

# DISTANT SPECULATORS AND ASSET BUBBLES IN THE HOUSING MARKET

ALEX CHINCO AND CHRIS MAYER

**ABSTRACT.** We investigate the role of out of town second house buyers (so-called “distant speculators”) in bubble formation during the recent housing boom. Distant speculators are likely to have an excessive reliance on capital gains for financial returns and be less informed about local market conditions much like noise traders in many financial models. Using transactions level data that allows us to identify the address of the property owner, we show that increases in purchases by distant speculators (but not local speculators) are strongly correlated with high house price appreciation rates and log implied to actual rent ratios—a proxy for mispricing in the housing market. We develop a simple model that helps us address the issue of reverse causality and separate out circumstances when out of town second house buyers are simply responding to unobserved changes in home values and when they help cause house price appreciation rates and log implied to actual rent ratios to rise. Consistent with the model, we show that the size of the investment market that out of town second house buyers come from is positively related to the impact of distant speculators from that MSA on house price appreciation rates and log implied to actual rent ratios in the target MSA. We conclude by showing the large impact that distant speculators have on the local economy, with out of town second house purchases representing as much as 5% of total output in Las Vegas during the boom.

---

*Date:* April 11, 2012.

Alex Chinco: Finance Department; New York University Stern School of Business. email: alexchinco@gmail.com. Chris Mayer: Columbia Business School and NBER; Visiting Scholar, Federal Reserve Bank of New York. email: cm310@columbia.edu. Daniel Hubbard, Laura Vincent and James Witkin provided dedicated research assistance. Moshe Cohen, Xavier Gabaix, Robin Greenwood, Gur Huberman, Tomasz Piskorski, Jeremy Stein and Reed Walker as well as seminar participants at Harvard Business School, NYU Stern and Columbia Business School have offered extremely helpful comments and suggestions. We are grateful for DATAQUICK, 1010DATA and an anonymous firm for providing data and assistance that were crucial for this project.

## 1. INTRODUCTION

How do bubbles form? Beginning with the work of Black (1985 [4]) and De Long et al. (1990 [7]), many authors have conjectured that the trading behavior of overconfident or uninformed speculators can destabilize financial markets and create bubbles. According to these models, other traders may not be able to restore equilibrium because of limits to arbitrage such as capital constraints<sup>1</sup>, informational frictions<sup>2</sup> or a limited supply of tradeable shares<sup>3</sup>. Real estate researchers have also long puzzled over the inefficiency of housing prices.<sup>4</sup> Several papers specifically point to the possible role of second house buyers in inflating house prices during the recent boom.<sup>5</sup>

In order to test whether or not some combination of speculative trading and arbitrageur constraints generates a bubble, an economist must confront three key challenges: first, identifying a group of overconfident or uninformed speculators; second, showing that an increase in the trading volume of these speculators predicts future mispricing; and third, demonstrating that these speculators are not simply responding to unobserved variation in asset values—i.e., ruling out reverse causality. This is a tall order.

Consider the 500% growth in the price of CISCO SYSTEM, INC. (ticker: CSCO) from Jan. 1998 to Mar. 2000 during the Dot-com boom. Anecdotal evidence<sup>6</sup> suggests that a

---

<sup>1</sup>See Schleifer and Vishny (1997 [37]).

<sup>2</sup>See Scheinkman and Xiong (2003 [36]).

<sup>3</sup>See Ofek and Richardson (2003 [31]) or Hong et al. (2006 [19]).

<sup>4</sup>See the seminal paper by Case and Shiller (1989 [6]) as well as a recent survey Mayer (2011 [27]) for a discussion of the literature on housing bubbles.

<sup>5</sup>Bayer et al. (2011 [3]) document the role of “speculators” that sell a small number of houses trying (unsuccessfully) to time the market in Los Angeles. These authors find that speculator trading behavior is strongly associated with neighborhood price instability. Haughwout et al. (2011 [17]) examine credit report data and show that mortgages on second homes represented nearly one-half of all mortgages originated in the 4 states with the highest price appreciation at the peak of the market. Gao and Li (2012 [25]) present theoretical results that second home buyers can fuel a boom and empirical evidence showing that second home buyers both are more likely to be present in MSAs with high price appreciation and subsequently default at higher rates. The results in these papers are complimentary to ours in that all of these papers document the large growth in second home purchases in the highest appreciating MSAs; however, none of the existing papers is able to directly address the issue of causality. Our work, below, also extends this analysis to differentiate between local second home buyers and out of town second home buyers and shows that only the purchases of the latter group appear to be causing some degree of mispricing.

<sup>6</sup>See MacKay (1841 [26]) and Greenwood and Nagel (2009 [14]).

large number of inexperienced traders increased their holdings of technology and communications stocks during this time period while many traders began active stock trading for the first time. It might seem obvious that this rapid growth in the price of CISCO's stock must have been driven by this influx of overconfident or uninformed speculators. After all, CISCO did not come close to delivering a dividend stream that warranted its price in early 2000.<sup>7</sup>

Take the first challenge: identifying a group of misinformed traders engaged in price speculation. In the late 1990s the internet was a new and revolutionary technology, so some might argue that traders buying shares of CISCO were not engaged in price speculation but rather believed that the company would in fact pay out large dividends in the future. Furthermore, it is not obvious that an increase in the demand of overconfident or uninformed speculators actually predicts future price appreciation in this example. Brunnermeier and Nagel (2004 [5]) document that a group of hedge funds did not simply exert a corrective force on market prices. Instead the hedge funds bought technology stocks during the boom and sold them just prior to the crash, likely to another group of misinformed speculators (possibly day traders). Thus, in regards to the second challenge, speculators are a heterogeneous group and it is challenging to link the timing of speculative trading to mispricing of individual stocks. Finally, with respect to the third challenge, even if a large number of speculators such as the hedge funds documented above were trading CISCO's stock from Jan. 1998 to Mar. 2000, it is difficult to sort out whether speculative trading caused the extraordinary 500% price increase or instead that speculators were attracted to CISCO's stock by the perception that it was undervalued.<sup>8</sup> Looking at a time series of a single stock or multiple stocks whose share prices are highly correlated limits the identification strategies available to an econometrician.

Like the stock market, the US residential housing market exhibits strong booms and busts resembling bubbles and offers researchers detailed microdata on home buyers. As well, the

---

<sup>7</sup>While CISCO's stock price peaked at a price of \$79.37 in Mar. 2000, it fell precipitously from this level over the course of the next year and as of Mar. 2012 remains at \$15.78.

<sup>8</sup>This argument involving reverse causality is commonly referred to as the Friedman critique and dates to Friedman (1953 [11]). See Abreu and Brunnermeier (2003 [1]) for an example of a model where traders arrive in a market in order to earn profits from excess price appreciation.

housing market is geographically segmented into metropolitan statistical areas (henceforth, MSAs) making it an attractive laboratory to study bubble formation because home prices do not follow the same time series pattern. As documented in Ferreira and Gyourko (2012 [10]) the recent boom began at different times, and prices exhibited sharply different appreciation rates across the various housing markets in the US. Even the subsequent decline in prices differed by a year or more in different markets. We make use of these features to test for speculative bubbles in 21 MSAs from Jan. 2000 to Dec. 2007.

We describe the datasets used in our analysis in Section 2. These data include sales and mortgages transactions for every housing unit in this 21 MSA sample as well as monthly indexes for real house prices (henceforth, HPI) and implied-to-actual rent ratios (henceforth, IAR ratios) for each of these MSAs. The IAR ratio data is taken from Himmelberg et al. (2005 [18]) and is a measure of mispricing in the housing market derived by comparing the cost of renting a house and the imputed rent to an owner occupant (the annual after-tax cost of owning a home).

Section 3 outlines a simple economic model of speculation. This model illustrates how we employ housing data to address the three challenges listed above and study the price impact of a specific group of overconfident or uninformed speculators—namely, out of town second house buyers. We discuss information available to local and out of town second house buyers and how it might affect their trading and the determination of market prices.

Next, Section 4 addresses the first of the three key challenges. We show that out of town second house buyers, i.e. traders that buy a house in a different MSA from which they live, behave much like overconfident or uninformed speculators. For out of town second house buyers (so-called “distant speculators”<sup>9</sup>), future house price appreciation rates are likely to

---

<sup>9</sup>In the analysis below, we assign precise definitions to the terms second house buyer and local and distant speculators. We refer to all traders who purchase a home they do not reside in as “second house buyers” or “speculators.” Such a home might in fact represent a second, third, fourth (etc. . .) house in addition to their primary residence, or even just a first home if they do not own their primary residence. We use the term “speculator” because second home buyers are less able to consume the full dividend stream from their purchases relative to owner-occupants due to tax disadvantages and agency conflicts on maintenance. We specifically focus on distant speculators who face additional cost of gathering information and in conducting repairs and maintenance, and thus are more dependent on capital appreciation for their return, as discussed in Section 4. This term is not a synonym for irrational traders. We avoid using the term investors in that all house buyers are investors.

play a larger role in purchasing decisions. As well, the opinions of these distant speculators about future house price appreciation rates are likely to be less informed than the opinions of local second house buyers or owner occupants.

Consider buyers living in San Francisco and purchasing houses in Las Vegas. Out of town house buyers can only visit houses bought in Las Vegas for a fraction of the year meaning that the consumption dividend stream from living in the second house is unused for much of the time. Distant speculators trying to rent their second houses are more likely to hire an agent and thus faces additional agency costs as well as higher property management fees. Out of town second house buyers also face higher property tax rates in many jurisdictions. In contrast with owner occupants or a local second house buyers, out of town second house buyers must also travel and spend additional time and energy learning about the new housing market before making his/her purchase. Consistent with the view that distant speculators may be less informed, we show that such buyers entered the market in much larger numbers just prior to the peak in house price levels and earned negative internal rates of return on their investment relative to local speculators, with the worst relative performance in the most mispriced markets. This evidence suggests that distant speculators were not optimally timing the market.

In Section 5, we address the second key challenge by estimating a number of predictive regressions. The results confirm that a rise in the number of purchases made by out of town second house buyers in an MSA predicts both higher house price appreciation rates and log IAR ratios over the next year. Specifically, we find in Tables 7 and 8 that a 1%/yr increase in the number of purchases made by out of town second house buyers in an MSA in a given month as a fraction of monthly sales is associated with an 0.99%/yr increase in the house price appreciation rate over the next year. Likewise, a 1%/yr increase in the number of purchases made by out of town second house buyers in an MSA as a fraction of monthly sales is associated with a 1.43% increase in the log IAR ratio. Thus, an increase in the number of distant speculators in a market is correlated with both higher house price appreciation rates and higher log IAR ratios, an indication of greater mispricing. By contrast, results for

local second house buyers are mixed and do not suggest a large effect of local speculators in predicting bubbles. The lagged share of local second house buyers is correlated with slightly higher log IAR ratios, but an increase in the lagged share of local second house buyers actually predicts a decline in house price appreciation rates.

Having defined a group of traders in the US residential housing market that behave like overconfident or uninformed speculators and appear to predict mispricing, Section 6 takes on the third and final challenge—the question of reverse causation. Here we exploit the geographic segmentation of the housing market to tease apart two hypotheses: the null hypothesis that distant speculators chose to purchase second houses in MSAs that were underpriced based on the information available to them at the time and the alternative hypothesis that variation in the purchasing decisions of out of town second house buyers caused increases in house price appreciation rates and log IAR ratios in some MSAs. Our key observation is that, if the null hypothesis is true and unobserved variation in the house values in MSAs such as Las Vegas and Miami caused extreme house price appreciation, then out of town second house buyers should purchase in roughly equal proportions from all other MSAs after controlling for factors such as distance and information transmission. A change in the value of housing in MSA  $i$  is a common shock to speculators living in all other MSAs.

In our analysis, we find that house price appreciation rates and log IAR ratios in MSA  $j$  are highest when second house buyers in an MSA  $i$  with lots of potential distant speculators purchase houses in MSA  $j$ . These regressions control for both MSA pair specific factors and macroeconomic factors with ordered MSA pair and time fixed effects. The coefficient on the interaction between the share of speculators purchasing houses in a distant market and the overall number of speculators in that market is positive and significant. In other words, the size of the investment market that out of town second house buyers come from is positively related to the impact of distant speculators from that MSA on house price appreciation rates and log IAR ratios in other MSAs. This violation of symmetry allows us to reject the null hypothesis and is thus consistent with the hypothesis that distant speculators themselves helped push up home prices when they enter markets in great numbers. The particular way

that the symmetry is violated suggests a commonality in the beliefs of second house buyers within each MSA.

We conclude by pointing out similarities between the US housing bubble and housing bubbles in other countries such as Spain, where a large influx of distant speculators from Germany and Britain appear to have driven up prices. In particular, we note that purchases by distant speculators represented as much as 5% of local output in Las Vegas, a similar estimate to the share of foreign direct investment in Spain during the housing bubble of 2007 and 2008.

## 2. DATA DESCRIPTION

We use data drawn from 3 main sources: county deeds records, HPI data from ZILLOW, INC. and IAR ratio data computed according to the algorithms developed in Himmelberg et al. (2005 [18]). Subsections 2.1, 2.2 and 2.3 describe each of these data sources and present summary statistics. Once cleaned, our data represents 21 MSAs indexed by  $i = 1, 2, \dots, I$  over the time period  $t = 1, 2, \dots, T$  with  $t = 1$  denoting Jan. 2000 and  $t = T$  denoting Dec. 2007. Appendix A provides a more detailed description of the deeds data.

**2.1. Transaction Level Deeds Records.** A deed is a written legal instrument that passes the rights to a particular property (in our case a single family house) from one owner to the next. The deeds records are public in most states due to information disclosure acts and are maintained by the local county. Deeds records document any time a property is sold or a new mortgage is taken out by an owner using the property as collateral. Together, these data contain a complete sales history of any parcel of land in each of the 21 MSAs which we track. Below, we define variables denoting the number of properties and number of sales in an MSA in a given month.

**Definition (Sales).** *Define  $X_{i,t}$  as the annualized number of single family houses sold in MSA  $i$  at month  $t$  in units of houses per year.*

While the term speculator is often tossed around in everyday conversation, as discussed in Section 1 the identity, motives, and information available to traders are generally hard

to isolate. One advantage of using the U.S. residential real estate market to study bubble formation is that we can obtain information on buyers and sellers via county deeds records. Namely, for each property transaction in our database, we observe not only an address for the property itself but also a billing address where the county sends the tax bill for the property. Below, we define variables denoting the identity of house purchases by speculators in an MSA in a given month. In Section 4 we give evidence that distant speculators rely more heavily on capital appreciation than either local speculators or owner occupants and are also likely less informed about local market conditions.

**Definition** (Second House Purchases). *Define  $S_{i \rightarrow j, t}$  as the annualized number of single family houses sales in MSA  $j$  at month  $t$  where:*

- (1) *The mailing address of the tax bill and the property address recorded in the deeds records do not match, and*
- (2) *The mailing address is located in an MSA  $i$ .*

*where  $S_{i \rightarrow j, t}$  has units of houses per year.*

**Definition** (Distant Speculator Purchases). *Define  $S_{j, t}^{\text{Distant}} = \sum_{j \neq i} S_{i \rightarrow j, t}$  as the annualized number of second house purchases in MSA  $j$  at month  $t$  where the mailing address is located in an MSA  $i$  with  $j \neq i$ .  $S_{i, t}^{\text{Distant}}$  has units of houses per year.*

**Definition** (Local Speculator Purchases). *Define  $S_{j, t}^{\text{Local}} = S_{j \rightarrow j, t}$  as the annualized number of second house purchases in MSA  $j$  at month  $t$  where both the mailing and property addresses are located in MSA  $j$ .  $S_{j, t}^{\text{Local}}$  has units of houses per year.*

Table 1 gives an example of an owner occupant, local second house buyer and out of town second house buyer in our data. In the mid 2000s, the number of purchases by distant speculators in MSAs like Las Vegas, Miami, and Phoenix grew appreciably relative to their level at the beginning (and end) of our sample period as evidenced by the sparkline plots in Table 2, which gives summary statistics for the number of distant speculator purchases in each MSA  $i$  as a fraction of the total number of properties in MSA  $i$ . At peak, distant



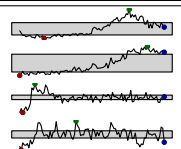

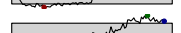
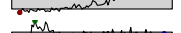

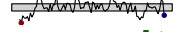
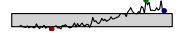





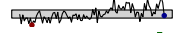

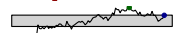

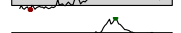




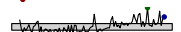
	Property Address	Tax Bill Address	Price	Date
1	1 Telegraph Hill Blvd, SF	1 Telegraph Hill Blvd, SF	\$151	04/15/2002
2	200 Fremont St, LV	888 W Bonneville Ave, LV	\$154	10/20/2003
3	200 Fremont St, LV	709 N La Brea Ave, LA	\$300	05/01/2006

**Table 1.** *This table displays 3 fictitious observations from the deeds records illustrating the basic structure of the data. The columns display the reported property address, tax bill address, sales price and sales date. Row 1 represents a purchase by an owner occupant, row 2 represents a purchase by a local second house buyer and row 3 represents a purchase by an out of town second house buyer.*

speculators always represent a minority of house purchases. In the most extreme market, Las Vegas, distant speculators purchased 17% of all housing units in 2004, up from roughly 7% percent in the early 2000s. Each of the MSA specific plots displays a similar hump-shaped pattern in the number of distant speculator purchases as a fraction of properties. A key insight for our analysis is that the scale of the patterns differ by orders of magnitude. For example, while both Miami and Milwaukee show similar percent change rises in the fraction of all houses bought by out of town second house buyers from 2002 to 2006, at the peak of the housing boom Miami had around 3 times as large a fraction of purchases made by out of town second house buyers as Milwaukee.

