`RR118`). Observations from the transaction data are also dropped such as those where the `propertyusestndcode` variable equals agricultural (`AG`), apartment (`AP`), commercial (`CM`), mobile homes (`MB`), mixed use (`MX`), unimproved (`UL`), multifamily (`MF`).

Lastly, to control for outliers, transactions with prices below $15,000, combined loan-to-value ratios (first plus second mortgage) above 120 percent, and annualized capital gains above 50% are dropped. To avoid the influence of house flipping, all pairs of sales that are less than 180 days apart are dropped. Some properties that were sold twice in the same year, but more than 180 days apart remain in the sample. If the property was sold more than once in the same year, then these transactions are dropped. If the property was sold in another year then the earlier sale date in the year is dropped. Given that house prices are rising throughout this period, keeping the later sale should bias capital gains downward. Given the high house price growth observed at the MSA-level in Phoenix, AZ, robustness checks were run to increase the annualized capital gain threshold from 50% to 60% and 70% with little change to the estimates. Similarly, second family homes were included in a separate robustness check with little alteration to the estimates.

[7]This is achieved by keeping observations with `buyercode` equal to domestic partners (`DP`), formerly known as (`FK`), her husband (`HH`), husband and wife (`HW`), individual (`ID`), married man (`MM`), minor (`MN`), married person (`MP`), married woman (`MW`), single man (`SM`), single person (`SP`), single woman (`SW`), unmarried man (`UM`), unmarried woman (`UW`), widowed (`WW`) and dropping those with `buyercode` equal to affiant (`AF`), borrower or trustor in default (`BR`), estate (`ES`), executor (`EX`), government (borough, city, village, etc.) (`GV`), surviving joint tenant (`SJ`), personal representative (`PR`), agent (`AG`), not provided (`NP`).

[8]The `propertylanduse` variable form the assessment data and `propertyusestndcode` variable from the transaction data are mostly but not always consistent. I use the assessment data as the main variable to denote property use because it has fewer blank observations and is more stable across time.

[9]Although it may be debatable as to whether or not condominiums should be included as single family homes, dropping condominiums decreases the number of observations from 84,076 to 53,463 for San Diego, CA and the estimates of expected capital gains differ to a larger extent from those of Landvoigt et al. (2015) shown in table (1).

Table (1) compares the estimates of average and cross-sectional capital gains for San Diego, CA to those of Landvoigt et al. (2015). Overall, the estimated coefficients from equation (2) resemble those of Landvoigt et al. (2015). A few minor discrepancies likely arise from the differences in source data used (Trulia vs. Zillow). My sample is larger than theirs with 84,076 repeat sales compared to their 70,315.

	2000	2001	2002	2003	2004	2005	2006	2007
$a^{LPS}_{t+1,t}$	1.29	1.41	1.30	0.87	0.60	-0.56	-1.09	-3.18
	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)	(0.07)	(0.10)	(0.12)
$b^{LPS}_{t+1,t}$	-0.093	-0.10	-0.09	-0.05	-0.04	0.04	0.07	0.22
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.01)	(0.01)	(0.01)
$a^{author}_{t+1,t}$	1.04	1.37	1.23	0.71	0.55	-0.33	-1.84	-3.16
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.07)	(0.10)	(0.12)
$b^{author}_{t+1,t}$	-0.07	-0.10	-0.08	-0.04	-0.03	0.02	0.13	0.21
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.008)	(0.009)

Table 1: Upper panel contains estimates from Landvoigt et al. (2015) (LPS) of 70,315 repeat sales in San Diego County during the years 1999-2008 using transaction data from Trulia. The lower panel contains a replication of their estimates using ZTRAX (Zillow, 2019) assessment and transaction data for 84,076 repeat seals in San Diego County. The numbers in parentheses are standard errors.

References

Aastveit, Knut Are, Bruno Albuquerque, and Andre K. Anundsen, "Changing Supply Elasticities and Regional Housing Booms," *Journal of Money, Credit and Banking*, 2023, *55* (7), 1749–1783.

Adelino, Manuel, Antoinette Schoar, and Felipe Severino, "Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class," *The Review of Financial Studies*, 03 2016, *29* (7), 1635–1670.

