DISTANT SPECULATORS AND ASSET BUBBLES IN THE HOUSING MARKET

ALEX CHINCO AND CHRIS MAYER

Abstract. We investigate the role that out of town second house buyers ("distant speculators") played in bubble formation in the US residential housing market. Distant speculators are likely to be more reliant on capital gains rather than dividend consumption for financial returns as well as less informed about local market conditions. Using transactions level data that identify the address of both the purchased property and the primary residence of the buyer, we show that an increase in purchases by distant speculators (but not local speculators) is strongly correlated with appreciation in both house price and implied-to-actual rent ratios (IAR)—a proxy for mispricing in the housing market. We develop a simple model that helps us address the issue of reverse causality. Consistent with this model, we show that the size of the MSA that out of town second house buyers come from is positively related to the impact of distant speculators on house price and IAR appreciation rates in the target MSA suggesting that out of town second house buyers are not simply responding to unobserved changes in housing values in the target MSA. We conclude by demonstrating the large impact that distant speculators have on the local economy, with out of town second house purchases equalling as much as 5% of total output in Las Vegas during the boom.

Date: August 18, 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. We thank Viral Acharya, Moshe Cohen, Xavier Gabaix, Ed Glaeser, Robin Greenwood, Gur Huberman, Aurel Hizmo, Andrew Paciorek (discussant), Tomasz Piskorski, Alexi Savov, David Sraer (discussant), Jeremy Stein, Stijn Van Nieuwerburgh, Reed Walker and Jeff Wurgler as well as seminar participants at Harvard Business School, NYU Stern, Columbia Business School, the NBER Behavioral Finance/Housing Bubbles joint session, the NBER Conference Project on Housing and the Financial Crisis, the Federal Reserve Bank of Cleveland, and Riksbank for 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 (1986) and De Long et al. (1991), 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, informational frictions, or a limited supply of tradable shares.[1] Real estate researchers have also long puzzled over the inefficiency of housing prices,[2] and several papers specifically point to the possible role of second house buyers in inflating house prices during the recent boom.[3]

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, identify a group of overconfident or uninformed speculators; second, show that an increase in the trading volume of these speculators predicts future mispricing; and third, demonstrate that these speculators are not simply responding to unobserved variation in asset values—i.e., address the question of reverse causality. This is a tall order using standard financial datasets. 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[4] suggests that a large number of inexperienced traders increased their holdings of technology and communications stocks during this period while many traders began active stock trading for the first time. It might

[1]See Shleifer and Vishny (1997), Scheinkman and Xiong (2003), and Ofek and Richardson (2003) respectively.
[2]See the seminal paper by Case and Shiller (1989) as well as a recent survey Mayer (2010) for a discussion of the literature on housing bubbles.
[3]Bayer et al. (2011) 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) examine credit report data and show that mortgages on second houses represented nearly half of all mortgages originated in the 4 states with the highest price appreciation at the peak of the market. Li and Gao (2012) present theoretical results that second house buyers can fuel a boom as well as empirical evidence showing that second house buyers are both more likely to be present in MSAs with high house price appreciation and also more likely to 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 house purchases in the highest appreciating MSAs; however, none of the existing papers is able to directly address the issue of reverse causality. Our work, below, also extends this analysis to differentiate between local second house buyers and out of town second house buyers and shows that only the purchases of the latter group appear to be causing some degree of mispricing.
[4]See Greenwood and Nagel (2009).

seem obvious that this rapid growth in the price of Cisco's stock must have been driven by this influx of overconfident or uniformed speculators. After all, Cisco did not come close to delivering a dividend stream that warranted its price in early 2000.[5]

With regards to the first challenge, identifying a group of misinformed speculators in the stock market is difficult because, for the most part, traders are anonymous. Turning to the second challenge, we note that where it is possible to identify particular types of traders in the data, it is not obvious there is a linear relationship between the trading volume of misinformed speculators and future price increases. For instance, Brunnermeier and Nagel (2004) document that a group of hedge funds bought technology stocks during the Dot-Com boom and strategically sold them just prior to the crash. Finally, with respect to the third challenge, even if a large number of misinformed speculators were trading Cisco's stock, it is difficult to sort out whether the speculative trading caused the extraordinary 500% price increase or whether they were attracted to Cisco's stock by the perception that it was undervalued.[6] 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 boom and bust cycles that resemble bubbles. However, unlike the stock market, the housing market offers researchers detailed microdata on traders (i.e., house buyers). As well, the housing market is geographically segmented into metropolitan statistical areas (henceforth, MSAs) making it an attractive laboratory to study bubble formation because house prices do not follow the same time series pattern[7] and home buyers in different MSAs may use somewhat different information when making their purchases. We make use of these features to test for speculative bubbles due to an increase in demand from out of town second house buyers in 21 MSAs from Jan. 2000 to Dec. 2007.

[5]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.

[6]This argument involving reverse causality is commonly referred to as the Friedman Critique and dates to Friedman (1953). See Abreu and Brunnermeier (2003) for an example of a model where traders arrive in a market in order to earn profits by riding excess price appreciation.

[7]As documented in Ferreira and Gyourko (2011) the recent boom began at different times in different MSAs, and house prices exhibited different appreciation rates across these markets. Even the start dates of the subsequent decline in prices differed by a year or more.

In Section 2, we describe the datasets used in our analysis which include sales and mortgages transactions for every single family housing unit in this sample as well as monthly indexes for real house prices (henceforth, HPI) and implied-to-actual rent ratios (henceforth, IARs) for each of these MSAs. The IAR data uses the methodology from Himmelberg et al. (2005) to compute a measure of mispricing in the housing market that compares the cost of renting a house and the imputed rent to an owner occupant (the annual after-tax cost of owning a house). Section 3 then outlines a simple economic model of speculation. This model illustrates how we employ housing data to address the three challenges listed above when studying the price impact of a specific group of overconfident or uninformed speculators—namely, out of town second house buyers.

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 second house in a different MSA from which they live, behave much like overconfident or uninformed speculators. Out of town second house buyers (so-called "distant speculators"[8]) appear less informed about local market conditions. These buyers entered markets such as Phoenix, Las Vegas, Miami and Tampa in much larger numbers just prior to the peak in house price levels and earned lower capital gains on their investment relative to local speculators. However, capital gains are only part of a return computation. We suggest that out of town second house buyers were likely less able to consume the dividend stream from their housing purchase as compared to local second house buyers or owner occupants. After all, out of town second house buyers can only live in their houses for a fraction of the year, face higher property taxes and have difficulties monitoring agents who maintain their property.

In Section 5, we address the second key challenge by estimating a set of panel vector autoregressions, showing that an increase in the fraction of all sales made by out of town second

[8]In the analysis below, we assign precise definitions to the terms second house buyer, local speculator and distant speculator. We refer to all traders who purchase a house they do not reside in as "second house buyers" or "speculators." Such a house might in fact represent a second, third, fourth (etc. . .) house in addition to their primary residence, or even just a first house if they do not own their primary residence. We use the term "speculator" because second house buyers are less able to consume the full dividend stream from their purchases relative to owner occupants and thus may be more dependent on capital appreciation for their return. This term is not a synonym for irrational traders. We avoid using the term "investors" in that all house buyers are investors.

house buyers in an MSA in a given month predicts increases in the house price and IAR appreciation rates in the next month. We estimate that the 3 standard deviation increase in distant speculator demand observed in Phoenix in 2004 explains around one sixth of the 30%/yr increase in house price and IAR appreciation rates in that market. By contrast, the lagged share of local second house buyers has little impact on either house price or IAR appreciation rates.

Section 6 addresses the third and final challenge, the issue of reverse causation, by exploiting geographic segmentation of information across metropolitan areas as well limits to arbitrage in the housing market. We examine the null hypothesis that distant speculators are responding to a common set of shocks about the value of housing in a target market. The key insight is that, if out of town second house buyers are responding to common shocks, then buyers living in each other MSA should purchase houses in the target market in roughly equal proportions after controlling for factors such as distance and ease of information transmission. These common shocks could be information that impacts the fundamental value of housing (e.g., the Friedman Critique) or a common behavioral factor. Yet, we find that the size of the MSA that out of town second house buyers come from is positively related to the impact of these distant speculators on house price and IAR appreciation rates in the target MSA. These regressions control for both MSA pair specific factors and macroeconomic factors with ordered MSA pair and time fixed effects. This violation of symmetry allows us to reject the null hypothesis of a common information shock and is thus consistent with the alternative hypothesis that distant speculators themselves helped push up house prices.