Research on the role of investors in housing bubbles typically treats local and distant speculators in the same way. However, as demonstrated in Table 3, purchases by local speculators exhibit a very different time series pattern than distant speculators. The overall share of purchases by local speculators varies much less across markets compared to the variability in house price appreciation. As well, in most cases (Las Vegas is an appreciable exception), the share of local speculators does not exhibit a hump with a peak at or near the peak of home prices.

**2.2. House Price Changes.** We obtain the monthly house price index (HPI) level from ZILLOW, INC. at the MSA level for each of the 21 MSAs used in our analysis from Jan. 2000 to Dec. 2007. Zillow data are available for a larger number of locations than S&P/Case and Shiller and include a number of innovations that make the index less sensitive to changes in the mix of properties that sell at a given point in time. In general, the Zillow indexes behave

$i$		Distant Speculators as a Percent of Sales						
		Mean	Sd	Min	Q25	Q50	Q75	Max
Baltimore		4.76	1.94	2.34	3.19	4.34	6.32	9.69
Charlotte		3.33	2.32	0.497	1.39	2.38	5.88	7.80
Cincinnati		6.27	1.18	2.78	5.82	6.29	6.84	9.46
Cleveland		5.37	0.942	3.03	4.87	5.38	5.98	7.49
Denver		2.20	1.28	0.676	1.08	1.70	3.29	5.54
Jacksonville		5.92	2.62	2.20	3.82	4.97	7.51	12.3
Las Vegas		11.0	3.83	4.67	7.03	12.0	14.5	17.1
Los Angeles		1.15	0.437	0.224	0.877	1.13	1.40	2.19
Miami		4.59	1.52	2.02	3.23	4.41	5.88	7.39
Milwaukee		1.28	0.599	0.193	0.857	1.19	1.63	2.94
Minneapolis		1.38	0.813	0.177	0.708	1.32	2.00	3.39
Orlando		9.86	3.41	3.16	7.48	9.99	12.4	15.7
Philadelphia		2.63	1.29	0.757	1.46	2.57	3.56	5.58
Phoenix		7.67	2.95	3.58	5.52	6.73	9.40	15.5
Riverside		8.33	1.30	5.62	7.41	8.23	9.41	11.4
Sacramento		6.49	0.951	4.31	5.88	6.56	7.37	8.28
San Diego		3.07	1.46	1.38	1.87	2.57	4.02	7.51
San Francisco		2.33	0.393	1.60	2.06	2.31	2.53	3.65
San Jose		1.86	0.451	0.693	1.56	1.83	2.13	3.06
Tampa		7.74	2.49	3.66	5.71	7.34	9.86	12.5
Washington		1.35	0.426	0.591	1.05	1.24	1.63	2.52
Mean		4.57	1.50					

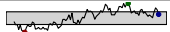

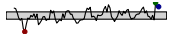




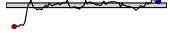

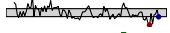



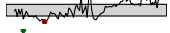

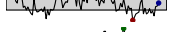




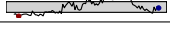
**Table 2.** This table displays the percentage of single family house purchases made by distant speculators in each MSA  $i$  in each month  $t$  over the time interval from Jan. 2000 to Dec. 2007. The shaded region in the sparkline graphs covers the interquartile range for each MSA and is not a constant scale.

quite similarly to S&P/Case and Shiller indexes during the boom, but show less of a sharp decline in 2007 and 2008 relative to S&P/Case and Shiller.<sup>10</sup>

**Definition** (House Price Appreciation Rate). Define the  $\Delta \log P_{i,t \rightarrow (t+12)} = \log P_{i,t+12} - \log P_{i,t}$  as the house price appreciation rate in MSA  $i$  at month  $t$  in units of  $1/\text{yr}$ , where  $P_{i,t}$  is the HPI index level normalized to be unity in a base year.

Table 4 gives summary statistics for the house price appreciation rate in each of the 21 MSAs in our analysis over the time interval from Jan. 2000 to Dec. 2007. A number of

<sup>10</sup>S&P/Case and Shiller indexes show a sharper decline than Zillow indexes in the immediate bust, but then a run-up in 2009-2010. Both indexes converge by 2011, but the Zillow index has less volatility by a slow steady decline in prices from 2009 to 2011. For more information about Zillow Price Index computation, see <http://www.zillow.com/blog/research/2012/01/21/zillow-home-value-index-methodology/>.

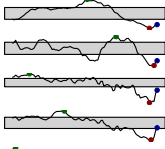





















$i$	Local Speculators as a Percent of Sales							
	Mean	Sd	Min	Q25	Q50	Q75	Max	
Baltimore		13.1	3.14	7.12	10.3	13.6	15.5	18.9
Charlotte		9.49	1.46	6.72	8.38	9.38	10.5	12.6
Cincinnati		12.3	1.77	6.87	11.1	12.2	13.6	16.7
Cleveland		10.5	1.64	6.84	9.19	10.4	11.4	15.2
Denver		9.94	2.42	6.41	7.99	9.17	11.7	16.2
Jacksonville		17.0	1.73	13.1	15.8	16.9	18.0	24
Las Vegas		12.8	3.36	7.30	10.3	12.8	14.7	19.7
Los Angeles		10.3	2.29	2.95	9.92	10.8	11.6	13.5
Miami		14.6	1.95	10.8	13.2	14.3	16.1	18.8
Milwaukee		10.1	1.96	5.46	8.63	9.82	11.6	16.3
Minneapolis		13.3	4.32	5.82	9.04	14.1	16.5	22.6
Orlando		15.9	2.16	10.9	14.1	15.9	17.3	22.7
Philadelphia		16.0	3.08	9.99	13.8	16.4	18.1	22.6
Phoenix		16.2	2.60	11.9	13.9	16.1	18.1	22.1
Riverside		10.4	1.08	8.38	9.58	10.2	11.2	13.3
Sacramento		11.6	1.37	8.65	10.6	11.5	12.7	14.1
San Diego		12.7	2.41	7.78	10.5	13.5	14.6	17.7
San Francisco		9.97	1.51	6.70	8.66	10.0	11.1	14.2
San Jose		8.05	1.92	5.03	6.78	7.68	8.87	15.2
Tampa		17.7	2.55	12.9	15.4	17.3	19.3	24.9
Washington		8.98	1.78	6.52	7.51	8.76	10.2	13.9
Mean		13.8	2.32					

**Table 3.** This table displays the percentage of single family house purchases made by local speculators in each MSA  $i$  in each month  $t$  over the time interval from Jan. 2000 to Dec. 2007. The shaded region in the sparkline graphs covers the interquartile range for each MSA and is not a constant scale.

the markets saw annual appreciation rates above 20%/yr, with house price appreciation rates exceeding 35%/yr in Las Vegas and Phoenix near the peak of their booms.

**2.3. Implied to Actual Rent Ratio.** Beginning with Poterba (1984 [34]), many authors have priced residential real estate by comparing the price of a house to the present value of its stream of rental payments, taking into account the favorable tax treatment given to owner occupants and mortgage rates. This pricing strategy is similar to the dividend discount model for the stock market. We refer to models that price housing along this margin as user cost models.

Unlike the stock market where analysts have actual dividends and share prices, in the housing market it is quite unusual to have matched data on the sale price and rental rate over

$i$		Annualized House Price Appreciation Rates in Percent						
		Mean	Sd	Min	Q25	Q50	Q75	Max
Baltimore		6.49	9.31	-12.3	-1.77	10.1	12.0	21.4
Charlotte		0.951	3.43	-6.52	-1.56	1.30	3.87	6.39
Cincinnati		0.048	2.39	-6.32	-1.05	0.796	1.70	3.23
Cleveland		-1.87	4.23	-11.2	-4.98	-0.322	1.09	4.48
Denver		-0.309	4.40	-9.61	-2.63	-0.237	0.976	11.3
Jacksonville		4.59	9.94	-19.3	2.26	8.20	10.2	17.0
Las Vegas		3.69	18.0	-34.5	-3.66	5.10	7.83	44.3
Los Angeles		6.21	14.9	-28.6	-4.49	10.9	16.6	27.6
Miami		6.48	16.7	-31.1	-1.66	12.9	14.9	27.3
Milwaukee		1.49	5.53	-9.36	-2.20	1.15	4.58	14.2
Minneapolis		1.86	7.00	-13.2	-2.99	4.43	7.07	10.5
Orlando		5.25	15.4	-28.7	0.897	7.25	12.6	32.3
Philadelphia		5.62	6.69	-8.54	0.499	7.70	11.4	12.9
Phoenix		3.53	16.3	-25.4	-7.37	3.67	7.28	39.4
Riverside		6.49	19.1	-40.2	-2.73	11.6	16.8	33.2
Sacramento		5.06	15.7	-26.8	-12.0	13.4	17.0	22.7
San Diego		4.47	14.7	-25.2	-7.63	9.04	15.7	27.9
San Francisco		2.29	12.0	-25.2	-5.46	3.96	11.3	21.9
San Jose		1.33	11.0	-19.2	-5.65	0.327	9.23	25.8
Tampa		4.66	14.3	-26.5	-2.03	9.44	11.5	23.9
Washington		6.43	12.8	-21.7	-6.27	11.1	15.9	20.7
Mean		3.56	11.14					

**Table 4.** This table displays the house price appreciation rate in each MSA  $i$  from month  $t$  to month  $t + 12$  in units of %/yr over the time interval from Jan. 2000 to Dec. 2007. The shaded region in the sparkline graphs covers the interquartile range for each MSA and is not a constant scale.

the next year for a particular house. Himmelberg et al. (2005 [18]) suggest a methodology that allows us to create an index of mispricing by comparing the ratio of the imputed rent level to the actual rent level, where the imputed rent is calculated by multiplying the user cost times the price of an owner-occupied house. We use the user cost of housing data from Himmelberg et al. (2005 [18]) updated through Dec. 2007.

**Definition** (User Cost of Housing). Define  $U_{i,t \rightarrow (t+12)}$  as the user cost of housing in MSA  $i$  in month  $t$  which reflects the fraction of the price of a house that an owner must pay in order to live in that house over the next year from time  $t$  to time  $t + 12$ :

$$U_{i,t \rightarrow (t+12)} = \rho_t + \omega_{i,t} - \tau_{i,t} \cdot \{\mu_t + \omega_{i,t}\} + \delta - \mathbb{E} \left[ \Delta \log P_{i,t \rightarrow (t+12)} \right] \quad (1)$$

where the user cost of housing has units of 1/yr.

Table 5 gives the data sources and set of short descriptions for the input variables used to compute the user cost of housing in the equation above.

Variable	Source	Description
$\rho_t$	CRSP	Riskfree rate computed as annualized 10yr T-Bill.
$\omega_{i,t}$	Emrath (2002 [9])	Property tax rate.
$\mu_t$	Federal Reserve Bank of St. Louis	Mortgage interest rate.
$\tau_{i,t}$	NBER	Federal marginal tax rate.
$\delta$	Harding et al. (2000 [16])	Housing capital depreciation rate.
$E[\Delta \log P_{i,t \rightarrow (t+12)}]$	Gyourko et al. (2011 [10]), the US Census, and the Livingston Survey	Expected house price appreciation rate equals historical long-term real growth rates by MSA plus expected inflation.

**Table 5.** This table gives both the data source and a short description of the input variables used to compute the user cost of housing in Himmelberg et al. (2005 [18]). All variables have units of 1/yr, reflect rates over the time interval  $t$  to  $t + 12$  and are known at time  $t$ .

In the standard user cost model, the price of a house in an MSA  $i$  at month  $t$  multiplied by the prevailing user cost of housing should equal the rental rate over the next year, or  $P_{i,t} \cdot U_{i,t \rightarrow (t+12)} = R_{i,t \rightarrow (t+12)}$ . We collect monthly estimates of the annualized rent for a 2-bedroom apartment in each of the 21 MSAs in our analysis from Jan. 2000 to Dec. 2007 from REIS.

**Definition** (Apartment Rental Rate Index). Define  $R_{i,t \rightarrow (t+12)}$  as the apartment rental rate index in MSA  $i$  at month  $t$  which reflects the annual rent payment required to live in 2-bedroom apartment in MSA  $i$  from month  $t$  to  $t + 12$  in units of 1/yr.

The log IAR ratio can be thought of as the excess return over the apartment rental rate of a trading strategy whereby an agent borrows money at rate  $\rho_{t/\text{yr}}$  to buy a house, lives in the house for a year while paying a constant proportion of the house value in depreciation costs  $\delta/\text{yr}$  and earning the tax shield  $\tau_{i,t/\text{yr}}$  on his debt and then sells the home after one year getting capital gains at the expected price appreciation rate of  $E[\Delta \log P_{i,t \rightarrow (t+12)}] / \text{yr}$ .<sup>11</sup>

<sup>11</sup>Himmelberg et al. (2005 [18]), do not allow the risk premium or leverage to change over time. Thus the computation can be thought of as a long-run measure of the relative price of owning versus renting,

When the IAR ratio in a given metropolitan area exceeds unity, owning a house is more expensive than renting relative to the average value over the sample period. If the index equals 1.2, for example, it means that purchasing a house is about 20%/yr more expensive than renting relative to the average of the ratio between Jan. 1980 and Dec. 2007.

**Definition** (Implied to Actual Rent (IAR) Ratio). *Define  $Z_{i,t}$  as the IAR ratio in MSA  $i$  at month  $t$  which reflects the ratio of the cost to a potential owner of borrowing money, purchasing a house and then selling it in 1yr to the cost at which he can rent a comparable property for the same amount of time:*

$$\begin{aligned} Z_{i,t} &= \frac{1}{\bar{Z}_i} \cdot \left( \frac{P_{i,t} \cdot U_{i,t \rightarrow (t+12)}}{R_{i,t \rightarrow (t+12)}} \right) \\ \bar{Z}_i &= \frac{1}{T} \cdot \sum_{t=1}^T \left( \frac{P_{i,t} \cdot U_{i,t \rightarrow (t+12)}}{R_{i,t \rightarrow (t+12)}} \right). \end{aligned} \tag{2}$$

*The IAR ratio is scaled to equal 1 relative to the average value of the ratio from Jan. 1980 to Dec. 2007.*

The IAR ratio is computed using HPI data from both the Federal Housing Finance Administration and Zillow since the Zillow house price indexes are not available prior to 1996. Table 6 gives summary statistics for the log IAR ratio in MSA  $i$  over the time interval from Jan. 2000 to Dec. 2007. This measure of mispricing varies substantially across markets such as Phoenix and Denver, respectively. At the peak in Phoenix, a tenant renting an apartment for 1000\$/mo would have to pay 1658\$/mo in mortgage payments and other costs in order to buy an equivalent house and live in it from Jan. 2004 to Dec. 2004. By comparison, in Denver, this ratio was around 1.267 between 2004 and 2006, so a tenant would have paid about 1267\$/mo to purchase a house that rented for 1000\$/mo and live in it from Jan. 2004 to Dec. 2004. While houses in Denver were still priced at a small premium relative to renting at the peak of the national boom, they appeared much less overpriced than houses in Phoenix at the same time.