Albanesi, Stefania, Giacomo DeGiorgi, and Jaromir Nosal, "Credit growth and the financial crisis: A new narrative," *Journal of Monetary Economics*, 2022, *132*, 118–139.

Bayer, Patrick, Kyle Mangum, and James W. Roberts, "Speculative Fever: Investor Contagion in the Housing Bubble," *American Economic Review*, February 2021, *111* (2), 609–51.

Chinco, Alex and Chris Mayer, "Misinformed Speculators and Mispricing in the Housing Market," *Review of Financial Studies*, 2016, *29* (2), 486–522.

Corbae, Dean and Erwan Quintin, "Leverage and the Foreclosure Crisis," *Journal of Political Economy*, 2015, *123*, 1–65.

Cox, Josue and Sydney C. Ludvigson, "Drivers of the great housing boom-bust: Credit conditions, beliefs, or both?," *Real Estate Economics*, 2021, *49* (3), 843–875.

Dong, Ding, Zheng Liu, Pengfei Wang, and Tao Zha, "A theory of housing demand shocks," *Journal of Economic Theory*, 2022, *203*, 105484.

Favilukis, Jack, Sydney C. Ludvigson, and Stijn Van Nieuwerburgh, "The Macroeconomic Effects of Housing Wealth, Housing Finance, and Limited Risk Sharing in General Equilibrium," *Journal of Political Economy*, 2017, *125* (1).

Ferreira, Fernando and Joseph Gyourko, "Anatomy of the Beginning of the Housing Boom across U.S. Metropolitan Areas," *The Review of Economics and Statistics*, 11 2023, *105* (6), 1442–1447.

Gao, Zhenyu, Michael Sockin, and Wei Xiong, "Economic Consequences of Housing Speculation," *The Review of Financial Studies*, 03 2020, *33* (11), 5248–5287.

Graham, James, "House Prices, Investors, and Credit in the Great Housing Bust," 2024. Working Paper.

Jacobson, Margaret M., "Beliefs, Aggregate Risk, and the U.S. Housing Boom," 2024. Manuscript.

Johnson, Stephanie, "Mortgage Leverage and House Prices," 2019. Working Paper.

Kaplan, Greg, Kurt Mitman, and Gianluca Violante, "The Housing Boom and Bust: Model Meets Evidence," *Journal of Political Economy*, 2020, *128* (9).

Kindermann, Fabian, Julia Le Blanc, Monika Piazzesi, and Martin Schneider, "Learning about Housing Cost: Survey Evidence from the German House Price Boom," 2024. Working Paper.

Kuchler, Theresa, Monika Piazzesi, and Johannes Stroebel, "Chapter 6 - Housing market expectations," 2023, pp. 163–191.

Landvoigt, Tim, Monika Piazzesi, and Martin Schneider, "The Housing Market(s) of San Diego," *American Economic Review*, 2015, *105* (4), 1371–1407.

Louie, Schuyler, John A. Mondragon, and Johannes Wieland, "Supply Constraints do not Explain House Price and Quantity Growth Across US Cities," 2025. Working paper.

Mian, Atif and Amir Sufi, "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *The Quarterly Journal of Economics*, 2009, *124* (4), 1449–1496.

_ **and** _ , "House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis," *The American Economic Review*, 2011, *101* (5), 2132–2156.

_ **and** _ , "Credit Supply and Housing Speculation," *The Review of Financial Studies*, 03 2022, *35* (2), 680–719.

Oh, Hyunseung, Choongryul Yang, and Chamna Yoon, "Land Development and Frictions to Housing Supply Over the Business Cycle," 2025. Working Paper.

Saiz, Albert, "The Geographic Determinants of Housing Supply," *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Soo, Cindy, "Quantifying Sentiment with News Media Across Local Housing Markets," *The Review of Financial Studies*, 2018, *31* (10), 3689–3719.

Zillow, "Zillow Transaction and Assessment Database (ZTRAX), United States, 1940-2020," 2026. Inter-university Consortium for Political and Social Research [distributor], 2026-01-27. https://doi.org/10.3886/ICPSR39652.v2.