We conclude by pointing out similarities between the US housing bubble and housing bubbles in other countries such as Spain, where commentators have pointed to a large influx of distant speculators from Germany and Britain as an important contributor to the large increase in prices. 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 three main sources: county deeds records obtained from Dataquick and an anonymous data provider, HPI data from Zillow, Inc. and IAR data computed according to the procedure developed in Himmelberg et al. (2005). 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 $i = 1, 2, \ldots, I$ over the time period $t = 1, 2, \ldots, T$ with $t = 1$ denoting Jan. 2000 and $t = T$ denoting Dec. 2007.

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. Below, we define variables denoting the 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, a trader's identity, motives, and information are generally hard to isolate. One advantage of using the US residential real estate market to study bubble formation is that we can obtain information on all 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 various type of house purchasers in an MSA in a given month.

Definition (Second House Purchases)**.** *Define $S_{i \to j,t}$ 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.*

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

Definition (Distant Speculator Purchases)**.** *Define* $S^{\text{Distant}}_{j,t} = \sum_{j\neq i} S_{i\to j,t}$ *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 $j \neq i$.* $S^{\text{Distant}}_{i,t}$ *has units of houses per year.*

Definition (Local Speculator Purchases)**.** *Define* $S^{\text{Local}}_{j,t} = S_{j\to j,t}$ *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.* $S^{\text{Local}}_{j,t}$ *has units of houses per year.*

Table 1 gives an example of an owner occupant, a local second house buyer (a "local speculator"), and an out of town second house buyer (a "distant speculator") 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 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 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. Many of these MSA specific sparkline plots display 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

	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.*

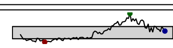
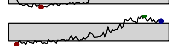
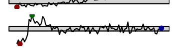
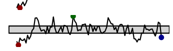
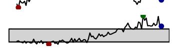
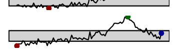
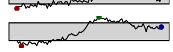
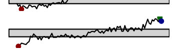
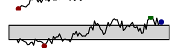
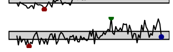
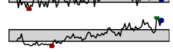
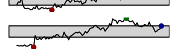
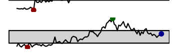
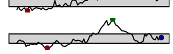
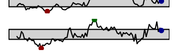
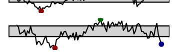
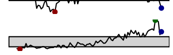
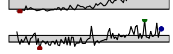
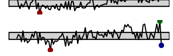
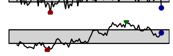
		Distant Speculator Purchases as % of Sales						
		Mean	Sd	Min	Q25	Q50	Q75	Max
Baltimore		4.76	1.94	2.34	3.19	4.34	6.32	9.69
Charlotte		3.33	2.32	0.497	1.39	2.38	5.88	7.80
Cincinnati		6.27	1.18	2.78	5.82	6.29	6.84	9.46
Cleveland		5.37	0.942	3.03	4.87	5.38	5.98	7.49
Denver		2.20	1.28	0.676	1.08	1.70	3.29	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 each sparkline graph covers the interquartile range for each MSA and is not a constant scale.*

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 house prices.

	Local Speculator Purchases as % 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 each sparkline graph covers the interquartile range for each MSA and is not a constant scale.*

2.2. **House Price Growth Rate.** We obtain monthly house price index (HPI) data from Zillow, Inc. at the MSA level. Zillow data are available for a larger number of locations than S&P/Case and Shiller index and uses a methodology that potentially makes the index less sensitive to changes in the mix of properties that sell at a given point in time. The Zillow indexes behave quite similarly to S&P/Case and Shiller indexes during the boom, but show less of a sharp decline in 2007 and 2008.

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

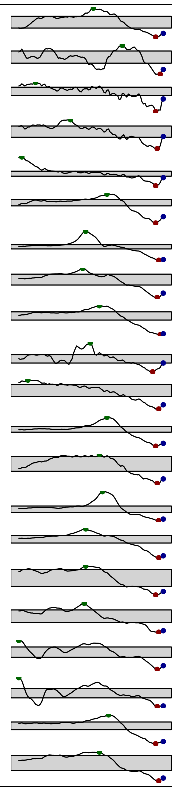
		House Price Appreciation Rates in %/yr						
		Mean	Sd	Min	Q25	Q50	Q75	Max
Baltimore		6.49	9.31	−12.3	−1.77	10.1	12.0	21.4
Charlotte		0.951	3.43	−6.52	−1.56	1.30	3.87	6.39
Cincinnati		0.048	2.39	−6.32	−1.05	0.796	1.70	3.23
Cleveland		−1.87	4.23	−11.2	−4.98	−0.322	1.09	4.48
Denver		−0.309	4.40	−9.61	−2.63	−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.*

Table 4 gives summary statistics for the house price appreciation rate in units of percent per year. A number of the markets saw annual house price appreciation rates above 20%/yr, with house price appreciation rates exceeding 35%/yr in Las Vegas and Phoenix near the peak of their booms. What's more, the sparkline plots show that the timing of these peaks varied substantially from MSA to MSA with the house price appreciation rate peak in Las Vegas arriving more than a year prior to the peak in Phoenix.

2.3. **Implied-to-Actual Rent Ratio Growth Rate.** Beginning with Poterba (1984), 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 occupied properties and mortgage interest payments. 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 the next year for a particular house. Himmelberg et al. (2005) 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) updated through Dec. 2007. Table 5 gives the data sources and a set of short descriptions for the input variables used to compute the user cost of housing in Equation (1).

Definition (User Cost of Housing). *Define* $U_{i,t\to(t+12)}$ *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* $t+12$*:*

$$U_{i,t\to(t+12)} = \rho_t + \omega_{i,t} - \kappa_{i,t} \cdot \{\mu_t + \omega_{i,t}\} + \delta - \mathrm{E}\left[\Delta \log P_{i,t\to(t+12)}\right] + \pi \tag{1}$$

where the user cost of housing has units of $1/\mathrm{yr}$*.*

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 $P_{i,t} \cdot U_{i,t\to(t+12)} = R_{i,t\to(t+12)}$. REIS collects monthly estimates of the annualized rent for a 2-bedroom apartment.

Definition (Apartment Rental Rate Index). *Define* $R_{i,t\to(t+12)}$ *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* $1/\mathrm{yr}$*.*

The log IAR can be thought of as the excess return over the apartment rental rate of a trading strategy whereby an agent borrows money at rate ρ_t per year to buy a house, lives

Variable	Source	Description
ρ_t	CRSP	Risk-free rate computed as annualized 10yr T-Bill.
$\omega_{i,t}$	Emrath (2002)	Property tax rate.
μ_t	Federal Reserve Bank of St. Louis	Mortgage interest rate.
$\kappa_{i,t}$	NBER	Federal marginal tax rate.
δ	Harding et al. (2000)	Housing capital depreciation rate.
$\mathrm{E}\left[\Delta \log P_{i,t\to(t+12)}\right]$	Ferreira and Gyourko (2011), the US Census, and the Livingston Survey	Expected house price appreciation rate equals historical long-term real growth rates by MSA plus expected inflation.
π	Flavin and Yamashita (2002)	Risk premium associated with owning a house.

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). All variables have units of* 1/yr *except for the federal marginal tax rate, $\kappa_{i,t}$, which is dimensionless. All variables reflect rates over the time interval t to $t+12$ and are known at time t.*

in the house for a year while paying a constant proportion of the house value in depreciation costs δ per year and earning the tax shield $\kappa_{i,t}$ on his debt payments of $(\mu_t+\omega_{i,t})$ per year and then sells the house after one year getting capital gains at the expected price appreciation rate of $\mathrm{E}\left[\Delta \log P_{i,t\to(t+12)}\right]$ per year[9] while enduring a constant risk premium of π per year.[10] When the IAR 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 Ratio (IAR) Appreciation Rate)**.** *$Z_{i,t}$ denotes the IAR in MSA i at month t reflecting 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*

[9]All IAR appreciation rate results are robust to using various specifications for the expected house price appreciation rate including estimates of the 6mo, 1yr or 2yr rolling averages using both Zillow and OFHEO price indexes.

[10]Himmelberg et al. (2005), 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, abstracting from important short-run considerations like easy and cheap leverage in the mid-2000s and time varying risk premia.

property for the same amount of time:

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

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

$\Delta \log Z_{i,t \to (t+\tau)} = \log Z_{i,t+\tau} - \log Z_{i,t}$ *denotes the IAR appreciation rate in units of* $1/\tau$mo.