---

abstracting from important short-run considerations like easy and cheap leverage in the mid-2000s and time varying risk premia.

Researchers have critiqued the user cost approach in a number of ways. For example, Glaeser and Gyourko (2007 [13]) point out that very few single family houses are rented, so any rental index is not assured to match up with the price index. Also, the user cost model as estimated above is inherently static, so it cannot easily incorporate time varying factors like risk premia, the expected growth rates of house prices, mean-reverting interest rates, credit constraints, and mobility.<sup>12</sup>

Nonetheless, a simple analysis of the user cost suggests it is well-suited for the purposes of our paper in that it allows us to estimate a single index value that proxies for overpricing.<sup>13</sup> Hubbard and Mayer (2009 [20]) estimate the log-linearized version of the user cost model:

$$\log P_{i,t} = \alpha + \beta \cdot \log R_{i,t \rightarrow (t+12)} + \gamma \cdot \log U_{i,t \rightarrow (t+12)} + \varepsilon_{i,t} \quad (3)$$

over the time interval from Jan. 1980 to Dec. 2007 with both MSA and year fixed effects.<sup>14</sup> The authors find coefficients of  $\gamma = 0.93$  and  $\beta = -0.75$ , which are very close to the values of 1.0 and  $-1.0$  respectively as predicted by the static user cost model. Thus, even though it has many imperfections, the user cost appears to provide a simple benchmark for what housing prices might be in a long-term equilibrium.<sup>15</sup>

### 3. A SIMPLE MODEL OF SPECULATION

In this section we develop a simple rational expectations model of the US residential housing market in order to clarify the empirical strategy used in our analysis below. We begin in Subsection 3.1 by outlining the basic economic framework. Then, in Subsection 3.2

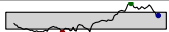
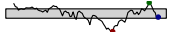
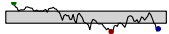
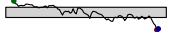
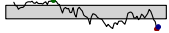





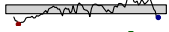



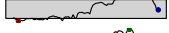

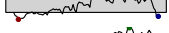
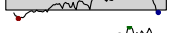
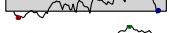
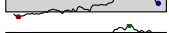

<sup>12</sup>See Glaeser et al. (2010 [12]) for a model that attempts to correct the simple user cost model for some of these time-varying features. Mayer (2011 [27]) provides a discussion of the pros and cons of the user cost model and other possible alternative measures of mispricing for housing.

<sup>13</sup>Comparing house prices to variables like employment and income has no firm theoretical prediction; for example, failing to adjust to changes in economic fundamentals like interest rates and variable land supply across locations. Comparing house prices to construction costs only works in markets where land has very low value and thus is in abundant supply relative to demand. Even in locations with low land prices, house prices should still equal the present value of rents.

<sup>14</sup>See Hubbard and Mayer (2009 [20]), Table 2.

<sup>15</sup>In all of the specifications below, we repeat our analysis with both house price appreciation rate as well as the log IAR ratio and report both sets of coefficients. The findings are quite similar for both measures. As well, all of our results involving the log IAR ratio are robust to computing this measure with a variety of different assumptions about the expected future house price appreciation rate.



$i$	Log IAR Ratio in Percent							
	Mean	Sd	Min	Q25	Q50	Q75	Max	
Baltimore		14.4	13.4	-1.34	3.41	9.26	28.2	41.2
Charlotte		18.7	6.88	0.714	14.4	20.5	24.0	30.5
Cincinnati		3.78	5.16	-6.41	-0.278	3.97	8.45	14.5
Cleveland		9.21	6.62	-9.62	4.04	8.78	15.3	22.2
Denver		15.1	7.43	0.103	9.58	14.7	21.9	26.7
Jacksonville		27.0	11.1	11.0	19.8	22.0	35.9	52.2
Las Vegas		56.2	18.9	31.8	38.1	45.6	76.8	86.4
Los Angeles		30.3	14.2	10.0	17.5	28.2	43.7	57.2
Miami		35.7	21.3	7.59	18.0	27.7	58.4	73.3
Milwaukee		13.9	5.23	3.35	10.5	13.2	16.6	24.4
Minneapolis		29.5	5.14	17.9	26.1	29.8	33.0	40.4
Orlando		30.0	18.0	12.1	15.3	20.7	47.7	65.6
Philadelphia		15.0	11.6	0.802	4.84	12.2	26.8	36.1
Phoenix		30.2	16.5	6.80	19.3	22.1	48.6	65.8
Riverside		35.2	18.0	13.1	18.0	29.0	53.2	67.2
Sacramento		30.9	15.3	5.30	17.2	28.4	44.7	56.7
San Diego		32.7	9.53	15.7	24.1	33.1	40.2	50.5
San Francisco		40.3	11.1	17.4	30.5	40.6	48.8	61.2
San Jose		46.5	10.5	28.1	37.2	47.2	54.1	67.8
Tampa		37.8	16.1	15.4	25.2	33.2	52.8	70.4
Washington		16.0	16.2	-7.62	0.949	13.4	31.3	44.3
Mean		27.07	12.3					

**Table 6.** This table displays the log IAR ratio in each MSA  $i$  from month  $t$  to month  $t + 12$  in units of % over the time interval from Jan. 2000 to Dec. 2007. The shaded region in the sparkline graphs covers the interquartile range for each MSA and is not a constant scale.

we study the pricing implications in two alternative regimes. The first regime admits only fully informed traders while the second allows for misinformed traders as well. Finally, in Subsection 3.3, we use this economic model to discuss the challenges facing an econometrician trying to identify a speculative bubble and describe how our study of US residential housing addresses these challenges.

**3.1. Economic Framework.** Consider a static housing market with  $I \geq 1$  MSAs. The price of a house in MSA  $i$  is  $P_i$  and the true value of a house in MSA  $i$  is  $V_i$  where both  $P_i$  and  $V_i$  are measured as dollars per house. We model the true value of housing in each MSA  $i$  as an iid a random variable drawn from a normal distribution  $V_i \sim \mathcal{N}(\mu_v, \sigma_v^2)$ .



There are  $Q_i$  traders in each MSA  $i$ . Let  $S_{i \rightarrow j}$  denote the number of houses in MSA  $j$  which traders from MSA  $i$  demand and let  $\theta_{i \rightarrow j} = S_{i \rightarrow j}/Q_i$  denote the number of houses in MSA  $j$  demanded per trader in MSA  $i$ . We then model the total demand for housing in MSA  $j$  denoted  $X_j$  as the sum of the housing demand from each MSA  $i \in \{1, 2, \dots, I\}$  plus an MSA specific demand shock  $\varepsilon_j$ :

$$X_j = \sum_{i=1}^I S_{i \rightarrow j} + \varepsilon_j = \sum_{i=1}^I (Q_i \cdot \theta_{i \rightarrow j}) + \varepsilon_j \quad (4)$$

where  $\varepsilon_j$  is an iid draw from a normal distribution  $\varepsilon_j \sim \mathcal{N}(0, \sigma_\varepsilon^2)$  and  $X_j$  has units of houses. Thus, in terms of computing the aggregate demand, we can think about  $\theta_{i \rightarrow j}$  as either the number of houses in MSA  $j$  that each trader in MSA  $i$  demands or the probability that a randomly selected trader in MSA  $i$  owns a house in MSA  $j$ .

There is a collection of market makers who operate under perfect competition. These agents only observe the aggregate demand  $X_j$  in each MSA and as a result of perfect competition set the price level equal to the expected value of housing in MSA  $j$  given the realized aggregate demand:

$$P_j = \mathbf{E} [ V_j \mid X_j ] = \alpha + \beta \cdot X_j. \quad (5)$$

The coefficient  $\beta$  can be interpreted as the \$ change in the price of housing in MSA  $j$  when traders demand 1 additional unit of housing in MSA  $j$ . We can think about a market makers as developer or property managers who either build new housing units to match demand or reclaim unused housing units by turning them into rental properties or razing them to build office or industrial space.

Traders in each MSA  $i$  know the true value of housing in every other MSA  $j$ . For instance, in this view of the world a trader living in San Francisco that purchases a second house in Las Vegas knows the true value of housing in Las Vegas. The competitive market makers

assume<sup>16</sup> that traders use a linear demand rule given by:

$$\theta_{i \rightarrow j} = \gamma_{i \rightarrow j} + \delta_{i \rightarrow j} \cdot V_j. \quad (6)$$

The coefficient  $\gamma_{i \rightarrow j}$  has units of houses per trader and the coefficient  $\delta_{i \rightarrow j}$  has units of houses per trader dollar. Traders in each MSA  $i$  optimize their value function  $W_i$  by choosing how many houses to buy in each MSA  $j$ :

$$W_i = \sum_{j=1}^I W_{i \rightarrow j} \quad (7)$$

$$W_{i \rightarrow j} = \max_{\theta_{i \rightarrow j}} \mathbf{E} [ (V_j - P_j) \cdot \theta_{i \rightarrow j} \mid V_j ].$$

**Definition (Equilibrium).** *An equilibrium consists of price parameters  $(\alpha^*, \beta^*)$  and demand parameters  $\{(\gamma_{i \rightarrow j}^*, \delta_{i \rightarrow j}^*)\}_{i,j \in I}$  for each ordered MSA pair such that:*

- (1) *Given that the market makers follow the pricing rule in Equation (5), the housing demand schedule  $\{\theta_{i \rightarrow j}\}_{i,j \in I}$  dictated by the demand rule parameters  $\{(\gamma_{i \rightarrow j}^*, \delta_{i \rightarrow j}^*)\}_{i,j \in I}$  solves the traders' optimization problem in Equation (7).*
- (2) *Given that the traders follow the demand rule in Equation (6), the price parameters  $(\alpha^*, \beta^*)$  satisfy the expectations equality in Equation (5).*

**3.2. Equilibrium Housing Prices.** First, we solve for the equilibrium in this economy when all traders are fully informed. This equilibrium is identical to the standard Kyle (1985 [22]) equilibrium in all aspects except for the fact that each trader represents only  $1/\sum_{i'=1}^I Q_{i'}$  of the total market demand. Thus parameters defining the number of houses demanded per trader,  $\theta_{i \rightarrow j}$ , as well as the price impact of each trader's demand decisions,  $(\gamma_{i \rightarrow j}, \delta_{i \rightarrow j})$ , are both deflated by a factor of  $1/\sum_{i'=1}^I Q_{i'}$ .

**Proposition 1 (Fully Informed Equilibrium).** *When traders in all markets have correct beliefs about the true value of housing  $V_j$  in MSA  $j$ , traders in all MSAs demand the same*

---

<sup>16</sup>This assumption can easily be verified in equilibrium.

number of houses in MSA  $j$ ,

$$\bar{\theta}_j = \theta_{1 \rightarrow j} = \theta_{2 \rightarrow j} = \dots = \theta_{I \rightarrow j} \quad (8)$$

where  $\bar{\theta}_j$  is given by:

$$\bar{\theta}_j = \left( \frac{1}{\sum_{i'=1}^I Q_{i'}} \right) \cdot \left\{ -\frac{\alpha}{2 \cdot \beta} + \left( \frac{1}{2 \cdot \beta} \right) \cdot V_j \right\} \quad (9)$$

The equilibrium pricing rule is characterized by the constants:

$$\alpha^* = \mu_V, \quad \beta^* = \frac{\sigma_V}{2 \cdot \sigma_\varepsilon} \quad (10)$$

The price impact parameters are constant across ordered MSA pairs  $i \rightarrow j \in I \times I$  so that  $(\bar{\gamma}^*, \bar{\delta}^*) = (\gamma_{i \rightarrow j}^*, \delta_{i \rightarrow j}^*) = (\gamma_{i' \rightarrow j'}^*, \delta_{i' \rightarrow j'}^*)$  for all  $i \rightarrow j, i' \rightarrow j' \in I \times I$  where the equilibrium demand rules are:

$$\bar{\gamma}^* = - \left( \frac{1}{\sum_{i'=1}^I Q_{i'}} \right) \cdot \frac{\mu_V \cdot \sigma_\varepsilon}{\sigma_V}, \quad \bar{\delta}^* = \left( \frac{1}{\sum_{i'=1}^I Q_{i'}} \right) \cdot \frac{\sigma_\varepsilon}{\sigma_V} \quad (11)$$

*Proof.* See Appendix B. □

The key implication of this framework is that, in a world where all traders are fully informed, the proportion of traders from MSA  $i$  investing in MSA  $j$  is the same for each  $i = 1, 2, \dots, I$ . i.e., variation in the housing demand in MSA  $j$  per person in MSA  $i$  is proportional to variation in the value of housing in MSA  $j$  as fluctuations in  $V_j$  represent a common shock.

Next, we solve for an equilibrium when traders in some MSA  $i$  are misinformed about the value of housing in MSA  $j$ . Specifically, suppose that traders in MSA  $i$  believe that the value of a house in MSA  $j$  is  $\tilde{V}_j = V_j + \eta$  dollars with  $\eta > 0$  rather than the true value of  $V_j$  dollars. Let  $\tilde{P}_j^{(i)}$  denote the price of housing in MSA  $j$  when traders from MSA  $i$  have overconfident beliefs about  $V_j$ .

**Proposition 2** (Price Distortion). *Suppose that misinformed traders in MSA  $i$  believe that the value of housing in MSA  $j$  is  $\tilde{V}_j = V_j + \eta$  with  $\eta > 0$ . Then the price of a house in MSA*

$j$  will be distorted by an amount proportional to the number of traders in MSA  $i$ :

$$\tilde{P}_j^{(i)} - P_j = \left( \frac{Q_i}{\sum_{i'=1}^I Q_{i'}} \right) \cdot \frac{\eta}{2} \quad (12)$$

*Proof.* See Appendix B. □

This proposition is easiest to interpret via a short numerical example. Suppose that there are  $55 \times 10^6$  traders split across 10 MSAs with the largest MSA  $i'$  containing  $10 \times 10^6$  traders and the smallest MSA  $i''$  containing only  $1 \times 10^6$  traders. Then, the price increase in MSA  $j$  when traders from MSA  $i'$  or  $i''$  alternately believe that housing values in MSA  $j$  are  $\tilde{V}_j = V_j + \$5000$  are:

$$\tilde{P}_j^{(\text{MSA})} - P_j = \begin{cases} \left( \frac{10 \times 10^6 \mathbf{p}}{55 \times 10^6 \mathbf{p}} \right) \cdot \frac{\$5000}{2} = \$454.55 & \text{if MSA} = i' \\ \left( \frac{1 \times 10^6 \mathbf{p}}{55 \times 10^6 \mathbf{p}} \right) \cdot \frac{\$5000}{2} = \$45.45 & \text{if MSA} = i'' \end{cases} \quad (13)$$

In other words, when misinformed traders from a larger market attempt to purchase investment properties, they have a bigger impact on prices than misinformed traders from a smaller market.

**3.3. Microdata and Market Segmentation.** In order to better understand the empirical strategy below, we now map the empirical setting described in Section 2 onto this economic framework. Consider the three challenges outlined in the introduction.

First, we must identify a group of overconfident or uninformed speculators. Within the model, this task corresponds to identifying a group of traders who are likely to have misinformed beliefs about future price levels, i.e. an  $\eta > 0$ . In Section 4 we give a variety of pieces of evidence suggesting that out of town second house buyers satisfy this criteria. Thus, the transaction level deeds records available in the US residential housing market allow us to identify a group of potentially overconfident or uninformed speculators.

Second, we must show that an increase in demand from this group of misinformed speculators actually predicts increases in house price appreciation rates and log IAR ratios. Within the model, this task is tantamount to testing to see if homes appear overpriced (e.g.,

$P_j > E[P]$ ) when distant speculators have above average demand. In Section 5 we show that an increase in the number of out of town second house buyers predicts higher house price appreciation rates and log IAR ratios over the next year.

Finally, we must rule out the hypothesis that out of town second house buyers are not simply reacting to unobserved variation in housing values, which would likely attract distant speculators and might also increase house price appreciation rates and/or log IAR ratios. Within the model, this task corresponds to identifying whether a high realized price in MSA  $j$  was due to a high realized housing value  $V_j$  or to some group of traders in MSA  $i$  having misinformed beliefs  $\eta > 0$ . We exploit the natural geographic segmentation in the housing market to address this challenge. Proposition 1 demonstrates that if an increase in the price of housing in MSA  $j$  is due to an unobserved increase in house values, then out of town second house buyers from each other MSA should increase their demand for housing in MSA  $j$  in equal proportions. In Section 6 we test for this symmetry and show it to be violated. From this evidence, we conclude that out of town second house buyers are not simply responding to unobserved information when making their purchases.

However, this symmetry may be broken for a variety of reasons. In Proposition 2 we show that if out of town second house buyers from MSA  $i$  share a common belief distortion  $\eta$  about the value of housing in MSA  $j$ , then the size of the resulting price distortion should be proportional to the share of traders residing in MSA  $i$ . We find exactly this pattern in the data; the correlation between the log IAR ratio (or the rate of house price appreciation) and the share of distant speculators going from MSA  $i$  to MSA  $j$  is bigger when the total number of distant speculators living in MSA  $i$  is larger. We interpret these results as evidence that MSA specific variation in out of town second house buyer beliefs about MSA  $j$  (perhaps due to common news sources or word of mouth) contributing to the realized increases in house price appreciation rates and higher log IAR ratios.

#### 4. OVERCONFIDENT OR UNINFORMED SPECULATORS

In this section, we address the first empirical challenge and use data from transactions level deeds records to show that out of town second house buyers behaved like overconfident or uninformed speculators in the US residential housing market during the period from Jan. 2000 to Dec. 2007.

In Subsection 4.1 we give evidence suggesting that out of town second house buyers consume less of the housing dividend generated by their purchase and thus their returns likely depend more on future house price appreciation rates when making their trading decisions. Then, in Subsection 4.2, we show that, relative to local second house buyers and owner occupants, out of town second house buyers are likely less informed about local market conditions as these traders must travel and spend valuable time and energy learning about a new housing market before making their purchase. Supporting the claim that distant speculators are less informed, we show out of town second house buyers often earned negative returns on their second house purchases in MSAs such as Las Vegas, Phoenix and Miami, especially relative to local second house buyers who were better able to time the market.

**4.1. Dividend Consumption.** Out of town second house buyers may purchase houses for a number of reasons: e.g., a buyer might want to live in the house for part of the year, rent the property out as an additional source of income, or renovate the house and sell it for a profit at sometime in the future. In each of these instances, an out of town second house buyer gets lower dividends from the purchased house relative to a local second house buyer or owner occupant.

Part time residents can only consume the dividend (e.g., live in the home) for part of each year. Property maintenance is likely more expensive at a distance lowering the net present value of any stream of housing dividends an out of town second house buyer does consume. It is costly and difficult to supervise contractors or maintenance people from far away. What's more, we exclude most condominiums sales and locations that are primarily vacation markets from our data. As further evidence, in Figure 3 we plot the median value of the primary residences of out of town second house buyers and compare these prices to the

median in the MSA. We find that in the cities with the highest maximum earnings such as San Francisco, San Jose and New York, the median home price of out of town second house buyers is actually below that of the MSA as a whole. Thus, it is not true that the bulk of second house buyers we study are simply rich occupants in coastal cities that are deriving large utility gains from owning a vacation house in the Phoenix or a weekend getaway in Miami.

Out of town second house buyers who wish to rent out their purchased house must also contend with higher costs of property management due to the challenges of monitoring and managing a house from a distance. To give an idea of the order of magnitude of these costs, they typical property manager charges a fee of one months rent plus an additional 8% of the annual rent each year to lease a house and manage relations with the tenant. Direct costs to maintain and pay for repairs to appliances and the house itself are extra. In addition, any second house buyer wishing to rent out their property faces the prospect of higher physical depreciation costs as rental tenants may treat the house relatively poorly as compared to owner occupants.

Out of town second house buyers who plan on renovating a house and selling it for a profit (also known as “flipping” the house) do not live in the property and are thus almost entirely motivated by future capital gains. We find evidence suggesting that second house purchasers are much more likely to flip their houses relative to owner-occupants. In Figure 1 we plot the percent of single family house purchases each year from 2000 to 2007 which resell within 24 months of the original purchase split by occupancy type. We find that, for instance, in 2004, 35% of houses bought by local second house buyers in Las Vegas were resold within 24 months of the original purchase date. Figure 1 shows that both kinds of second house buyers are much more likely to resell within 24<sub>mo</sub>. These findings are broadly consistent with the results in Bayer et al. (2012 [3]), who suggest that house flipping in Los Angeles may have contributed to an increase in house price appreciation rates in that market.

Finally, both local and out of town second house buyers pay higher property taxes relative to owner occupants due to the different federal tax status of the imputed rent on second

houses and higher property tax rates imposed on second houses by many local communities. This evidence, when taken together, suggests that out of town second house buyers are more interested in the capital appreciation and thus treat housing more like a financial asset relative to local speculators or owner-occupants.

**4.2. Informational Disadvantage and Returns.** Out of town second house buyers are also likely less informed than either local second house buyers or owner occupants about location market conditions, and thus resemble uninformed or noise traders in the finance literature. Out of town second house buyers, by definition, live farther away than local second house buyers or owner occupants. Thus, these traders don't "know the neighborhood" and may be less aware of local market conditions.

In addition, out of town buyers face a difficult principal agent problem when dealing with local real estate agents who are paid on commission. Levitt and Syverson (2008 [24]) find that real estate agents have substantial discretion in the timing and price of house sales. Brokers receive about 3.7% more than other local owner occupants when selling their own homes suggesting that out of town second house buyers with higher monitoring costs likely face even more extreme agency costs.

As further evidence, we show that out of town second house buyers are also less successful in timing the market, especially when compared to local second house buyers. Figure 2 shows the average annualized internal rate of return on single family house purchases made by local and out of town second house buyers in MSA  $i$  in each month  $t$  in units of percent per year over the time interval from Jan. 2000 to Dec. 2007.<sup>17</sup> We compute this return by taking the weighted average of the annualized house price appreciation rates earned by all second house buyers who purchased a property in MSA  $i$  in month  $t$  and then resold it in month  $t + \tau$  for  $\tau \in [1, \bar{\tau}]$  where  $\bar{\tau}$  represents the number of months between Dec. 2007 and  $t$  where our data are right censored. We give assign agents who are right censored the house

---

<sup>17</sup>Of course the internal rate of return should include any rental revenues (or consumption benefits from owning a home) and also subtract costs. As noted above, we believe that distant speculators also face cost and revenue disadvantages relative to local speculators or owner occupants. Thus these IRR computations likely understate the poor returns for distant speculators relative to other buyers.



price appreciation rate from  $t$  to  $t + \bar{\tau}$ . The width of each line is scaled to represent the annualized number of second house purchases made by each type of buyer in MSA  $i$  in month  $t$  as a fraction of all properties in units of percent per year. In key markets such as Las Vegas and Phoenix, out of town second house buyers earned lower returns on their investments relative to local second house buyers. The width of the distant speculator line suggests that the majority of the out of town second house purchases in these markets yielded a negative return.

One potentially offsetting benefit to out of town second home purchasers is the possible diversification benefits from purchasing a second home in a market where housing returns are less correlated with other assets in the portfolio. Of course, this portfolio benefit might be mitigated to the extent that purchases of out of town housing generates lower-than-average returns. As well, for most home owners, purchases of stocks or bond might generate even more diversification with lower trading and holding costs, so diversification benefits from out of town housing purchases are likely limited.

Finally, if capital gains played a more critical role on the financial returns to out of town second house purchases, these types of investments might have attracted traders who were susceptible to overly exuberant expectations of house price appreciation rates. For instance, De Long et al. (1990 [7]) writes that “noise traders falsely believe that they have special information about the future price of the risky asset. . . in formulating their investment strategies, they may exhibit the fallacy of excessive subjective certainty. . .” Such excessively certain traders may actually seek out investments in fast appreciating markets.

## 5. PREDICTIVE REGRESSIONS

In this section, we address the second empirical challenge and show that an increase in the number of house purchases in an MSA made by out of town second house buyers predicts an increase in both house price appreciation rates as well as log IAR ratios. In Subsection 5.1 we look at the impact of additional out of town second house purchases in the current month on house price appreciation rates over the next year, while in Subsection 5.2 we look

at the impact on the log IAR ratio over the next year. We find that the number of out of town second house purchases as a percent of total sales in an MSA in month  $t$  is a strong positive predictor of house price appreciation rates and log IAR ratios over the next year.

**5.1. House Price Appreciation Rate Regressions.** We begin by estimating the relationship between the house price appreciation rate in an MSA  $i$  from month  $t$  to  $t + 12$  and the numbers of local and out of town second house purchases as a percent of sales in MSA  $i$  in month  $t$ . These regressions employ a panel data set indexed by MSA and month ranging over the 21 MSAs listed in Table 11 and the time period from Jan. 2001 to Dec. 2007 for a total of 1764 observations.

First, we estimate the regression specified in Equation (14) below:

$$\begin{aligned} \Delta \log P_{i,t \rightarrow (t+12)} = & \beta \cdot \Delta \log P_{i,(t-12) \rightarrow t} + \gamma \cdot \left( \frac{S_{i,t}^{\text{Distant}}}{X_{i,t}} \right) + \delta \cdot \left( \frac{S_{i,t}^{\text{Local}}}{X_{i,t}} \right) \\ & + \alpha_i + \eta_{t \rightarrow (t+12)} + \varepsilon_{i,t \rightarrow (t+12)} \end{aligned} \quad (14)$$

In this equation, we include both MSA specific and month/year specific fixed effects denoted by  $\alpha_i$  and  $\eta_{t \rightarrow (t+12)}$  respectively. The coefficients on distant and local speculator fractions are measured as annual changes in house prices. The MSA specific fixed effects account for across MSA variation in the mean house price appreciation rate during the sample period. We report the point estimates and standard errors for  $\beta$ ,  $\gamma$  and  $\delta$  in Table 7.

Dependent Variable: House Price Appreciation Rate				
	Estimate	Std. Error		
Lagged House Price Appreciation Rate	0.354	0.029	0.085	0.126
Distant Speculator Fraction	0.990	0.140	0.259	0.671
Local Speculator Fraction	-0.749	0.095	0.144	0.359
	Clustering	None	$t$	$i$
	$N$	1764		
	$R^2$	0.146		

**Table 7.** This table presents the coefficient estimates from Equation (14) using monthly data on the 21 MSAs listed in Table 11 from Jan. 2001 to Dec. 2007. The regression includes month/year and MSA fixed effects. Standard errors are estimated three different ways to account for clustering over time or across MSAs.

We find that a 1%/yr increase in the house price appreciation rate in an MSA over the period from month  $t - 12$  to month  $t$  is associated with a 0.354%/yr increase in the house price appreciation rate over the period from month  $t$  to month  $t + 12$  indicating some positive autocorrelation in prices. In addition, we again find that increases in the number of second house purchases predicts future an increase in the house price appreciation rate with a 1% increase in the number of out of town second house purchases as a fraction of all purchases in an MSA  $i$  in month  $t$  associated with a 0.990%/yr increase in the house price appreciation rate over the next year. However, we find the exact opposite relationship when looking at local second house purchases. Here, we find that a 1% increase in the local speculator share in an MSA  $i$  in month  $t$  is associated with a 0.749%/yr decline in the house price appreciation rate over the next year. Thus, if anything, an increase in the number of local speculator purchases appears to precede price declines.

**5.2. Log IAR Ratio Regressions.** Next, we estimate the relationship between the log IAR ratio in an MSA  $i$  in month  $t$  and the percent of home sales made to local and out of town second house buyers in that MSA in month  $t$ . These regressions employ the same panel data set indexed by MSA and month ranging over 21 MSAs during the time period from Jan. 2001 to Dec. 2007 for a total of 1764 observations.

$$\log Z_{i,t} = \beta \cdot \log Z_{i,t-12} + \gamma \cdot \left( \frac{S_{i,t}^{\text{Distant}}}{X_{i,t}} \right) + \delta \cdot \left( \frac{S_{i,t}^{\text{Local}}}{X_{i,t}} \right) + \alpha_i + \eta_t + \varepsilon_{i,t} \quad (15)$$

Confirming the findings in the subsection above, we find that the number of distant speculator purchases in an MSA is a strong predictor of future log IAR ratios. A 1% increase in the number of local second house purchases as a fraction of all sales in an MSA  $i$  in month  $t$  is associated with only a 0.510% increase in the log IAR ratio from in MSA  $i$  from month  $t$  to month  $t + 12$ . By contrast, we find that the number of out of town second house purchases as a fraction of all sales in an MSA  $i$  at time  $t$  is an even stronger predictor of future log IAR ratios than the current IAR ratio.

Dependent Variable: Log IAR Ratio				
	Estimate	Std. Error		
Lagged Log IAR Ratio	0.780	0.013	0.022	0.038
Distant Speculator Fraction	1.43	0.094	0.085	0.276
Local Speculator Fraction	0.510	0.066	0.096	0.259
	Clustering	None	$t$	$i$
	$N$	1764		
	$R^2$	0.701		

**Table 8.** *This table presents the coefficient estimates from Equation (15) using monthly data on the 21 MSAs listed in Table 11 from Jan. 2001 to Dec. 2007. Fixed effect estimates are omitted for clarity. Standard errors are estimated three different ways to account for clustering over time or across MSAs.*

## 6. CAUSAL ANALYSIS

In this section, we address the third empirical challenge. We use the geographic segmentation of the US residential real estate market to show that out of town second house buyers are not simply responding to an unobserved increase in the value of housing in MSAs like Las Vegas and Miami, but appear to directly contribute to an increase in the house price appreciation rates and log IAR ratios in these MSAs.

Our goal is to distinguish between two alternative hypotheses: the null hypothesis that distant speculators chose to purchase second houses in MSAs due to unobserved fluctuations in housing values and the alternative hypothesis that an increase in distant speculator purchases caused an increase in house price appreciation rates and log IAR ratios. Our key observation is that, if the null hypothesis is true and housing value fluctuations drove the extreme price appreciation in MSAs like Las Vegas, then out of town second house buyers from each MSA should have increased their demand for housing in Las Vegas in equal proportions since variations in Las Vegas housing values represent a common shock to all distant speculators after controlling for MSA pair specific factors such as distance and information transmission. This intuition is spelled out in Proposition 1. On the other hand, if the null hypothesis is not true then this proportional symmetry should be violated.

However, this symmetry can be violated in a variety of ways and only some of these violations are consistent with the alternative hypothesis. If an increase in demand from out

of town second house buyers caused an increase in the house price appreciation rates and log IAR ratios in Las Vegas, an increase in the demand per trader for second houses in MSA Las Vegas by out of town second house buyers living in MSA  $i$  should have the largest effect when there are relatively more second house buyers in MSA  $i$  as dictated in Proposition 2.

We start in Subsection 6.1 by describing how we empirically implement the model outlined in Section 3. We frame the predictions of this model as a pair of estimable equations. In Subsection 6.2 we estimate these equations. We find evidence consistent with the hypothesis that an increase in distant speculators in an MSA causes an increase in that MSA's house price appreciation rate and log IAR ratio over the subsequent year. Finally, in Subsection 6.3 we consider alternative explanations.

**6.1. Empirical Predictions.** We need to estimate the total number of distant speculators in each MSA  $i$  and the demand per distant speculator in MSA  $i$  for second houses in each other MSA  $j$  at time  $t$ . We begin by defining the number of speculators in each MSA  $i$  corresponding to the variable  $Q_i$  in Section 3.