The IAR 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 annual IAR appreciation rates. 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.

As shown in Equation (3) below, the monthly IAR appreciation rate has the attractive interpretation of being the house price appreciation rate from month t to month $t+1$ deflated by the rate at which owning became more attractive than renting over that same period:

$$
\Delta \log Z_{i,t \to (t+1)} = \Delta \log P_{i,t \to (t+1)} - \left(\Delta \log R_{i,t \to (t+1)} - \Delta \log U_{i,t \to (t+1)} \right) \tag{3}
$$

Thus, if rental rates rose by 1%/mo and the user cost of house ownership fell by 2%/mo over the interval from month t to month $t+1$, a 5%/mo increase in the house price appreciation rate would only represent a 2%/mo increase in the IAR appreciation rate over that same period due to the offsetting effect of the increase in the attractiveness of owning relative to renting.[11]

Researchers have critiqued the user cost approach in a number of ways. For example, Glaeser and Gyourko (2007) 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.[12]

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.[13] Hubbard and Mayer (2009) estimate the log-linearized version of the user cost model:

$$\log P_{i,t} = \alpha_i + \kappa_t + \beta \cdot \log R_{i,t\to(t+12)} + \gamma \cdot \log U_{i,t\to(t+12)} + \varepsilon_{i,t} \tag{4}$$

over the time interval from Jan. 1980 to Dec. 2007 with both MSA and year fixed effects.[14] 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.[15]

[11]The motivation for using the IAR rather than the price to rent ratio comes from the fact that the dividend streams earned by house buyers systematically vary across locations and times due to factors such as tax treatment and prevailing interest rates. In the stock market, there are circumstances in which shares of the same stock might confer different dividend streams. For instance, one would expect that shares of the same stock with voting rights would command a higher price to dividend ratio than shares without voting rights and that the size of this premium might vary over time as documented in Zingales (1995).

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

[13]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.

[14]See Hubbard and Mayer (2009), Table 2.

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

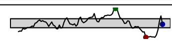
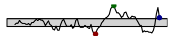
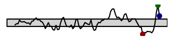
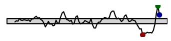
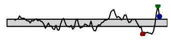
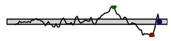
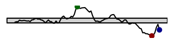
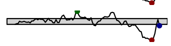
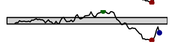
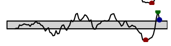
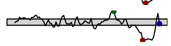
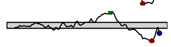
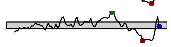
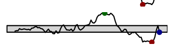
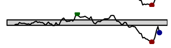
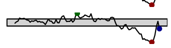
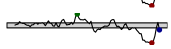
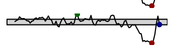
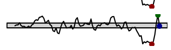
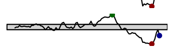
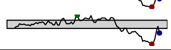
		IAR Appreciation Rates in %/yr						
		Mean	Sd	Min	Q25	Q50	Q75	Max
Baltimore		2.04	9.75	−18.5	−4.14	2.39	9.36	24.8
Charlotte		−1.00	8.53	−16.4	−6.42	−1.83	4.06	21.9
Cincinnati		−1.72	5.31	−12.6	−5.26	−1.55	1.76	14.9
Cleveland		−3.41	5.70	−18.2	−6.04	−2.63	−0.545	13.1
Denver		−2.42	7.51	−17.5	−7.18	−1.91	2.36	20.4
Jacksonville		0.792	11.8	−25.9	−4.13	0.494	6.11	31.4
Las Vegas		0.640	16.6	−39.4	−5.42	0.678	6.85	36.9
Los Angeles		0.210	14.0	−39.6	−4.44	3.12	8.47	24.7
Miami		2.72	14.9	−33.1	−0.856	4.66	12.1	24.7
Milwaukee		0.047	6.67	−15.1	−3.43	0.548	4.54	14.3
Minneapolis		−0.095	7.37	−19.6	−2.46	0.872	4.76	13.5
Orlando		1.84	14.7	−32.5	−4.01	1.66	9.34	32.6
Philadelphia		2.37	7.11	−14.5	−0.852	2.36	6.65	17.8
Phoenix		0.205	16.6	−32.3	−8.55	−1.65	7.01	36.6
Riverside		0.369	16.4	−46.3	−3.20	3.69	10.7	27.1
Sacramento		0.938	13.0	−35.2	−4.28	4.78	9.30	20.1
San Diego		−1.38	12.8	−37.2	−4.87	1.64	5.91	25.0
San Francisco		−0.385	13.8	−41.1	−3.47	3.68	7.85	18.2
San Jose		0.20	13.2	−36.1	−2.39	1.81	7.42	23.9
Tampa		1.57	13.5	−30.9	−2.26	3.61	8.49	27.8
Washington		1.64	11.6	−29.1	−3.04	4.59	10.5	18.6
Mean		0.246	11.5					

Table 6. *This table displays the IAR 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.*

3. A Simple Model of Speculation

In this section we develop a simple noisy 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 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. In Subsection 3.3, we use this simple economic model to frame the challenges facing an econometrician in 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 $V_i \overset{\text{iid}}{\sim} \mathrm{N}(\mu_v, \sigma_v^2)$.

There are Q_i traders in each MSA i indexed by $q = 1, 2, \ldots, Q_i$. Let $\vartheta_{q,i \to j}$ denote the number of houses in MSA j demanded by the q^{th} trader in MSA i and let $S_{i \to j}$ denote the total number of houses in MSA j demanded by traders in MSA i. We denote the average demand for houses in MSA j by traders living in MSA i as $\theta_{i \to j} = (1/Q_i) \cdot \sum_{q=1}^{Q_i} \vartheta_{q,i \to j}$ and can interpret this quantity as the probability that a randomly selected trader in MSA i buys a house in MSA j. Total demand for housing in MSA j, denoted X_j, is defined as the sum of the housing demand from each MSA i plus an MSA specific demand shock ε_j:

$$X_j = \sum_{i=1}^{I} S_{i \to j} + \varepsilon_j = \sum_{i=1}^{I} \left(\sum_{q=1}^{Q_i} \vartheta_{q,i \to j} \right) + \varepsilon_j = \sum_{i=1}^{I} (Q_i \cdot \theta_{i \to j}) + \varepsilon_j \tag{5}$$

where ε_j is an iid draw from a normal distribution $\varepsilon_j \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ and X_j has units of houses.

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:

$$P_j = \mathrm{E}[V_j | X_j] = \alpha + \beta \cdot X_j \tag{6}$$

The coefficient β can be interpreted as the dollar change in the price of housing in MSA j when traders demand one additional unit of housing in MSA j. Market makers might be developers 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[16] that traders use a linear demand rule given by:

$$\vartheta_{q,i\to j} = \gamma_{q,i\to j} + \delta_{q,i\to j} \cdot V_j \tag{7}$$

The coefficient $\gamma_{q,i\to j}$ has units of houses per trader and the coefficient $\delta_{q,i\to j}$ has units of houses per trader dollar. Each individual trader optimizes their value function $W_{q,i}$ by choosing how many houses to buy in each MSA j:

$$\begin{aligned} W_{q,i} &= \sum_{j=1}^{I} W_{q,i\to j} \\ W_{q,i\to j} &= \max_{\vartheta_{q,i\to j}} \mathrm{E}\left[(V_j - P_j) \cdot \vartheta_{q,i\to j} | V_j\right] \end{aligned} \tag{8}$$

Definition (Equilibrium)**.** *An equilibrium consists of price parameters* (α^*, β^*) *and demand parameters* $\{(\gamma^*_{q,i\to j}, \delta^*_{q,i\to j})\}$ *for each trader over every ordered MSA pair such that:*

(1) *Given market makers follow the pricing rule in Equation* (6)*, the housing demand schedule* $\{\vartheta_{q,i\to j}\}_{i,j\in I}$ *dictated by the demand rule parameters* $\{(\gamma^*_{q,i\to j}, \delta^*_{q,i\to j})\}_{i,j\in I}$ *solves each trader's optimization problem in Equation* (8)*.*

(2) *Given all traders follow the demand rules specified in Equation* (7)*, the price parameters* (α^*, β^*) *satisfy the expectations equality in Equation* (6)*.*

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) 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 $\theta_{i\to j}$ as well as the price impact of each trader's demand decisions $(\gamma_{i\to j}, \delta_{i\to j})$ 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*

[16] This is the standard ansatz for Kyle (1985) type models and can easily be verified in equilibrium.

number of houses in MSA j:

$$\bar{\theta}_j = \theta_{1\to j} = \theta_{2\to j} = \cdots = \theta_{I\to j} \tag{9}$$

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 $i = 1, 2, \ldots, I$. 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. While full information is perhaps the most natural benchmark, note that the symmetry in Proposition 1 still holds if traders are not fully informed but instead similarly misinformed. For instance, if potential second house buyers in every MSA all over-valued housing in Phoenix by 10%, then traders in all MSAs would still demand the same number of houses in Phoenix—this common demand per trader would just be too high.

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 assuming that traders in MSA i think that all other traders share the same beliefs. 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:*

$$\tilde{P}_j^{(i)} - P_j = \left(\frac{Q_i}{\sum_{i'=1}^{I} Q_{i'}} \right) \cdot \frac{\eta}{2} \tag{10}$$

This proposition is easiest to interpret via a short numerical example. Suppose that there are 55×10^6 traders split across 10 MSAs with the largest MSA i' containing 10×10^6 traders and the smallest MSA i'' containing only 1×10^6 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:

$$\tilde{P}_j^{(\text{MSA})} - P_j = \begin{cases} \left(\frac{10\times10^6}{55\times10^6}\right)\cdot\frac{\$5000}{2} = \$454.55 & \text{if MSA} = i' \\ \left(\frac{1\times10^6}{55\times10^6}\right)\cdot\frac{\$5000}{2} = \$45.45 & \text{if MSA} = i'' \end{cases} \tag{11}$$

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. **Empirical Strategy.** The goal of this simple model is to provide a scaffolding within which to better understand the empirical strategy we employ. With this goal in mind, we now map the empirical setting described in Section 2 onto this model. 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 and IAR appreciation rates. Within the model, this task is tantamount to testing to see if housing appears overpriced—i.e., that $P_j/\mathrm{E}[P_j] > 1$ or $\log P_j - \log \mathrm{E}[P_j] > 0$ after taking logs—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 and IAR appreciation rates. While the model is cast in levels, in the empirical implementation we study $\log P_{j,t} - \log P_{j,t-1}$ in place of $\log P_j - \log \mathrm{E}[P_j]$ under the assumption that $\mathrm{E}[P_j] = P_{j,t-1}$.

Finally, we must address the issue of reverse causality. 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 (from the point of view of an econometrician) 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.

In Proposition 2 we show that if out of town second house buyers from MSA i have a belief distortion η 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 house price and IAR appreciation rates 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 local news sources or word of mouth) is contributing to the realized price distortion.

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 Subsection 4.1, we show that out of town second house buyers are likely less informed about local market conditions relative to local second house buyers and owner occupants. Supporting the claim, we show out of town second house buyers earned lower capital gains on their second house purchases in MSAs such as Las Vegas, Phoenix and Miami relative to local second house buyers who were better able to time the market. Of course, returns are composed of both capital gains and dividends. In Subsection 4.2, we then argue that out of town second house buyers are either less able or less motivated to consume the dividend generated by their housing purchase. Thus, the expected return calculations of distant speculators likely depend more on their beliefs about future house price appreciation rates.

4.1. **Informational Disadvantage.** Out of town second house buyers resemble uninformed or overconfident traders relative to local second house buyers or owner occupants. By definition, out of town second house buyers live farther away from the houses they have purchased than local second house buyers or owner occupants. Thus, these traders don't "know the neighborhood" as well as local buyers. 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) find that real estate agents have substantial discretion in the timing and pricing of house sales with brokers receiving about 3.7% more than other local owner occupants when selling their own houses. Out of town second house buyers with higher monitoring costs likely face an even larger distortion.

As more direct evidence, we show that out of town second house buyers are less successful in timing their exit from the market when compared to local second house buyers. Figure 1 shows the average realized capital gains 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. We compute this capital gain 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 assign observations that are right censored the house price appreciation rate from t to $t + \bar{\tau}$. The width of the out of town second house buyers line is scaled to represent the number of distant speculator purchases in MSA i in month t as a fraction of all sales in units of percent.

In key markets such as Las Vegas, Phoenix, Miami and Tampa, out of town second house buyers earned lower capital gains on their investments relative to local second house buyers. For instance, distant speculators purchasing in Las Vegas in Mar. 2004 earned an 8%/yr capital gain on average; whereas, local speculators buying houses in the same month earned a 17%/yr capital gain on average. In addition, the average capital gain 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. While distant speculators realized 3%/yr lower capital gains than local speculators in Las Vegas during the entire sample period, this gap is largest for buyers who bought near the peak of the housing boom in Las Vegas. These patterns exist only for "boom" markets and are either absent or reversed in other markets such as San Francisco or Cleveland which traditionally have either very cyclical or very flat house price appreciation rates.

Since both distant and local speculators bought houses at the same time in Figure 1, the differences in capital gains earned by each group of traders must stem from differences in exit timing. Put differently, the figure suggests that local second house buyers in markets such as Las Vegas, Phoenix, Miami and Tampa were better able to time the market downturn that distant second house buyers. To quantify this intuition, we estimate the regression specification in Equation (12) below which captures the extent to which distant and local speculators were able to recognize the most appropriate time to sell their house prior to the crash. In particular, we estimate the probability that a speculator "flips" their house within 6 months as a function of (a) the buyer type, (b) whether house prices have hit their peak, (c) the extent to which house prices are rising or falling in the upcoming year and (d) the interaction of these terms:

$$\begin{aligned}
F_{n,i,t-6} = \alpha_i &+ \hat{\alpha}_i \cdot 1_n^{\text{Distant}} + \kappa_t + \hat{\kappa}_t \cdot 1_n^{\text{Distant}} + \hat{\kappa} \cdot 1_n^{\text{Distant}} \\
&+ \beta \cdot \Delta \log P_{i,t\to(t+12)} + \hat{\beta} \cdot \Delta \log P_{i,t\to(t+12)} \cdot 1_n^{\text{Distant}} \\
&+ \gamma \cdot 1_{i,t}^{\text{PostPeak}} + \hat{\gamma} \cdot 1_{i,t}^{\text{PostPeak}} \cdot 1_n^{\text{Distant}} \\
&+ \delta \cdot \Delta \log P_{i,t\to(t+12)} \cdot 1_{i,t}^{\text{PostPeak}} + \hat{\delta} \cdot \Delta \log P_{i,t\to(t+12)} \cdot 1_{i,t}^{\text{PostPeak}} \cdot 1_n^{\text{Distant}} \\
&+ \varepsilon_{n,i,t}
\end{aligned} \tag{12}$$

If local speculators are better informed about future house price appreciation rates, then this knowledge should be revealed in their resale timing. These buyers should be more likely to exit the each market immediately before the house price appreciation rate begins to collapse. Naïvely, we might expect that more informed traders would always flip at a higher rate over the interval $(t-6) \to t$ when house price appreciation rates are lower over the interval from $t \to (t+12)$. However, quickly reselling a house is difficult when house prices are collapsing.

Thus, this naïve estimate of a $\beta\%$ response to a $1\%/\text{yr}$ increase in the house price appreciation rate in MSA i from $t \to (t+12)$ is a weighted average of the decline in the flipping rate in order to earn the capital gains and the increase in the flipping rate due to market liquidity. To disentangle these two offsetting effects, we interact the house price appreciation rate in MSA i from month $t \to (t+12)$ with a dummy variable $1_{i,t}^{\text{PostPeak}} \in \{0,1\}$ which is 1 if the house price appreciation rate in MSA i peaked in months $(t-6) \to t$ and house price appreciation rates in MSA i reached at least $20\%/\text{yr}$ to ensure we are not identifying small local peaks, but rather the culmination of a large increase in prices.

Table 7 displays the estimated regression coefficients from Equation (12). In all of our regression specifications with both time and group fixed effects, we report unclustered standard errors as well as standard errors clustered at along both the time and group dimensions. Reporting each of these three values allows both verifies the robustness of the coefficient estimates and also allows readers to diagnose potential problems with the specification as suggested in Petersen (2009).