**Definition** (Number of Speculators). *Let  $Q_i$  denote the number of distant speculators in MSA  $i$  measured as the average annualized number of second house purchases made by buyers living in MSA  $i$  each month over the period from Jan. 2000 to Dec. 2007 so that  $T = 96$ :*

$$Q_i = \frac{1}{96} \cdot \sum_{t=1}^{96} \left( \sum_{i \neq j} S_{i \rightarrow j, t} \right) \quad (16)$$

*Let  $\hat{Q}_i$  denote the number of distant speculators in MSA  $i$  similarly defined but measured over the period from Jan. 2000 to Dec. 2001 so that  $T = 12$ :*

$$\hat{Q}_i = \frac{1}{12} \cdot \sum_{t=1}^{12} \left( \sum_{i \neq j} S_{i \rightarrow j, t} \right) \quad (17)$$

The first definition of the number of distant speculators in each MSA  $i$  represents the sample average over the entire period from Jan. 2000 to Dec. 2007. Since this variable is computed using the entire time series, it is potentially simultaneously determined with investment opportunities in the largest markets for distant speculators that appear attractive

later in the sample period. (e.g., Some distant speculators might only have entered the housing market because MSAs like Las Vegas and Phoenix appeared to have had great investment opportunities.) This observation motivates the use of the second definition that includes only data from the year 2000, which is otherwise omitted from the regressions below due to 12<sub>mo</sub> lagged values of other variables in the analysis.

Next, we define time varying demand per distant speculator in MSA  $i$  for second houses in MSA  $j$  at month  $t$  corresponding to the variable  $\theta_{i \rightarrow j}$  in Section 3.

**Definition** (Speculator Share). *Let  $\theta_{i \rightarrow j,t}$  and  $\hat{\theta}_{i \rightarrow j,t}$  denote the demand for houses in MSA  $j$  at time  $t$  by buyers in MSA  $i$  as a fraction of the number of second house buyers in MSA  $i$ :*

$$\theta_{i \rightarrow j,t} = \frac{S_{i \rightarrow j,t}}{Q_i}, \quad \hat{\theta}_{i \rightarrow j,t} = \frac{S_{i \rightarrow j,t}}{\hat{Q}_i} \quad (18)$$

Here, both  $\theta_{i \rightarrow j,t}$  and  $\hat{\theta}_{i \rightarrow j,t}$  are measured as number of houses per trader.

Unless explicitly stated, we use  $\theta_{i \rightarrow j,t}$  to denote both  $\theta_{i \rightarrow j,t}$  and the alternative variable  $\hat{\theta}_{i \rightarrow j,t}$  in the text below for brevity.

We estimate all regression equations in this section using a panel dataset at a monthly frequency from Jan. 2001 to Dec. 2007 on the  $21 \times 20 = 420$  possible ordered MSA pairs generated from Table 11 with all  $i = j$  pairs removed. Observations from Jan. 2000 to Dec. 2000 are removed due to the missing 12<sub>mo</sub> lag values yielding a balanced panel of 35,280 observations.

We begin by estimating Equations (19) and (22) which study the relationship between the house price appreciation rate from time  $t$  to time  $t + 12$  (or the log IAR ratio at time  $t$  in MSA  $j$ ) and the proportion of second house buyers in each MSA  $i$  that purchase an out of town second house in MSA  $j$  at time  $t$  represented by the coefficient  $\gamma$  on the variable  $\theta_{i \rightarrow j,t}$ .

$$\Delta \log P_{j,t \rightarrow (t+12)} = \beta \cdot \Delta \log P_{j,(t-12) \rightarrow t} + \gamma \cdot \theta_{i \rightarrow j,t} + \alpha_{i \rightarrow j} + \eta_{t \rightarrow (t+12)} + \varepsilon_{i \rightarrow j,t \rightarrow (t+12)}, \quad i \neq j \quad (19)$$

The ordered MSA pair dummy variables control for two key effects as displayed in Equation (20) below.

$$\alpha_{i \rightarrow j} = \bar{\alpha}_j - \gamma \cdot \mathbf{E}[\theta_{i \rightarrow j, t}] \quad (20)$$

First, each  $\alpha_{i \rightarrow j}$  accounts for the mean house price appreciation rate  $\bar{\alpha}_j$  in each MSA  $j$  over the sample period from Jan. 2001 to Dec. 2007 (or the mean IAR ratio over the same time period). The  $\bar{\alpha}_j$  component will be present in each of the  $\alpha_{i \rightarrow j}$  for  $i \in \{I \setminus j\}$ . Second, each  $\alpha_{i \rightarrow j}$  adjusts the predicted house price appreciation rate (or IAR ratio) in MSA  $j$  for the average rate at which second house buyers living in MSA  $i$  purchase second houses in MSA  $j$ . For instance,  $\gamma \cdot \mathbf{E}[\theta_{(\text{SFO}, j), t}]$  differentially controls for the tendency of San Francisco to purchase more second houses in Phoenix than in Milwaukee. For the same reason, this second component of the fixed effect controls for the tendency of San Francisco residents to purchase more second houses in Phoenix than vice versa:

$$\mathbf{E}[\theta_{(\text{SFO}, \text{PHX}), t}] \neq \mathbf{E}[\theta_{(\text{PHX}, \text{SFO}), t}] \quad (21)$$

We also estimate the specification outlined in Equation (22), below, which replaces the house price appreciation rate in MSA  $j$  from time  $t$  to time  $t + 12$  with the log IAR ratio at time  $t$ .

$$\log Z_{j, t} = \beta \cdot \log Z_{j, t-12} + \gamma \cdot \theta_{i \rightarrow j, t} + \alpha_{i \rightarrow j} + \eta_t + \varepsilon_{i \rightarrow j, t}, \quad i \neq j \quad (22)$$

Consistent with the results in Section 5, we expect to estimate a positive  $\gamma$  for both specifications indicating that, for instance, log IAR ratios rise by  $\gamma\%$  in MSA  $j$  when the proportion of second house buyers in MSA  $i \neq j$  that invest in MSA  $j$  increases by 1%.

Next, we augment this baseline specification in order to investigate whether or not the null hypothesis that second house buyers in all MSAs  $i \in \{I \setminus j\}$  proportionally increase their demand for houses in MSA  $j$  after appropriate controls is true. We do this by including an interaction between the number of second house buyers in MSA  $i$ ,  $Q_i$ , and the the proportion of these speculators buying houses in MSA  $j$ ,  $\theta_{i \rightarrow j, t}$ . Specifically, we define the three indicator

variables below which divide the set of 21 MSAs in our sample into terciles based on the number of second house buyers where  $1_{\{\text{Small MSA}\}}$  denotes one of the seven MSAs with the smallest number of distant speculators,  $1_{\{\text{Medium MSA}\}}$  denotes the next seven MSAs with a moderate number of distant speculators and  $1_{\{\text{Large MSA}\}}$  denotes one of the seven MSAs with a largest number of distant speculators.

We then estimate the two regression specifications, below, where  $\delta_2$  and  $\delta_3$  have units of houses per person per year in Equation (23) and units of houses per person in Equation (24).

$$\begin{aligned} \Delta \log P_{j,t \rightarrow (t+12)} &= \beta \cdot \Delta \log P_{j,(t-12) \rightarrow t} \\ &+ \gamma \cdot \theta_{i \rightarrow j,t} + \delta_2 \cdot 1_{\{\text{Medium MSA}\}} \cdot \theta_{i \rightarrow j,t} + \delta_3 \cdot 1_{\{\text{Large MSA}\}} \cdot \theta_{i \rightarrow j,t} \end{aligned} \quad (23)$$

$$+ \alpha_{i \rightarrow j} + \eta_{t \rightarrow (t+12)} + \varepsilon_{i \rightarrow j,t \rightarrow (t+12)}, \quad i \neq j$$

$$\log Z_{j,t} = \beta \cdot \log Z_{j,t-12}$$

$$+ \gamma \cdot \theta_{i \rightarrow j,t} + \delta_2 \cdot 1_{\{\text{Medium MSA}\}} \cdot \theta_{i \rightarrow j,t} + \delta_3 \cdot 1_{\{\text{Large MSA}\}} \cdot \theta_{i \rightarrow j,t} \quad (24)$$

$$+ \alpha_{i \rightarrow j} + \eta_t + \varepsilon_{i \rightarrow j,t}, \quad i \neq j$$

If the null hypothesis is true, we should find  $\delta_2 = \delta_3 = 0$ . i.e., a 1% increase in the demand per trader living in San Francisco for second houses in Phoenix should be equally predictive of an increase in house price appreciation rates in Phoenix as a 1% increase in the demand per trader from Denver for Phoenix housing. We can reject the null hypothesis that out of town second house buyers in both San Francisco and Denver are responding to the same unobservable value increase in Phoenix housing if  $\delta_2, \delta_3 \neq 0$ . Note that in Equations (23) and (24), the ordered MSA pair fixed effects have the decomposition  $\alpha_{i \rightarrow j} = \bar{\alpha}_j - \gamma \cdot \mathbf{E}[\theta_{i \rightarrow j,t}] - \delta_2 \cdot 1_{\{\text{Medium MSA}\}} \cdot \mathbf{E}[\theta_{i \rightarrow j,t}] - \delta_3 \cdot 1_{\{\text{Large MSA}\}} \cdot \mathbf{E}[\theta_{i \rightarrow j,t}]$ . Thus, these coefficients now capture the mean house price appreciation rates and log IAR ratios in MSA  $j$  as well as the full effect of the average number of houses demanded in MSA  $j$  by speculators living in MSA  $i$ .

Proposition 2 indicates that in order to confirm the alternative hypothesis that out of town second house buyers are causing increases in the house price appreciation rates and log IAR ratios, we should find  $\delta_3 > \delta_2 > 0$ . i.e., that house price appreciation rates and log IAR



ratios are the highest in MSA  $j$  in the 12 months following an increase in the demand per speculator in MSA  $i$  when MSA  $i$  contains the largest number of potential traders.<sup>18</sup>

**6.2. Estimation Results.** We report the estimated coefficients and standard errors from Equations (19) and (22) in Panel (a) of Tables 9 and 10. Panels (b) and (c) in Tables 9 and 10 house the estimated values and standard errors from Equations (23) and (24), which interact the share of distant speculators living in MSA  $i$  and purchasing second houses in MSA  $j$  with the number of distant speculators living in the “home” market of MSA  $i$ .

Panel (a) in both Table 9 and 10 indicate that  $\gamma$  is both positive and statistically significant. The point estimate for  $\gamma$  in Table 9 can be interpreted as saying that a 1% increase in the number of houses demanded in MSA  $j$  per trader in MSA  $i$  each year predicts a 0.094%/yr increase in the house price appreciation rate in MSA  $j$  over the next year. Similarly in Table 10 a 1% increase in the number of houses in MSA  $j$  demanded per distant speculator in MSA  $i$  results in a 0.110% increase in the log IAR ratio, suggesting mispricing grows when distant speculator demand grows.

Next, looking at Panel (b) in both Tables 9 and 10 we see that  $\delta_3 \neq 0$  in violation of the symmetry predicted by Proposition 1. The coefficients on the interaction between the MSAs with a large distant speculator population have a price impact that is almost twice as large as the price impact of investors coming from the smallest MSAs.

As well, the ordering of the interaction terms is consistent with the alternative hypothesis that demand from distant speculators causes house price appreciation rates and log IAR ratios to increase. In all specifications  $\delta_3 \geq \delta_2 \geq 0$ . We can interpret the coefficients  $\gamma$ ,  $\delta_2$  and  $\delta_3$  reported in Panel (b) of Table 9 as saying that while a 1% increase in the number of houses demanded in MSA  $j$  per trader in MSA  $i$  each year predicts an 0.046%/yr increase in the house price appreciation rate in MSA  $j$  over the next year when there are relatively few speculators in MSA  $i$ , that same 1% increase in houses demanded per trader is associated with a 0.121%/yr increase in the house price appreciation rate in MSA  $j$  over the next year when there are a relatively large number speculators in MSA  $i$ .

---

<sup>18</sup>This identification strategy is analogous to the front door criterion as outlined in Pearl (2000 [33]).

Dependent Variable: House Price Appreciation Rate

Panel (a), Ranking Period: Jan. 2000 to Dec. 2007

	Estimate	Std. Error		
Lagged House Price Appreciation Rate	0.353	0.006	0.075	0.026
Distant Speculator Share	0.094	0.005	0.013	0.015
	Clustering	None	$t$	$i \rightarrow j$
$N$	35280			
$R^2$	0.111			

Panel (b), Ranking Period: Jan. 2000 to Dec. 2007

	Estimate	Std. Error		
Lagged House Price Appreciation Rate	0.350	0.006	0.075	0.026
Distant Speculator Share	0.046	0.008	0.014	0.013
Medium MSA $\times$ Dist. Speculator Share	0.074	0.011	0.019	0.022
Large MSA $\times$ Dist. Speculator Share	0.075	0.011	0.018	0.038
	Clustering	None	$t$	$i \rightarrow j$
$N$	35280			
$R^2$	0.114			

Panel (c), Ranking Period: Jan. 2000 to Dec. 2000

	Estimate	Std. Error		
Lagged House Price Appreciation Rate	0.351	0.006	0.075	0.026
Distant Speculator Share	0.032	0.003	0.005	0.007
Medium MSA $\times$ Dist. Speculator Share	-0.013	0.004	0.003	0.012
Large MSA $\times$ Dist. Speculator Share	0.045	0.006	0.007	0.022
	Clustering	None	$t$	$i \rightarrow j$
$N$	35280			
$R^2$	0.113			

**Table 9.** Panel (a): Coefficient estimates from Equation (19). Panel (b): Coefficient estimates from Equation (23) using the full sample estimate of  $Q_i$ . Panel (c): Coefficient estimates from Equation (23) using the pre sample estimate of  $\hat{Q}_i$ . All regressions use monthly data from Jan. 2001 to Dec. 2007 on the 420 possible ordered MSA pairs generated from Table 11 with all  $i = j$  pairs removed. Fixed effect estimates of  $\alpha_{i \rightarrow j}$  and  $\eta_{t \rightarrow (t+12)}$  are omitted for clarity. Standard errors are estimated three different ways to account for clustering over time or across MSA pairs.

Panel (c) in both tables reports the results of measuring the number of distant speculators in each MSA using the beginning of ranking in 2000. This specification controls for any possible simultaneity between how we measure the number of distant speculators and subsequent investment opportunities. These results are a bit less robust than those in panel (b) with the interaction terms having smaller coefficients, but present a consistent story. In all cases, the coefficients on the interaction of distant speculator share and large MSAs is statistically different from zero no matter which clustering of standard errors we use. In

Dependent Variable: Log IAR Ratio

Panel (a), Ranking Period: Jan. 2000 to Dec. 2007

	Estimate	Std. Error		
Lagged Log IAR Ratio	0.770	0.003	0.024	0.008
Distant Speculator Share	0.110	0.003	0.008	0.016
	Clustering	None	$t$	$i \rightarrow j$
$N$	35280			
$R^2$	0.634			

Panel (b), Ranking Period: Jan. 2000 to Dec. 2007

	Estimate	Std. Error		
Lagged Log IAR Ratio	0.772	0.003	0.024	0.008
Distant Speculator Share	0.064	0.006	0.008	0.015
Medium MSA $\times$ Dist. Speculator Share	0.021	0.008	0.014	0.029
Large MSA $\times$ Dist. Speculator Share	0.124	0.008	0.009	0.029
	Clustering	None	$t$	$i \rightarrow j$
$N$	35280			
$R^2$	0.637			

Panel (c), Ranking Period: Jan. 2000 to Dec. 2000

	Estimate	Std. Error		
Lagged Log IAR Ratio	0.771	0.003	0.024	0.008
Distant Speculator Share	0.026	0.002	0.004	0.007
Medium MSA $\times$ Dist. Speculator Share	0.012	0.003	0.004	0.008
Large MSA $\times$ Dist. Speculator Share	0.067	0.004	0.005	0.021
	Clustering	None	$t$	$i \rightarrow j$
$N$	35280			
$R^2$	0.636			

**Table 10.** Panel (a): Coefficient estimates from Equation (22). Panel (b): Coefficient estimates from Equation (24) using the full sample estimate of  $Q_i$ . Panel (c): Coefficient estimates from Equation (24) using the pre sample estimate of  $\hat{Q}_i$ . All regressions use monthly data from Jan. 2001 to Dec. 2007 on the 420 possible ordered MSA pairs generated from Table 11 with all  $i = j$  pairs removed. Fixed effect estimates of  $\alpha_{i \rightarrow j}$  and  $\eta_t$  are omitted for clarity. Standard errors are estimated three different ways to account for clustering over time or across MSA pairs.

Table 9, the coefficient on the interaction with medium size cities is negative, but is not statistically different from zero when we cluster by ordered city pair  $i \rightarrow j$ .