First, we see that out of town second house buyers are 5% less likely than local second house buyers to resell their house within 6 months over the entire sample. Next, we find that while local second house buyers are 4.3% more likely to flip their house purchase within the 6 months immediately following the peak in local house price appreciation rates, distant second house buyers are only $4.3 - 3.1 = 1.2\%$ more likely to flip their house purchase during this key interval. What's more, a t-test reveals that the point estimate for distant speculators is not statistically different from zero, suggesting that the likelihood of flipping is nearly unchanged for distant speculators immediately after a house price peak. Finally, while local second house buyers are more likely to flip a second house purchase when prices are rising rapidly during the subsequent 12 months, this effect disappears immediately following the peak in house price appreciation rates further suggesting that local speculators are strategically changing their behavior in order to time the market. On the other hand, while out of town second house buyers tend to flip houses more often when house prices are declining during the entire sample, this effect disappears immediately following the peak in house price appreciation

Market Timing: Local vs. Distant Speculators

Dependent Variable: House resells within 6 Months

	Estimate	Std. Error		
Out of Town Second House Buyer	−0.050	0.015	0.019	0.015
Future House Price Apprec. Rate	0.123	0.014	0.031	0.032
Post Peak Resale	0.043	0.010	0.020	0.019
Post Peak × Future Apprec. Rate	−0.150	0.046	0.055	0.041
Out of Town × Future Apprec. Rate	−0.131	0.019	0.018	0.025
Out of Town × Post Peak	−0.031	0.015	0.020	0.021
Out of Town × Post Peak × Future Apprec. Rate	0.112	0.066	0.071	0.090
	Clustering	$\emptyset$	t	i
	N	1390118		
	R^2	0.083		

Table 7. *Estimated coefficients and standard errors from Equation* (12)*. Resale within* 6 *months is defined as one if a house purchase in month* $t-6$ *in MSA* i *resells during the interval* $(t-6,t]$*. Future house price appreciation rate is the house price appreciation rate in MSA* i *over the interval from* $t \to (t+12)$ *in units of percent per year. Post peak is a dummy variable which is* 1 *if the house price appreciation rate in MSA* i *peaked in months* $(t-6,t]$ *and MSA* i*'s house price appreciation rate peak reached* 20%/yr *or more. The regression uses monthly data from Jul.* 2000 *to Jun.* 2008 *on all house sales to local and out of town second house purchases the* 21 *MSAs weighted by the number of second house purchases in each MSA in each month. Fixed effect estimates of* α_i*,* $\hat{\alpha}_i$*,* κ_t *and* $\hat{\kappa}_t$ *are omitted for clarity. Standard errors are estimated three different ways to account for clustering over time and clustering across MSAs respectively.*

rates. Taken together, the evidence presented in Table 7 suggests that distant speculators are not using insights about future house price appreciation rates to strategically exit their investments in the local housing market to the same extent as local speculators.[17]

4.2. **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 than a local second house buyer or an owner occupant.

We first examine out of town purchasers who use the house only for weekends, holidays, and vacations. Part time residents can only consume the dividend (e.g., live in the house)

[17] These findings are broadly consistent with the results in Bayer et al. (2011), who suggest that house flipping in Los Angeles may have contributed to an increase in house price appreciation rates in that market.

for the portion of each year that they live in the house and thus get lower use than an owner occupant. One might be concerned about preference heterogeneity; perhaps 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. Yet, the data do not appear to be consistent with this hypothesis. For example, Figure 2 shows that the typical out of town house buyer is not a very rich household for whom such a loss of use might be immaterial. To better understand the socioeconomic status of distant speculators, we examine the price of the house that is the primary residence for distant speculators in the highest income cities including San Francisco, San Jose and New York. In Jan. 2005, the median value of all single family houses purchased in San Francisco was $600k. By contrast, the median value of primary residences of out of town second house buyers who live in San Francisco and bought a second house in another MSA in Jan. 2005 was only $555k. While the value of their primary residence is not a complete characterization of out of town second house buyers' wealth, this evidence suggests that the super-rich are not the only traders buying out of town second houses.

Next, consider buyers who wish to rent out their out of town second house purchase. Out of town purchasers face potentially higher costs of property maintenance, renovation, and rental management. It is costly and difficult to supervise contractors or maintenance people from far away. As a proxy for the full opportunity cost, note that a 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. Finally, 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. Finally, 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.

5. Predictive Regressions

In this section, we address the second empirical challenge and show that an increase in the number of house purchases made by out of town second house buyers in an MSA predicts an increase in both house price and IAR appreciation rates. We begin in Subsection 5.1 by comparing the roles of local and distant speculators in the US residential housing market. Then, in Subsection 5.2 we show that increases in demand from out of town second house buyers predicts increases in house price appreciation rates. In Subsection 5.3 we conclude by extending this analysis to look at the impact of out of town second house buyer demand shocks on IAR appreciation rates.

5.1. **Local vs. Distant Speculators.** We view local second house buyers as agents who are often engaging in a similar trade as out of town second house buyers, but who are better informed about future house price appreciation rates. To be clear, there are a variety of differences between the two groups. For instance, distant speculators are more likely to live in their second house part time, while local speculators are more likely to rent their second house out as a source of income.

Out of town second house buyers may potentially be interested in diversification benefits from purchasing a second house in a market where returns are less correlated with other assets in the portfolio.[18] 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 people who already own a house, purchases of stocks or bonds might generate even more diversification with lower trading and holding costs, so any portfolio benefits from purchasing out of town housing are likely limited.

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. Such excessively certain traders may actually seek out investments in fast appreciating markets. For instance,

[18] See Lustig and Van Nieuwerburgh (2005) for empirical evidence on the existence of a housing capital risk premia due to the covariance of its returns with the returns to the household's human capital.

De Long et al. (1991) 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... "

After acknowledging these differences, however, the fact remains that both groups of traders are less able to consume the dividend stream from their second house purchase relative to owner occupants and thus both groups are more reliant on capital gains to earn positive returns on their investments. Thus, we think of local second house buyers as a somewhat comparable "control group" of speculative traders who are more informed than out of town second house buyers when examining the predictive power of speculator demand shocks on future house price and IAR appreciation rates.

5.2. **House Price Appreciation Rate Regressions.** We estimate a panel VAR characterizing the relationship between the house price appreciation rate in an MSA i from month t to $t+1$ and the numbers of local and out of town second house purchases as a percent of sales in MSA i in month t using a panel data set indexed by MSA and month.

The state vector $\mathbf{Y}_{i,t}$ contains the house price appreciation rate from month $(t-1) \to t$ in MSA i as well as the fraction of all house purchases in MSA i in month t that were made by out of town and local second house buyers respectively:

$$\mathbf{Y}_{i,t} = \begin{bmatrix} \Delta \log P_{i,(t-1)\to t} & \frac{S^{\text{Distant}}_{i,t}}{X_{i,t}} & \frac{S^{\text{Local}}_{i,t}}{X_{i,t}} \end{bmatrix}^{\top} \tag{13}$$

The omitted category is the fraction of sales made by owner occupants. Using this state vector, we study the regression specified in Equation (14) below:

$$\mathcal{E}_{i,t} = (\mathbf{I} - \boldsymbol{\Theta}\mathrm{L}_1)(\mathbf{Y}_{i,t} - \mathbf{A}_i - \mathbf{K}_t) \tag{14}$$

In this representation, $\mathbf{I}$ denotes a 3×3 identity matrix, $\boldsymbol{\Theta}$ denotes the 3×3 transition matrix, L_1 denotes the 1 month lag operator, $\mathbf{A}_i$ and $\mathbf{K}_t$ denote 3×1 vectors of MSA and month specific fixed effects and $\mathcal{E}_{i,t}$ denotes a 3×1 vector of error terms.

We report the point estimates and standard errors for the elements of the $\boldsymbol{\Theta}$ transition matrix in Table 8. Panel (a) of Table 8 reveals that a 1% increase in the number of out of town

Predictive Regressions

(a) Dependent Variable: House Price Appreciation Rate

	Estimate	Std. Error		
Lagged House Price Apprec. Rate	0.864	0.012	0.022	0.025
Lagged Distant Spec. Fraction	0.022	0.007	0.009	0.009
Lagged Local Spec. Fraction	−0.010	0.004	0.006	0.007
	Clustering	$\emptyset$	t	i
	N	1995		
	R^2	0.75		

(b) Dependent Variable: Distant Speculator Fraction

	Estimate	Std. Error		
Lagged House Price Apprec. Rate	0.084	0.020	0.027	0.023
Lagged Distant Spec. Fraction	0.872	0.011	0.018	0.016
Lagged Local Spec. Fraction	0.012	0.007	0.007	0.014
	Clustering	$\emptyset$	t	i
	N	1995		
	R^2	0.80		

(c) Dependent Variable: Local Speculator Fraction

	Estimate	Std. Error		
Lagged House Price Apprec. Rate	0.116	0.038	0.045	0.035
Lagged Distant Spec. Fraction	0.074	0.021	0.027	0.032
Lagged Local Spec. Fraction	0.782	0.014	0.020	0.029
	Clustering	$\emptyset$	t	i
	N	1995		
	R^2	0.66		

Table 8. *Parameter values and standard errors of the transition matrix* $\mathbf{\Theta}$ *specified in Equation* (14) *estimated using three panel regressions on monthly data for the* 21 *MSAs from Feb.* 2000 *to Dec.* 2007*. Fixed effect estimates of* $\mathbf{A}_i$ *and* $\mathbf{K}_t$ *are omitted for clarity. Standard errors are estimated three different ways to account for clustering over time or across MSAs.*

second house purchases as a fraction of all purchases in an MSA i in month t is associated with a 0.02%/mo increase in the rate of house price appreciation. To get a better sense of the size of this relationship at the yearly horizon in the presence of the other variables, we compute the cumulative change in the house price appreciation in Phoenix in response to a 3 standard deviation increase in the fraction of purchases made by out of town second house buyers via an impulse response calculation. This 3 standard deviation increase matches the observed change distant speculator demand in Phoenix just prior to its sudden rise in house price appreciation rates. We find that this 3 standard deviation increase in the fraction of all purchases in Phoenix made by distant speculators corresponds to a 5.22%/yr increase in the annual house price appreciation rate in Phoenix. This change corresponds to roughly one sixth of the realized 30%/yr jump in the annual house price appreciation rate in Phoenix during the boom. What's more, this estimated response is likely biased downward since the group fixed effects tend to explain absorb too much variation in panel VARs with short time series as suggested in Nickell (1981).[19]

Note that while an increase in the fraction of purchase made by out of town second house buyers predicts and increase in house price appreciation rates over the next year, an increase in the fraction of purchases made by local second house buyers has a *negative* but statistically insignificant effect. Thus, it is clear that not all second house buyers have the same price impact in this market.

5.3. **IAR Appreciation Rate Regressions.** The results in the previous subsection suggest that an increase in the fraction of purchases made by distant speculators in a given month predicts an increase in the house price appreciation rate in the subsequent month. However, large price movements do not necessarily indicate mispricing; instead, these movements in price could be due to fluctuations in housing market fundamentals. In order to address this concern, we augment our analysis in the previous subsection with a similarly specified panel VAR regression using the monthly IAR appreciation rate rather than the monthly house price appreciation rate. We report these estimates in Table 9.

Comparing Panel (a) in Tables 8 and 9 reveals that the predictive power of an increase in the fraction of all purchases in a given month made by out of town second house buyers is nearly four times larger when examining IAR appreciation rates rather than house price appreciation rates. This evidence suggests that distant speculator demand shocks appreciably distort the own vs. rent calculus of people living in the target MSA. As before, we find very little effect that an increase in purchases by local second house buyers affects mispricing. The

[19]For intuition, recall that in finite samples principle component analysis over-estimates the size of the first principle component and leaves too little variation to be explained by subsequent factors. Similarly, by picking the group fixed effects that best explain the average level of the group, panel VARs on data with a short time series dimension tend to assign too much of the variation across groups to the group fixed effects $\mathbf{A}_i$ and leave too little to be explained by the transition matrix $\mathbf{\Theta}$.

Predictive Regressions

(a) Dependent Variable: IAR Appreciation Rate

	Estimate	Std. Error		
Lagged IAR Apprec. Rate	0.480	0.021	0.075	0.088
Lagged Distant Spec. Fraction	0.080	0.014	0.015	0.026
Lagged Local Spec. Fraction	0.000	0.009	0.011	0.009
	Clustering	$\emptyset$	t	i
	N	1995		
	R^2	0.26		

(b) Dependent Variable: Distant Speculator Fraction

	Estimate	Std. Error		
Lagged IAR Apprec. Rate	0.047	0.017	0.019	0.017
Lagged Distant Spec. Fraction	0.878	0.011	0.018	0.017
Lagged Local Spec. Fraction	0.011	0.008	0.007	0.015
	Clustering	$\emptyset$	t	i
	N	1995		
	R^2	0.80		

(c) Dependent Variable: Local Speculator Fraction

	Estimate	Std. Error		
Lagged IAR Apprec. Rate	0.085	0.032	0.033	0.036
Lagged Distant Spec. Fraction	0.080	0.021	0.026	0.033
Lagged Local Spec. Fraction	0.780	0.014	0.020	0.029
	Clustering	$\emptyset$	t	i
	N	1995		
	R^2	0.66		

Table 9. *Parameter values and standard errors of the transition matrix* $\mathbf{\Theta}$ *specified in Equation* (14) *estimated using three panel regressions on monthly data for the* 21 *MSAs from Feb.* 2000 *to Dec.* 2007 *but using the IAR appreciation rate rather than the house price appreciation rate. Fixed effect estimates of* $\mathbf{A}_i$ *and* $\mathbf{K}_t$ *are omitted for clarity. Standard errors are estimated three different ways to account for clustering over time or across MSAs. The coefficient in Panel* (a) *"Lagged Local Spec. Fraction" is* -0.00005.

predictive power of an increase in the fraction of purchases made by these local speculators is a tightly estimated zero in Table 9.

Computing the response to the same 3 standard deviation increase in the fraction of purchases made by out of town second house buyers on IAR appreciation rates in Phoenix rather than house price appreciation rates, we find that this shock also explains around one sixth of the increase in mispricing in Phoenix. This estimate remains relatively unchanged even though the point estimates on the lagged distant speculator purchase fraction in Table 9 are much larger than those in 9 because the IAR appreciation rate is substantially less predictable than the house price appreciation rate. The autoregressive coefficient falls from 0.864 in Table 8 to 0.480 in Table 9, while the adjusted R^2 drops from 75.4% to only 26.4%. As a check on the stability of the estimation procedure, we also note that switching from house price appreciation rates to IAR appreciation rates leaves the point estimates in Panels (*b*) and (*c*) of Table 9 nearly unchanged as compared to Table 8.

6. Reverse Causality

In this section, we address the third and most difficult empirical challenge—the issue of reverse causality. We distinguish between two hypotheses: the null hypothesis that distant speculators are reacting to a common signal about the target MSA and the alternative hypothesis that distant speculator purchase decisions are not entirely driven by unobserved common signals. Our key observation is that, if the null hypothesis is true and distant speculators were reacting in the mid-2000s to a positive shock to housing values in Las Vegas that we cannot observe as econometricians, then out of town second house buyers living in each other MSA should have increased their demand for housing in Las Vegas in equal proportions once we control for MSA pair specific factors such as distance and information transmission as spelled out in Proposition 1. On the other hand, if the null hypothesis is not true, then this proportional symmetry should not hold. In particular, increases in out of town second house buyer demand from largest markets should lead to the largest increases in house price and IAR appreciation rates in the target market as demonstrated in Proposition 2.

Subsection 6.1 describes how we empirically implement the model outlined in Section 3. We frame the predictions of this model as a pair of equations that we estimate in Subsection 6.2. We find evidence against an explanation based solely on reverse causality, but consistent with a causal effect of distant speculators on MSA house prices and IAR appreciation rates.

Finally, in Subsection 6.3, we consider additional specifications to investigate the robustness of the main results.

6.1. **Empirical Predictions.** To test the predictions from Section 3, we need to measure the number of distant speculators living in each MSA i every month t 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.

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 $T = 96$:*

$$Q_i = \frac{1}{96} \cdot \sum_{t=1}^{96} \left(\sum_{i \neq j} S_{i \to j,t} \right) \tag{15}$$

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 $\theta_{i \to j}$. We estimate all regression equations in this section using a panel dataset at a monthly frequency from Feb. 2000 to Dec. 2007 on the $21 \times 20 = 420$ ordered MSA pairs with all $i = j$ pairs removed. Observations from Jan. 2000 are removed due to the missing 1mo lagged values yielding a balanced panel of $39,900$ observations.

Definition (Speculator Share)**.** *Let $\theta_{i \to j,t}$ 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:*

$$\theta_{i \to j,t} = \frac{S_{i \to j,t}}{Q_i} \tag{16}$$

where $\theta_{i \to j,t}$ has units of houses per trader.