Finally, we observe that the empirical results are strongest in Table 10 where we use the IAR ratio as the dependent variable. In all cases the magnitude of the coefficients on the interaction terms with MSA size are increasing and all coefficients are different from zero. To the extent that the log IAR ratio proxies for mispricing, these results present a consistent picture that distant speculators contribute to mispricing.

**6.3. Alternative Explanation: Subprime Lending.** Most analysts point to the growth of subprime lending and lax credit standards in explaining the spike in house price appreciation in the US during the mid 2000s.<sup>19</sup> From a theoretical perspective, Stein (1995 [38]) shows how the credit supply channel can amplify price shocks leading to sudden jumps in prices. If easier credit from the subprime crisis served as an aggregate shock that increased home prices across the country, it would be adequately controlled for by the inclusion of time fixed effects in our earlier specification. However, easier credit might have also amplified the demand shock by impact distant speculators. For example, Haughwout et al. (2011 [17]) and Gao and Li (2012 [25]) document a strong growth in mortgages to second home buyers during the boom.

To get a preliminary sense of the impact of subprime lending on investor purchases, we examine the impact of leverage on local and distant second home buyers relative to owner occupants in Figure 4. This figure plots the average cumulative loan-to-value ratio (C-LTV) for the three groups of buyers over time and across MSAs.<sup>20</sup> Of particular interest is the fact that owner occupants always appear to use more leverage than local or distant speculators. Of course, these data do not allow us to determine whether second home buyer demand for leverage was lower than owner occupants or whether lenders were unwilling to provide comparable leverage to second home buyers.

Differences in the mean amount of leverage between the various groups of investors cannot explain our findings. Of greater relevance, it does not appear that distant speculators started using leverage relative to owner occupants or local investors at the peak of the bubble in boom markets. While distant speculator leverage rose sharply Tampa and Miami during the boom, it rose even faster in less overpriced markets like Charlotte and Philadelphia. Distant speculators in Las Vegas and Phoenix appear to have increased leverage a little over the time period, but the slow increase in leverage cannot explain the sharp increase in the share of

---

<sup>19</sup>See Mian and Sufi (2008 [29]; 2009 [30]) for evidence of the impact of leverage on housing demand. Mayer et al. (2009 [28]) and Demyanyk and Van Hemert (2011 [8]) show that subprime lending standards and downpayment requirements fell rapidly up to 2007. By contrast, Glaeser et al. (2011 [12]) express skepticism that declining lending standards played an important role in the housing bubble.

<sup>20</sup>Mortgage data are available for 20 of the 21 MSAs, with missing data for Minneapolis.

purchases to distant speculators over the same period. These results are similar to those of Lee et al. (2012 [23]) who show that second lien usage was less pronounced for second home buyers than owner occupants.

Finally, in future work we will examine the role of leverage and capital gains on the primary residence in fueling demand by distant speculators. We hope to examine whether distant speculators used higher leverage or second liens taken out on their primary residence to finance purchases of a second home. Figure 3 suggests that at least in the most expensive cities like San Francisco and Los Angeles, out of town second home buyers had smaller primary residences than average, suggesting this was not just buyers of the most expensive homes who were distant speculators. However, more work is necessary to address the issue of leverage on the primary residence. The analysis of the location of the first home versus the second home might also allow us to determine the extent to which out of town home purchases provided much portfolio diversification. Finally, we can examine other, more detailed, hypotheses about how distant speculators learned about investment opportunities. For example, we can examine a link between local media reports (e.g., a newspaper article in the *Chicago Tribune* on investment opportunities in Miami) and the timing of subsequent purchases. We can also examine whether purchases are made from concentrated neighborhoods in the "home" market and whether different types of purchasers achieve relatively better (or worse) returns.

## 7. CONCLUSION

Analyzing the asset pricing implications of speculative trading using data from the stock market is difficult because traders are anonymous and there is no natural market segmentation. In response to these difficulties, we analyze the price impact of speculative demand in the US residential housing market where we obtain detailed data on traders and the market is geographically segmented.

We show that out of town second house buyers (who we refer to as "distant speculators") behave like overconfident or uninformed speculators. These purchasers are less able to consume the dividend from their housing purchase and less likely to be informed about local

market conditions when compared to local second house buyers or owner occupants. We then show that an increase in the number of purchases by distant speculators as a fraction of total sales in an MSA predicts an increase in house price appreciation rates and higher log IAR ratios in the next year. We examine the issue of reverse causality and find that these distant speculators are unlikely to be responding to unobserved fluctuations in the value of housing. Rather our evidence is consistent with the hypothesis that demand from out of town second house buyers caused house price appreciation rates and log IAR ratios to rise.

We conclude by discussing some of the broader implications of our findings. First, we consider the impact that fluctuations in house prices caused by distant speculator demand might have on the real economy. To get a sense of the order of magnitude of the real effects of purchases by distant speculators relative to the size of the local economy, we examine the size of cash transfers from one MSA to the next as a result of out of town second house buyers. Figure 5 shows the total down payments made by distant speculators on single family house purchases in each MSA in units of billions of dollars per year in every month from Jan. 2000 to Dec. 2007. This figure shows, for example, that the sum of all down payments in Las Vegas by second house buyers living in another MSA exceeded \$1.25<sub>B11</sub> in 2004. Overall, billions of dollars moved across MSAs in the boom just for downpayments alone, with the largest transfers in some of the fastest appreciating markets.

Next we attempt to examine how total out of town purchases compare to the size of the local economy. Figure 6 plots the sum of the sales prices on distant speculator purchases as a percent of gross MSA product,  $G(\text{MSA})P$  from 2000 to 2007, where  $G(\text{MSA})P$  is reported by the Bureau of Economic Analysis. These calculations treat all purchases as being net new capital coming from outside the MSA, whether financed by debt or equity. This figure shows that the sum of the sales prices in Las Vegas exceeded 5% of the  $G(\text{MSA})P$  for the entire MSA in 2004. Thus demand shocks from distant speculators appear to be quite substantial when compared to the aggregate economic output of many MSA level economies, especially if such purchases resulted in more homes being built than might otherwise have been constructed.

Barro and Ursua (2008 [2]) define a 10% drop in the GDP of a country as an economic disaster while Javorcik (2004 [21]) examines firm level data in Lithuania and finds that foreign direct investment from the US on the order of 3.4% of the Lithuanian GDP in 2000 leads to substantial spillover effects in its real economy. We see an opportunity in future work to study the impact of these spillovers on local economies.

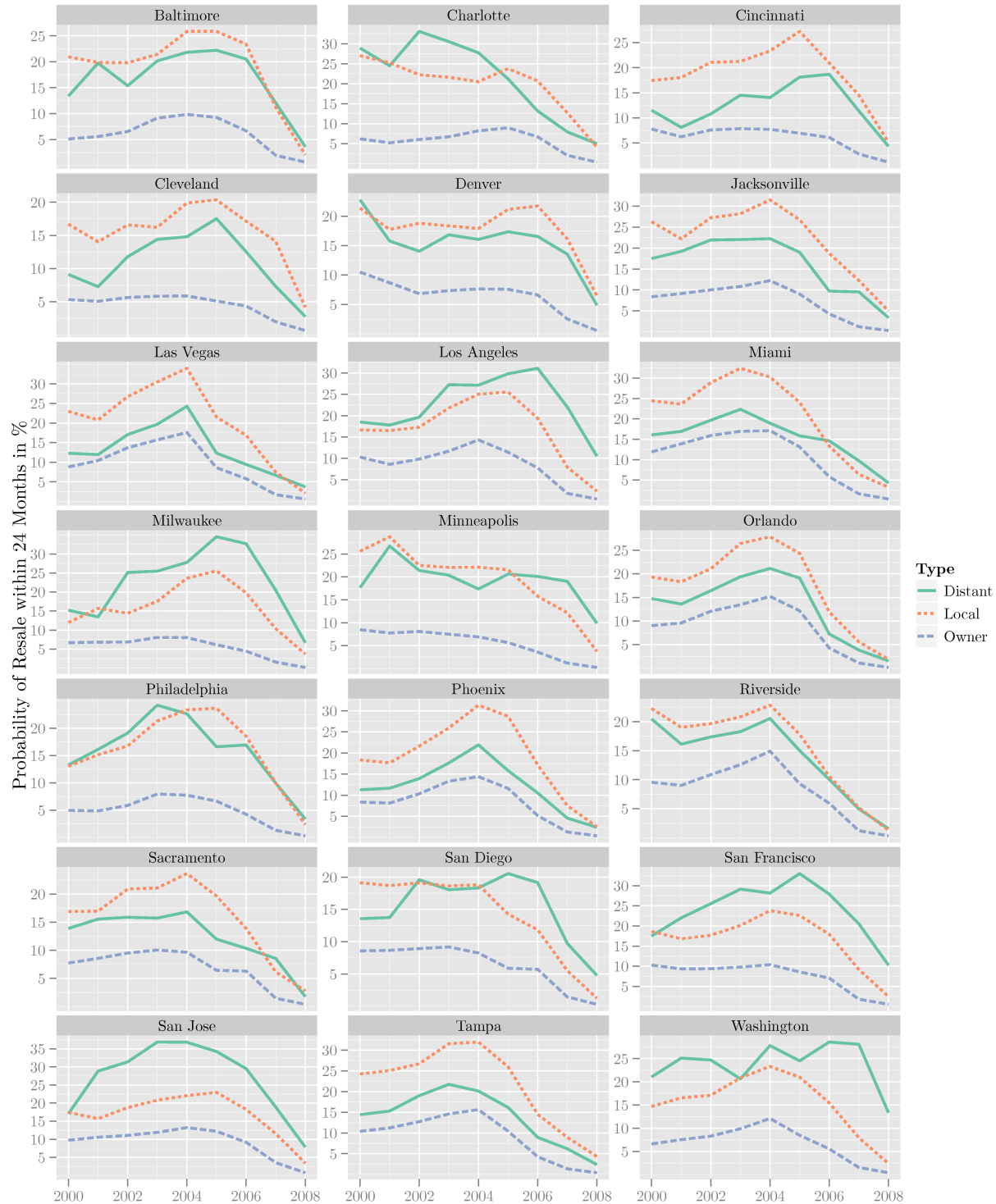
We conjecture that distant speculator demand driven bubbles may not be a phenomenon confined to the US residential real estate market. For instance, a 2009 Office for National Statistics<sup>21</sup> report found that 1.8<sub>M</sub> households in England owned a second home and, among these properties, 87<sub>k</sub> were in Spain and being used as part time residences during the peak of the Spanish housing boom. To give some idea of the scale of this investment expenditure by overseas second home buyers in Spain, in Figure 7 we plot the net foreign direct investment (henceforth, FDI) in Spain as a percent of Spain's GDP from 2003 to 2010 using data from the World Bank alongside the real HPI level in Spain over this same time period. We find that FDI as a percent of GDP spikes to just under 5% in 2008, a similar percentage to the total of outside purchases of homes in Las Vegas at peak, and that the timing of this spike corresponds to the peak of the HPI level. Data do not show a similar peak in FDI in other southern European counties.

A similar phenomenon occurred in the US commercial real estate market in the late 1980s when a 1986 tax code change made purchases of commercial real estate less attractive for US companies and invited a host of foreign investors from countries like Japan to large scale purchases of commercial office buildings.<sup>22</sup> Looking to examples outside the real estate domain, many authors such as Griffin et al. (2010 [15]) have suggested that an influx of day traders drove some of the price appreciation in technology stocks during the Dot-Com boom. Thus, distant speculators may be an important class of traders playing a role in bubble formation more generally.

---

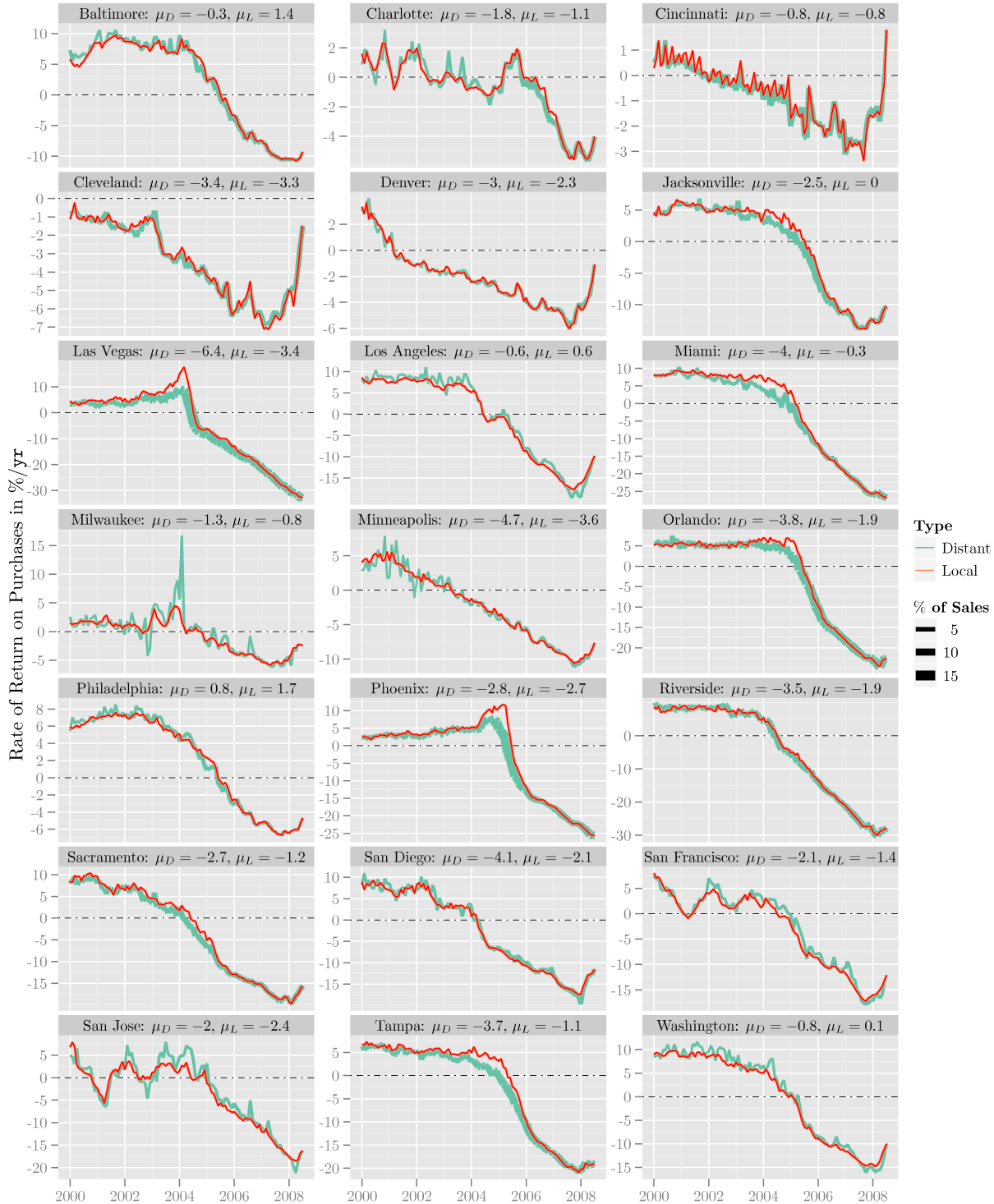
<sup>21</sup>See English Housing Survey (2009 [32]).

<sup>22</sup>See Sagalyn (1999 [35]), which discuss the purchase of Rockefeller Center by MITSUBISHI TRUST, CO. for more than \$1<sub>B</sub> in the late 1980.