Using these variables, we estimate Equation (17), which studies the relationship between the house price appreciation rate from time t to time $t+1$ 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 γ on the variable $\theta_{i\to j,t}$:

$$\Delta \log P_{j,t\to(t+1)} = \beta \cdot \Delta \log P_{j,(t-1)\to t} + \gamma \cdot \theta_{i\to j,t} + \alpha_{i\to j} + \kappa_t + \varepsilon_{i\to j,t}, \qquad i \neq j \tag{17}$$

The ordered MSA pair dummy variables control for two key effects as displayed in Equation (18) below:

$$\alpha_{i\to j} = \bar{\alpha}_j - \gamma \cdot \mathrm{E}[\theta_{i\to j,t}] \tag{18}$$

First, each $\alpha_{i\to j}$ accounts for the mean house price appreciation rate $\bar{\alpha}_j$ in each MSA j over the sample period (or the mean IAR appreciation rate over the same time period). Second, each $\alpha_{i\to j}$ adjusts the predicted house price appreciation rate (or IAR appreciation rate) in MSA j for the average rate at which second house buyers living in MSA i purchase second houses in MSA j. For instance, $\gamma \cdot \mathrm{E}[\theta_{(\mathrm{SFO},j),t}]$ differentially controls for the tendency of distant speculators living in San Francisco to purchase more second houses in Phoenix rather than in Milwaukee:

$$\mathrm{E}[\theta_{(\mathrm{SFO},\mathrm{PHX}),t}] \neq \mathrm{E}[\theta_{(\mathrm{PHX},\mathrm{SFO}),t}] \neq \mathrm{E}[\theta_{(\mathrm{SFO},\mathrm{MIL}),t}] \tag{19}$$

We also estimate the specification outlined in Equation (17) replacing the house price appreciation rate in MSA j from time t to time $t+1$ with the IAR appreciation rate from time t to $t+1$. Consistent with the results in Section 5, we expect to estimate a positive γ for both specifications indicating that, for instance, IAR appreciation rates rise by $\gamma\%/\mathrm{mo}$ 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 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. We do this by including an interaction between the number of second house buyers in MSA i and the proportion of these speculators buying houses in MSA j. 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_i^{Small} denotes one of the seven MSAs with the smallest number of distant speculators, 1_i^{Medium} denotes the next seven MSAs with a moderate number of distant speculators and 1_i^{Large} denotes one of the seven MSAs with a largest number of distant speculators. We then estimate the regression specification in Equation (20) below where δ_2 and δ_3 have units of houses per person per month:

$$\begin{aligned}\Delta \log P_{j,t\to(t+1)} = \alpha_{i\to j} + \kappa_t + \beta \cdot \Delta \log P_{j,(t-1)\to t} + \gamma \cdot \theta_{i\to j,t} \\ + \ \delta_2 \cdot 1_i^{\text{Medium}} \cdot \theta_{i\to j,t} + \delta_3 \cdot 1_i^{\text{Large}} \cdot \theta_{i\to j,t} + \varepsilon_{i\to j,t}\end{aligned} \qquad i \neq j \tag{20}$$

using both house price appreciation rates and IAR appreciation rates. 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 (a large market) 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 (a medium market) 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 Equation (20), the ordered MSA pair fixed effects control for variation in the mean house price and IAR appreciation rates in MSA j as well as differences across MSAs in the average number of houses demanded in MSA j by speculators living in MSA i. The alternative hypothesis in Proposition 2 states that if out of town second house buyers are causing increases in the house price and IAR appreciation rates, we should find $\delta_3 > \delta_2 > 0$. i.e., that house price and IAR appreciation rates are the highest in MSA j in the month following an increase in the demand per speculator in MSA i when MSA i contains the largest number of potential traders.[20]

At first glance it might appear that this relationship is mechanical. To see why this is not the case, consider a short example based on the insights from Section 3 where 100k traders living in Los Angeles and 10k traders living in Milwaukee consider whether or not to buy a second house in Las Vegas. First, suppose that the null hypothesis is true and a common

[20]This identification strategy is analogous to the front door criterion as outlined in Pearl (2000).

signal about the value of housing in Las Vegas that we can't observe as econometricians drives distant speculator purchase decisions. In such a world, if the signal warrants a 10% increase in the fraction of distant speculators that should purchase a second house in Las Vegas, we should see an 11k increase in demand for houses in Las Vegas—10k from distant speculators living in Los Angeles and 1k from distant speculators living in Milwaukee. Under the null hypothesis, a 10% increase in the fraction of distant speculators arriving from Los Angeles will be equally predictive of a rise in house price appreciation rates in Las Vegas as a 10% increase in the fraction of distant speculators arriving from Milwaukee because each will coincide with an 11k increase in demand. On the other hand, if the alternative hypothesis is true and distant speculators are not simply reacting to a common signal, then changes in demand by distant speculators in Los Angeles and Milwaukee will not generally coincide. In this world, a 10% increase in the fraction of distant speculators arriving from Los Angeles will predict a 10k trader demand shock while a 10% increase in the fraction of distant speculators arriving from Milwaukee will predict only a 1k trader demand shock. Thus, it is only under the alternative hypothesis that we should see a 1% increase in the demand per trader from a large MSA lead to a larger effect relative to a 1% increase in the demand per trader from a small MSA.

6.2. **Estimation Results.** Panel (a) in both Table 10 and 11 reports the estimated coefficients and standard errors from Equation (17) using both price and IAR appreciation rates as the dependent variable and indicates that γ is both positive and statistically significant. The point estimate for γ in Table 10 implies that a 1% increase in the number of houses demanded in MSA j per distant speculator living in MSA i predicts a $0.213 \times 12 = 2.556\%$ increase in the house price appreciation rate in MSA j over the next year. Similarly, in Table 11, a 1% increase in the number of houses in MSA j demanded per distant speculator in MSA i results in a $0.769 \times 12 = 9.228\%$ increase in the IAR appreciation rate over the next year, suggesting mispricing grows when distant speculator demand grows.

Next, looking at Panel (b) in both Tables 10 and 11 we see that the coefficient δ_3 in both Equation (20) is statistically different from zero in violation of the symmetry predicted by

Reverse Causality: Baseline Specification

(a) Dependent Variable: House Price Appreciation Rate

	Estimate	Std. Error		
Lagged House Price Appreciation Rate	0.853	0.003	0.021	0.005
Distant Speculator Share	0.213	0.024	0.059	0.035
	Clustering	$\emptyset$	t	$i \to j$
	N	39900		
	R^2	0.753		

(b) Dependent Variable: House Price Appreciation Rate

	Estimate	Std. Error		
Lagged House Price Appreciation Rate	0.869	0.003	0.021	0.006
Distant Speculator Share	0.052	0.038	0.052	0.040
Medium MSA $\times$ Dist. Speculator Share	0.253	0.055	0.058	0.057
Large MSA $\times$ Dist. Speculator Share	0.318	0.055	0.072	0.090
	Clustering	$\emptyset$	t	$i \to j$
	N	39900		
	R^2	0.755		

Table 10. *Panel (a): Coefficient estimates from Equation (17). Panel (b): Coefficient estimates from Equation (20). All regressions use monthly data from Feb. 2000 to Dec. 2007 on the 420 ordered MSA pairs with all $i = j$ pairs removed. Fixed effect estimates of $\alpha_{i \to j}$ and κ_t are omitted for clarity. Standard errors are estimated three different ways to account for clustering over time or across ordered MSA pairs.*

Proposition 1. The impact on house price appreciation rates of a 1% increase in the number of house demanded per distant speculator living in a large MSA is almost twice as large as that of a 1% increase in the number of houses demanded per distant speculator living in a small MSA. As well, the ordering of the interaction terms is consistent with the alternative hypothesis that demand from distant speculators causes house price and IAR appreciation rates to increase. In all specifications $\delta_3 \geq \delta_2 \geq 0$. We can interpret the coefficients γ, δ_2 and δ_3 reported in Panel (b) of Table 10 as saying that while a 1% increase in the number of houses demanded in MSA j per trader in MSA i predicts an $0.052 \times 12 = 0.624\%$ increase in the house price appreciation rates 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 $(0.052 + 0.318) \times 12 = 4.440\%$ 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.