**Figure 1.** *The percent of single family house purchases each year from 2000 to 2007 which resell within 24 months of the original purchase split by occupancy type. Reads: “In 2004, 35% of houses bought by local second house buyers in Las Vegas were resold within 24 months of the original purchase date.”*

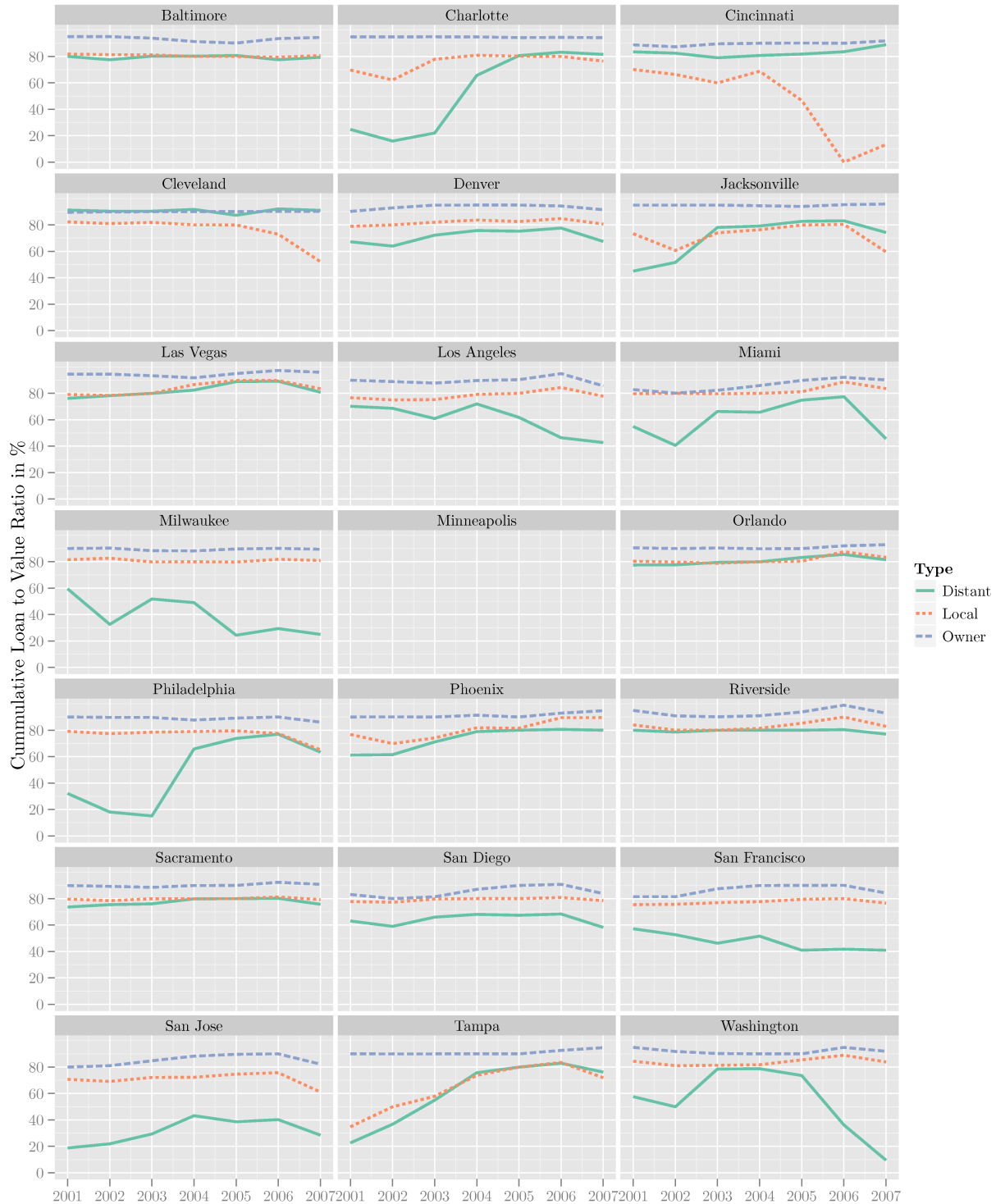




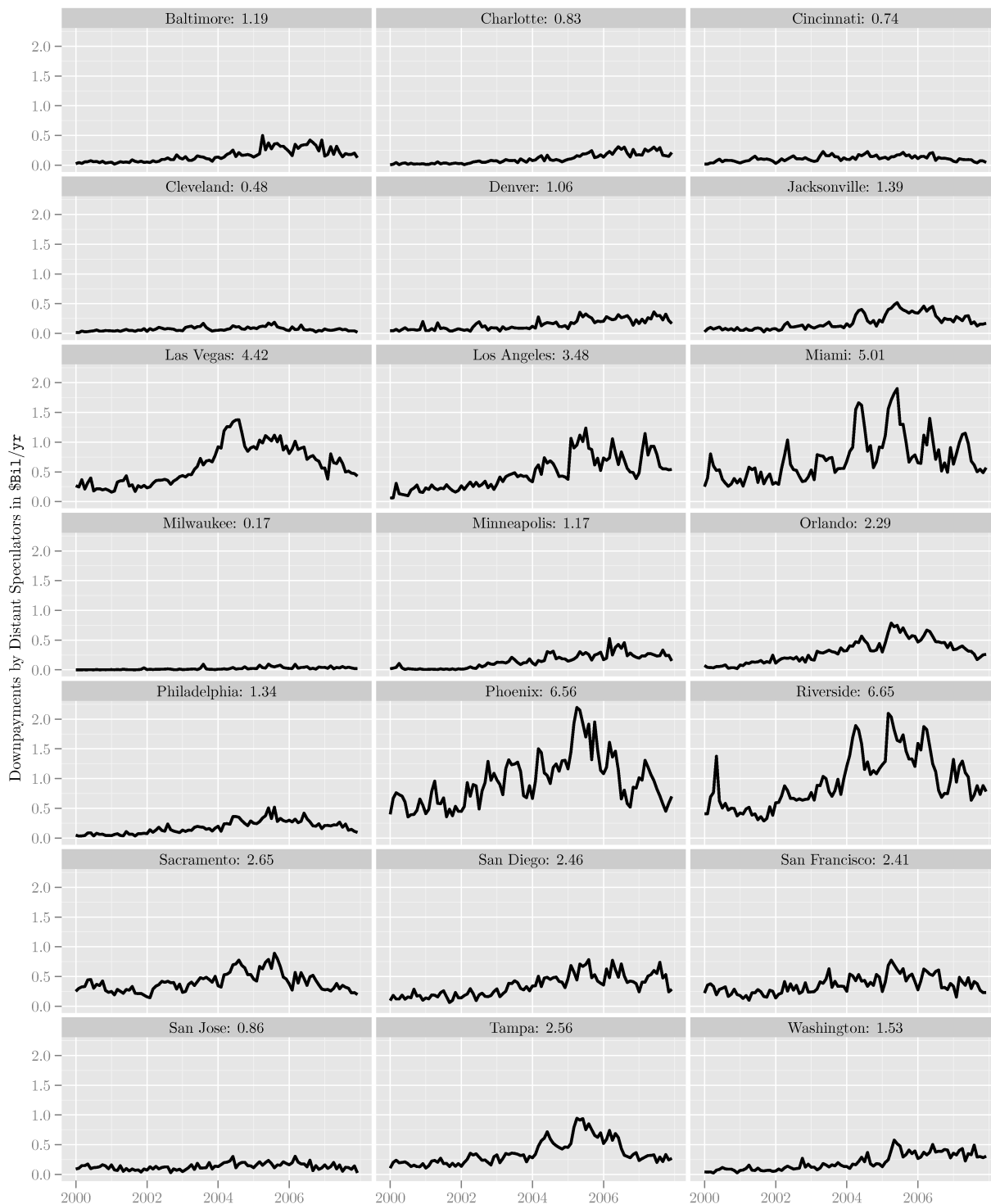
**Figure 2.** *The return on single family house purchases made by local and distant speculators from Jan. 2000 to Dec. 2007 in units of %/yr. The width of the distant speculator line is scaled by the number of second house purchases as a fraction of all sales in units of %. The width of the local speculator line is not scaled.  $\mu_D$  and  $\mu_L$  are the mean rates of return for distant and local speculators over the entire sample in units of %/yr. Reads: “Distant speculators purchasing in Las Vegas in Mar. 2004 earned an 8%/yr return on average; whereas, local speculators earned a 17%/yr return on average in Mar. 2004. The average return on distant speculators purchases decreased from 8%/yr to -15%/yr as the number of out of town second house purchases as a percent of all sales rose from 5% in Mar. 2004 to 13% in Jan. 2007.”*



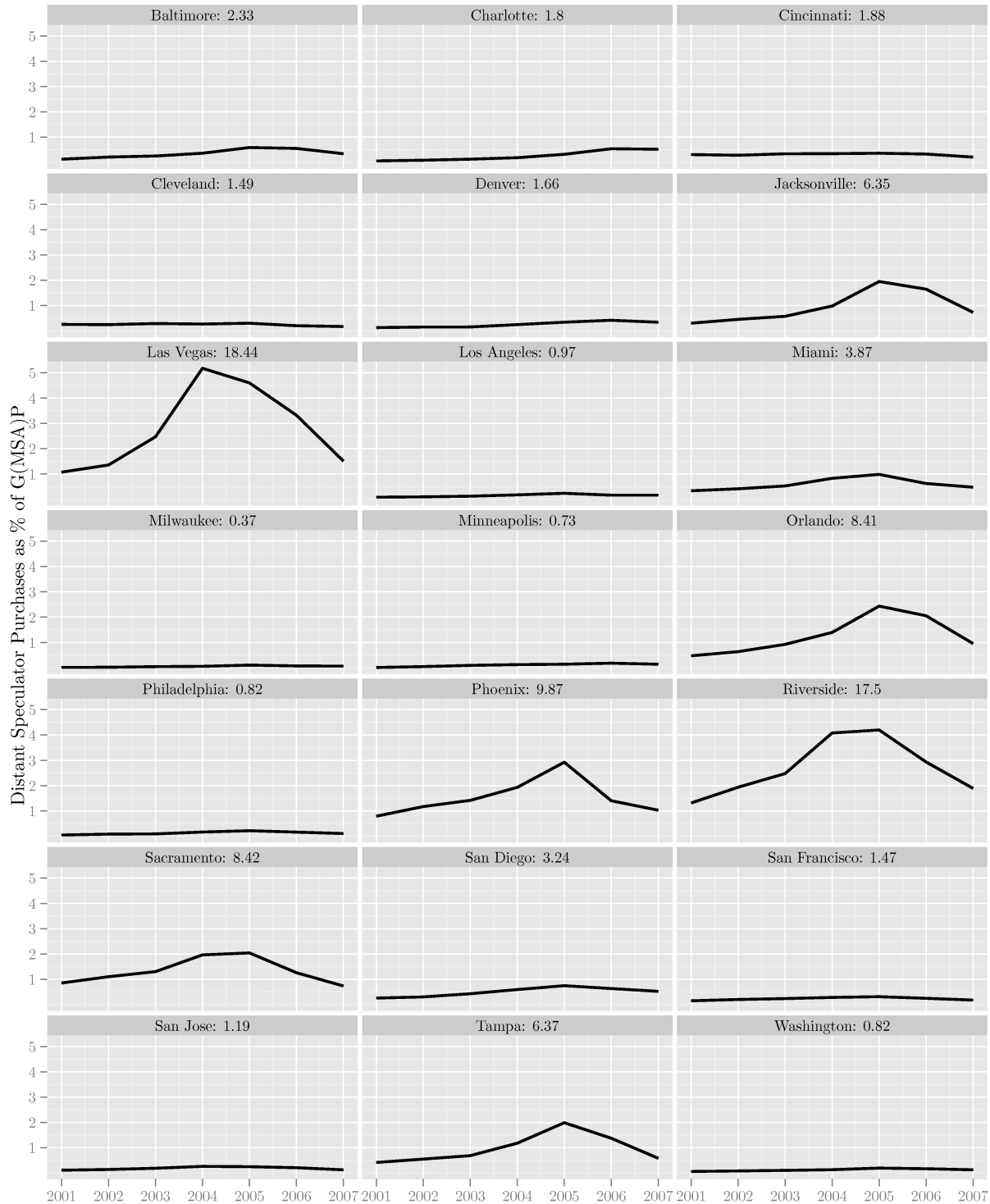
**Figure 3.** *The median and interquartile range of the primary residence house prices for the populations of distant speculators and of all buyers in units of \$100k in each month from Jan. 2000 to Dec. 2007. Reads: “The median price of the primary residence of a distant speculator living in San Francisco who made a second house purchase in some other MSA in Mar. 2005 was \$550k. The median price of a randomly selected home purchased in Mar. 2005 in San Francisco was \$600k.”*



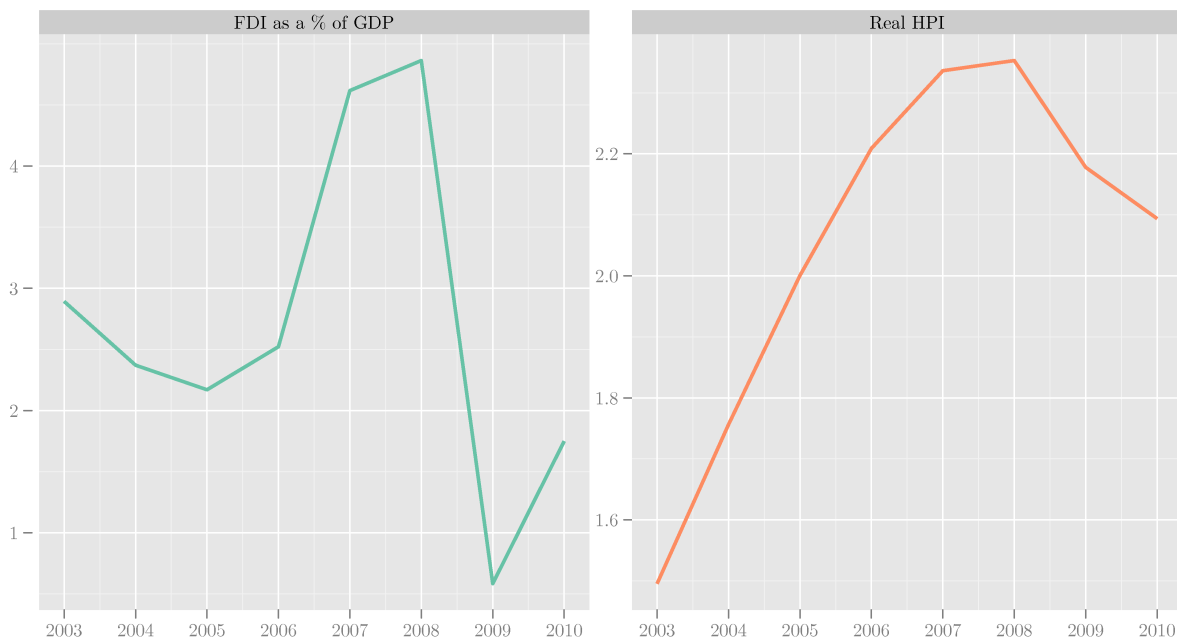
**Figure 4.** The mean cumulative loan to value ratio ( $C-LTV$ ) used to finance single family house purchases made in units of % from 2000 to 2007. We leave the panel for Minneapolis blank due to missing mortgage data in the raw deeds records. Reads: “The average distant speculator buying a \$100k house in Los Angeles in 2004 would have financed this purchase with a \$75k mortgage resulting in a 75%  $C-LTV$ . The average owner occupant buying the same house would have taken out a \$92k mortgage resulting in a 92%  $C-LTV$ .”



**Figure 5.** Total downpayments (annualized) made by distant speculators on single family house purchases each month in units of \$Bil/yr from Jan. 2000 to Dec. 2007. The number at the top of each panel represents the sum of the downpayments by distant speculators in each MSA from Jan. 2002 to Dec. 2007.



**Figure 6.** Sum of the sales prices of single family houses sold to distant speculators as a fraction of total  $G(MSA)P$  in each MSA in units of % from 2000 to 2007. We compute  $G(MSA)P$  using data from the BEA as the product of the per capita income in each MSA times the population. The number at the top of each panel represents the sum of the  $G(MSA)P$  shares in each MSA from 2002 to 2007.



**Figure 7.** *Left Panel: Net foreign direct investment (henceforth, FDI) in Spain from the World Bank as a percent of Spain’s GDP from 2003 to 2010. Reads: “Net FDI inflows into Spain amounted to a little less than 5% of Spain’s GDP in 2008.” Right Panel: Real HPI index level in Spain over this same time period. Reads: “The real HPI index level rose by just over 230% from a base of 1 in 2000.”*

## REFERENCES

- [1] ABREU, D., AND BRUNNERMEIER, M. Bubbles and crashes. *Econometrica* 71 (2003).
- [2] BARRO, R., AND URSÚA, J. Macroeconomic crises since 1870. *Brookings Papers on Economic Activity* (2008).
- [3] BAYER, P., GEISSLER, C., AND ROBERTS, J. Speculators and middlemen: The role of flippers in the housing market. *National Bureau of Economic Research Working Paper Series No. 16784* (2011).
- [4] BLACK, F. Noise. *The Journal of Finance* 41 (1986).
- [5] BRUNNERMEIER, M., AND NAGEL, S. Hedge funds and the technology bubble. *The Journal of Finance* 59 (2004).
- [6] CASE, K., AND SHILLER, R. The efficiency of the market for Single-Family homes. *The American Economic Review* 79 (1989).
- [7] DE LONG, B., SHLEIFER, A., SUMMERS, L., AND WALDMANN, R. Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance* 45 (1991).

- [8] DEMYANYK, Y., AND VAN HEMERT, O. Understanding the subprime mortgage crisis. *Review of Financial Studies* 24 (2011).
- [9] EMRATH, P. Property taxes in the 2000 census. *Housing Economics* 12 (2002).
- [10] FERREIRA, F., AND GYOURKO, J. Anatomy of the beginning of the housing boom: U.S. neighborhoods and metropolitan areas, 1993-2009. *National Bureau of Economic Research Working Paper Series No. 17374* (2011).
- [11] FRIEDMAN, M. *Essays in Positive Economics*. University of Chicago Press, 1953.
- [12] GLAESER, E., GOTTLIEB, J., AND GYOURKO, J. Can cheap credit explain the housing boom? *National Bureau of Economic Research Working Paper Series No. 16230* (2010).
- [13] GLAESER, E., AND GYOURKO, J. Arbitrage in housing markets. *National Bureau of Economic Research Working Paper Series No. 13704* (2007).
- [14] GREENWOOD, R., AND NAGEL, S. Inexperienced investors and bubbles. *Journal of Financial Economics* 93 (2009).
- [15] GRIFFIN, J., HARRIS, J., SHU, T., AND TOPALOGLU, S. Who drove and burst the tech bubble? *Journal of Finance* (2010).
- [16] HARDING, J., MICELI, T., AND SIRMANS, C. Do owners take better care of their housing than renters? *Real Estate Economics* 28 (2000).
- [17] HAUGHWOUT, A., LEE, D., TRACY, J., AND VAN DER KLAAUW, W. Real estate investors, the leverage cycle, and the housing market crisis. *Federal Reserve Bank of New York Staff Reports* 514 (2011).
- [18] HIMMELBERG, C., MAYER, C., AND SINAI, T. Assessing high house prices: Bubbles, fundamentals and misperceptions. *The Journal of Economic Perspectives* 19 (2005).
- [19] HONG, H., SCHEINKMAN, J., AND XIONG, W. Asset float and speculative bubbles. *The Journal of Finance* 61 (2006).
- [20] HUBBARD, G., AND MAYER, C. The mortgage market meltdown and house prices. *The B.E. Journal of Economic Analysis and Policy* 9 (2009).
- [21] JAVORCIK, B. Does foreign direct investment increase the productivity of domestic firms? in search of spillovers through backward linkages. *American Economic Review* 94 (2004).
- [22] KYLE, A. Continuous auctions and insider trading. *Econometrica* 53 (1985).
- [23] LEE, D., MAYER, C., AND TRACY, J. A new look at second liens. *SSRN* (2012).
- [24] LEVITT, S., AND SYVERSON, C. Market distortions when agents are better informed: The value of information in real estate transactions. *Review of Economics and Statistics* 90 (2008).
- [25] LI, W., AND GAO, Z. Real estate investors and the boom and bust of the us housing market. *SSRN* (2012).

- [26] MACKEY, C. *Extraordinary Popular Delusions and the Madness of Crowds*. Orig: Richard Bentley, 1841.
- [27] MAYER, C. Housing bubbles: A survey. *Annual Review of Economics* 3 (2010).
- [28] MAYER, C., PENCE, K., AND SHERLUND, S. The rise in mortgage defaults. *Journal of Economic Perspectives* 23 (2009).
- [29] MIAN, A., AND SUFI, A. The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis. *Quarterly Journal of Economics* 124 (2009).
- [30] MIAN, A., AND SUFI, A. House prices, home Equity-Based borrowing, and the U.S. household leverage crisis. *American Economic Review* 101 (2011).
- [31] OFEK, E., AND RICHARDSON, M. Dotcom mania: The rise and fall of internet stock prices. *The Journal of Finance* 58 (2003).
- [32] OFFICE FOR NATIONAL STATISTICS. Community housing survey, 2007.
- [33] PEARL, J. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.
- [34] POTERBA, J. Tax subsidies to Owner-Occupied housing: An Asset-Market approach. *The Quarterly Journal of Economics* 99 (1984).
- [35] SAGALYN, L. *Cases in Real Estate Finance and Investment Strategy*. Urban Land Institute, 1999.
- [36] SCHEINKMAN, J., AND XIONG, W. Overconfidence and speculative bubbles. *The Journal of Political Economy* 111 (2003).
- [37] SHLEIFER, A., AND VISHNY, R. The limits of arbitrage. *The Journal of Finance* 52 (1997).
- [38] STEIN, J. Prices and trading volume in the housing market: A model with Down-Payment effects. *The Quarterly Journal of Economics* 110 (1995).

## APPENDIX A. ADDITIONAL DATA DESCRIPTION

The raw data are housed in paper form (and sometimes in PDF files on websites), making them inaccessible to computational analysis on a large scale. Our data provider compiled information from the county registrar websites that house the public raw data. The primary data source for this analysis agreed to provide the data only if we guaranteed anonymity. However, such data are available from multiple sources such as DATAQUICK for researchers interested in replicating or expanding our analysis.

We conducted a number of validity checks to ensure that any differences we observe across MSAs are not due to variation in data standards. First, we cross checked the sales counts



reported in the deeds records against county and city level sales estimates reported by local realtor associations. We also compared our anonymous data source with records reported in DATAQUICK where the data overlapped. In addition, we also spot checked individual records against values reported by LEXISNEXIS, which provides a searchable database of property transaction records. We restrict our sample to sales of single family houses with valid sale amounts and addresses located within the 21 MSAs listed in Table 11, which also lists the FIPS codes for each county in our sample.

Note that the introduction of additional distant speculators into MSA  $i$  at month  $t$  does not mean that either owner occupants or local speculators must leave this market as new houses can be built for speculators to purchase, existing houses can sell more quickly with less transitional vacancy, or some local residents can choose to rent instead of own houses. All three factors were likely prevalent during the boom. For instance, while regional data for average the time on the market is not available, news stories suggest that the average sale took as little as 1 or 2 weeks in cities like Phoenix and Miami during the height of the boom. What's more, new construction also grew rapidly in these markets at the same time investor purchases were high. By contrast, homeownership rates in the US hit their peak during 2004, and were declining two years before house prices stopped rising. Thus, there is not a sharp adding up constraint linking the number of house owners in an MSA and its population.

## APPENDIX B. PROOFS

*Proof (Proposition 1).* Substituting both the functional form for the housing price in MSA  $j$  from Equation (5) and the functional form for the aggregate demand in MSA  $j$  from Equation (4) into the objective function for an individual trader from MSA  $i$  yields an expression:

$$\begin{aligned}
 W_{i \rightarrow j} &= \max_{\theta_{i \rightarrow j}} \mathbf{E} [ (V_j - \alpha - \beta \cdot X_j) \cdot \theta_{i \rightarrow j} \mid V_j ] \\
 &= \max_{\theta_{i \rightarrow j}} \mathbf{E} \left[ \left( V_j - \alpha - \beta \cdot \sum_{i'=1}^I (Q_{i'} \cdot \theta_{i',j}) - \beta \cdot \varepsilon_j \right) \cdot \theta_{i \rightarrow j} \mid V_j \right]
 \end{aligned} \tag{25}$$

$i$	$H_{i,T} \times 10^{-6}$	County FIPS Codes
Baltimore, MD	0.677	24003, 24005, 24013, 24025, 24027, 24035, 24510
Charlotte, NC-SC	0.565	37007, 37025, 37071, 37119, 37179, 45091
Cincinnati, IN-KY-OH	0.430	18029, 18047, 18115, 21015, 21023, 21037, 21077, 21081, 21117, 21191, 39015, 39017, 39025, 39061, 39165
Cleveland, OH	0.594	39035, 39055, 39085, 39093, 39103
Denver, CO	0.649	08001, 08005, 08014, 08019, 08031, 08035, 08039, 08047, 08059, 08093
Jacksonville, FL	0.418	12003, 12019, 12031, 12089, 12109
Las Vegas, NV	0.478	32003
Los Angeles, CA	1.978	06037, 06059
Miami, FL	1.058	12011, 12086, 12099
Milwaukee, WI	0.311	55079, 55089, 55131, 55133
Minneapolis, MN-WI	0.904	27003, 27019, 27025, 27037, 27053, 27059, 27123, 27139, 27141, 27163, 27171, 55093, 55109, 55109
New York, NJ-NY-PA	2.989	34003, 34013, 34017, 34019, 34023, 34025, 34027, 34029, 34031, 34035, 34037, 34039, 36005, 36047, 36059, 36061, 36079, 36081, 36085, 36087, 36103, 36119, 42103
Orlando, FL	0.595	12069, 12095, 12097, 12117
Philadelphia, DE-MD-NJ-PA	1.437	10003, 34005, 34007, 34015, 34033, 42017, 42029, 42045, 42091, 42101, 24015
Phoenix, AZ	1.149	04013, 04021
Riverside, CA	1.072	06065, 06071
Sacramento, CA	0.618	06017, 06061, 06067, 06113
San Diego, CA	0.622	06073
San Francisco, CA	0.892	06001, 06013, 06041, 06075, 06081
San Jose, CA	0.337	06069, 06085
Tampa, FL	0.804	12053, 12057, 12101, 12103
Washington, DC-MD-VA-WV	1.083	11001, 24009, 24017, 24021, 24031, 24033, 51013, 51043, 51059, 51061, 51107, 51153, 51177, 51179, 51187, 51510, 51600, 51610, 51630, 51683, 51685, 54037

**Table 11.** This table displays a list of the 21 MSAs used in our analysis.  $H_{i,T} \times 10^{-6}$  denotes the number of single family houses in each MSA  $i$  as of Dec. 2007 in millions of homes. Counties gives a list of the county FIPS codes defining each MSA. For brevity, we omit the state abbreviations for each MSA in all future references.

Since all traders know the true value of  $V_j$  and pricing and demand rules are linear, then  $\theta_{i \rightarrow j} = \theta_{i' \rightarrow j}$  for all  $i, i' \in \{1, 2, \dots, I\}$ . Enforcing this symmetry simplifies the above expression to:

$$\bar{W}_j = \max_{\theta_j} \mathbf{E} \left[ \left( V_j - \alpha - \beta \cdot \bar{\theta}_j \cdot \sum_{i'=1}^I Q_{i'} - \beta \cdot \varepsilon_j \right) \cdot \bar{\theta}_j \mid V_j \right] \quad (26)$$

Taking the first order condition with respect to  $\bar{\theta}_j$  and evaluating the expectations operator then delivers the relationship:

$$\begin{aligned} 0 &= \mathbf{E} \left[ V_j - \alpha - 2 \cdot \beta \cdot \bar{\theta}_j \cdot \sum_{i'=1}^I Q_{i'} - \beta \cdot \varepsilon_j \mid V_j \right] \\ &= V_j - \alpha - 2 \cdot \beta \cdot \bar{\theta}_j \cdot \sum_{i'=1}^I Q_{i'} \end{aligned} \quad (27)$$

Solving for  $\bar{\theta}_j$  yields the desired result linking  $V_j$  and  $\bar{\theta}_j$  in a linear equation.

Next, in order to solve for the pricing coefficients  $\alpha^*$  and  $\beta^*$  as well as the common demand per trader coefficients  $\bar{\gamma}^*$  and  $\bar{\delta}^*$  we use a coefficient matching strategy. First, we express the market makers' collective pricing rule as a projection to yield an expression for  $\beta$  as a function of  $\beta$  and the parameters  $\sigma_V$  and  $\sigma_\varepsilon$ :

$$\begin{aligned} P_j &= \mathbf{E} [ V_j \mid X_j ] \\ &= \mu_V + \left( \frac{\text{Cov}[V_j, X_j]}{\text{Var}[X_j]} \right) \cdot (X_j - \mu_X) \\ &= \mu_V + \left( \frac{\text{Cov} \left[ V_j, -\left(\frac{\alpha}{2\beta}\right) + \left(\frac{1}{2\beta}\right) \cdot V_j + \varepsilon_j \right]}{\text{Var} \left[ -\left(\frac{\alpha}{2\beta}\right) + \left(\frac{1}{2\beta}\right) \cdot V_j + \varepsilon_j \right]} \right) \cdot (X_j - \mu_X) \\ &= \mu_V - \left( \frac{\left(\frac{1}{2\beta}\right) \cdot \sigma_V^2}{\left(\frac{1}{2\beta}\right)^2 \cdot \sigma_V^2 + \sigma_\varepsilon^2} \right) \cdot \mu_X + \underbrace{\left( \frac{\left(\frac{1}{2\beta}\right) \cdot \sigma_V^2}{\left(\frac{1}{2\beta}\right)^2 \cdot \sigma_V^2 + \sigma_\varepsilon^2} \right)}_{\beta} \cdot X_j \end{aligned} \quad (28)$$

Solving for  $\beta$  yields the equilibrium choices of  $\alpha^*$  and  $\beta^*$  for the market makers. Next, we express aggregate demand for housing in MSA  $j$ ,  $X_j$ , as a linear function of value of housing in MSA  $j$ ,  $V_j$ , in order to get an expression for  $\bar{\gamma}^*$  and  $\bar{\delta}^*$  in terms of the equilibrium coefficients  $\alpha^*$  and  $\beta^*$  as well as the summation  $\sum_{i'=1}^I Q_{i'}$ :

$$\begin{aligned} X_j &= \sum_{i=1}^I Q_i \cdot \bar{\theta}_j + \varepsilon_j \\ &= \sum_{i=1}^I Q_i \cdot \left( \underbrace{-\left(\frac{\alpha^*}{2 \cdot \beta^* \cdot \sum_{i'=1}^I Q_{i'}}\right)}_{\bar{\gamma}^*} + \underbrace{\left(\frac{1}{2 \cdot \beta^* \cdot \sum_{i'=1}^I Q_{i'}}\right)}_{\bar{\delta}^*} \cdot V_j \right) + \varepsilon_j \end{aligned} \quad (29)$$

□

*Proof (Proposition 2).* If the market makers do not realize that traders may be overconfident or uninformed, they will adopt the same pricing rule as in Proposition 1. What's more, both traders with correct beliefs in MSAs  $i' \neq i$  and traders with overconfident beliefs in MSA  $i$  think that all other agents share their beliefs so that they anticipate a price in MSA  $j$  of:

$$\mathbb{E} [ P_j \mid \text{MSA} ] = \begin{cases} \alpha^* + \beta^* \cdot \sum_{i'=1}^I Q_{i'} \cdot (\bar{\gamma}^* + \bar{\delta}^* \cdot V_j) & \text{if MSA} \neq i \\ \alpha^* + \beta^* \cdot \sum_{i'=1}^I Q_{i'} \cdot (\bar{\gamma}^* + \bar{\delta}^* \cdot \{V_j + \eta\}) & \text{if MSA} = i \end{cases} \quad (30)$$

However, the realized total demand in MSA  $j$  given that traders in MSA  $i$  have inflated beliefs,  $\tilde{X}_j^{(i)}$ , will be given by:

$$\begin{aligned} \tilde{X}_j^{(i)} &= \sum_{i' \neq i} Q_{i'} \cdot (\bar{\gamma}^* + \bar{\delta}^* \cdot V_j) + Q_i \cdot (\bar{\gamma}^* + \bar{\delta}^* \cdot \{V_j + \eta\}) \\ &= \sum_{i'=1}^I Q_{i'} \cdot (\bar{\gamma}^* + \bar{\delta}^* \cdot V_j) + Q_i \cdot \bar{\delta}^* \cdot \eta \end{aligned} \quad (31)$$

Thus, the difference between the price levels in MSA  $j$  in the fully informed regime and the regime with misinformed speculators will be given by  $\tilde{P}_j^{(i)} - P_j = Q_i \cdot \beta^* \cdot \bar{\delta}^* \cdot \eta$ . Substituting in the functional forms for the equilibrium coefficients  $\beta^*$  and  $\bar{\delta}^*$  from Proposition 1 yields the desired result. □