Reverse Causality: Baseline Specification

(a) Dependent Variable: IAR Appreciation Rate

	Estimate	Std. Error		
Lagged IAR Appreciation Rate	0.505	0.005	0.073	0.020
Distant Speculator Share	0.769	0.049	0.115	0.090
	Clustering	$\emptyset$	t	$i \to j$
	N	39900		
	R^2	0.254		

(b) Dependent Variable: IAR Appreciation Rate

	Estimate	Std. Error		
Lagged IAR Appreciation Rate	0.503	0.005	0.073	0.020
Distant Speculator Share	0.356	0.079	0.088	0.100
Medium MSA $\times$ Dist. Speculator Share	0.493	0.115	0.117	0.173
Large MSA $\times$ Dist. Speculator Share	0.821	0.115	0.169	0.243
	Clustering	$\emptyset$	t	$i \to j$
	N	39900		
	R^2	0.255		

Table 11. *Panel (a): Coefficient estimates from Equation (17) using IAR appreciation rates rather than price appreciation rates as the dependent variable. Panel (b): Coefficient estimates from Equation (20) using IAR appreciation rates rather than price appreciation rates as the dependent variable. All regressions use monthly data from Feb. 2000 to Dec. 2007 on the 420 ordered MSA pairs with all $i = j$ pairs removed. Fixed effect estimates of $\alpha_{i \to j}$ and κ_t are omitted for clarity. Standard errors are estimated three different ways to account for clustering over time or across ordered MSA pairs.*

6.3. **Robustness Checks.** In this subsection, we discuss possible ways that the regression specified in Equation (20) might lead to spurious conclusions and then describe how we address these issues.

First, we observe that, while the baseline specification assumes that the house price and IAR appreciation rates in all MSAs realize a common shock κ_t in each month, macroeconomic forces during our sample period likely affected distant speculators living in different MSAs in different ways. For example, potential second house buyers living in New York City might always have more accurate beliefs about the fundamental value of housing than potential second house buyers living in Milwaukee. In Table 12 we re-run the specifications in Equation (20) using home MSA by month rather than simply month fixed effects to account for this concern. Thus, the $\kappa_{i,t}$ terms capture the time varying effect of shocks to different distant

Reverse Causality: (i,t) and $(i \to j)$ Fixed Effects

(a) Dependent Variable: House Price Appreciation Rate

	Estimate	Std. Error		
Lagged House Price Appreciation Rate	0.852	0.003	0.023	0.008
Distant Speculator Share	0.039	0.030	0.020	0.029
Medium MSA × Dist. Speculator Share	0.214	0.043	0.037	0.031
Large MSA × Dist. Speculator Share	0.404	0.040	0.032	0.053
	Clustering	$\emptyset$	(i,t)	$i \to j$
	N		39900	
	R^2		0.754	

(b) Dependent Variable: IAR Appreciation Rate

	Estimate	Std. Error		
Lagged IAR Appreciation Rate	0.511	0.006	0.080	0.021
Distant Speculator Share	0.297	0.080	0.097	0.097
Medium MSA × Dist. Speculator Share	0.390	0.111	0.130	0.165
Large MSA × Dist. Speculator Share	0.787	0.171	0.203	0.210
	Clustering	$\emptyset$	(i,t)	$i \to j$
	N		39900	
	R^2		0.257	

Table 12. *Panel* (a)*: Coefficient estimates from Equation* (20) *with* (i,t) *and* $i \to j$ *fixed effects. Panel* (b)*: Coefficient estimates from Equation* (20) *with* (i,t) *and* $i \to j$ *fixed effects using IAR appreciation rates rather than price appreciation rates as the dependent variable. All regressions use monthly data from Feb.* 2000 *to Dec.* 2007 *on the* 420 *ordered MSA pairs with all* $i = j$ *pairs removed. Fixed effect estimates of* $\alpha_{i \to j}$ *and* $\kappa_{i,t}$ *are omitted for clarity. Standard errors are estimated three different ways to account for clustering over home MSA by time or across ordered MSA pairs.*

speculator home MSAs. We find that including home MSA by month fixed effect, if anything, actually strengthens our results and interpret these findings as evidence that unobserved variation in the distant speculator populations in different home MSAs is unlikely to be driving our results.

In order for this sort of variation to confound our results, it would need to be the case that, for example, potential second house buyers in New York City got differentially better information about the fundamental value of housing in Miami than potential second house buyers living in Milwaukee in an extremely precise way: (a) New Yorkers would need to receive extremely good information about buying a second house in Miami during the period from 2004 through 2006 when Miami realized its highest house price appreciation rates, and (b) this information could only have applied to Miami and other cities that New Yorkers might

Reverse Causality: Jan. 2000 to Dec. 2000 Ranking Period

(a) Dependent Variable: House Price Appreciation Rate

	Estimate	Std. Error		
Lagged House Price Appreciation Rate	0.869	0.003	0.021	0.006
Distant Speculator Share	0.069	0.015	0.021	0.025
Medium MSA × Dist. Speculator Share	−0.012	0.020	0.013	0.032
Large MSA × Dist. Speculator Share	0.142	0.028	0.036	0.056
	Clustering	$\emptyset$	t	$i \to j$
	N	39900		
	R^2	0.754		

(b) Dependent Variable: IAR Appreciation Rate

	Estimate	Std. Error		
Lagged IAR Appreciation Rate	0.504	0.005	0.073	0.020
Distant Speculator Share	0.205	0.033	0.036	0.035
Medium MSA × Dist. Speculator Share	0.019	0.042	0.028	0.069
Large MSA × Dist. Speculator Share	0.436	0.059	0.082	0.149
	Clustering	$\emptyset$	t	$i \to j$
	N	39900		
	R^2	0.255		

Table 13. *Panel (a): Coefficient estimates from Equation (20) with Q_i estimated over the period from Jan. 2000 to Dec. 2000. Panel (b): Coefficient estimates from Equation (20) with Q_i estimated over the period from Jan. 2000 to Dec. 2000 using IAR appreciation rates rather than price appreciation rates as the dependent variable. All regressions use monthly data from Feb. 2000 to Dec. 2007 on the 420 ordered MSA pairs with all $i = j$ pairs removed. Fixed effect estimates of $\alpha_{i\to j}$ and κ_t are omitted for clarity. Standard errors are estimated three different ways to account for clustering over time or across ordered MSA pairs.*

have invested in. Thus, while ordered city pair by month specific correlation between both house price and log IAR appreciation rates and distant speculator shares would confound our results, it is difficult to think of such an explanation.

Second, perhaps the size of second house buyers belief distortions are not fixed as in Section 3, but instead are random variables. In such a world, covariance between the home MSA size and the size of second house buyers belief distortion may bias our results. To address this concern, in Table 13 we again re-run the specifications in Equation (20), only this time we instead compute the number of distant speculators in each MSA using the ranking in 2000. Let $\widehat{Q}_i$ denote the number of distant speculators in MSA i similarly defined but measured over the period from Jan. 2000 to Dec. 2001 so that $T = 12$:

$$\widehat{Q}_i = \frac{1}{12} \cdot \sum_{t=1}^{12} \left(\sum_{i \neq j} S_{i \to j,t} \right) \tag{21}$$

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 predates the rapid rise in house price appreciation rates in all MSAs and minimizes the possibility for correlation between home MSA size and the level of belief distortion.

This specification controls for 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 Tables 10 and 11 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 Table 13, 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 $i \to j$. Finally, we observe that the empirical results are strongest in Panel (b) where we use the IAR appreciation rate as the dependent variable. In all cases the δ_3 point estimates are different from zero. To the extent that the IAR appreciation rate proxies for mispricing, these results present a consistent picture that distant speculators contribute to mispricing.

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 impact of speculative demand in the US residential housing market where we obtain detailed microdata on traders and can exploit the housing market's is natural geographic segmentation since house prices do not follow the same time series pattern and home buyers in different MSAs use different information when making their purchases.

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 appear less well 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 and IAR appreciation rates. 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 and IAR appreciation rates 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 how total out of town purchases compare to the size of the local economy. Figure 3 plots the sum of the sales prices on distant speculator purchases as a percent of gross MSA product, G(MSA)P from 2000 to 2007, where G(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(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 Ursúa (2008) define a 10% drop in the GDP of a country as an economic disaster while Javorcik (2004) 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[21] report found that 1.8Mil households in England owned a second home and, among these properties, 87k 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 4 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.[22] Thus, distant speculators may be an important class of traders playing a role in bubble formation more generally and an interesting topic of future research.

[21]See Office for National Statistics (2007).
[22]See Sagalyn (1999), which discuss the purchase of Rockefeller Center by Mitsubishi Trust, Co. for more than $1Bil in the late 1980.

Appendix A. Proofs

Proof (Proposition 1). Substituting both the functional form for the housing price in MSA j from Equation (6) and the functional form for the aggregate demand in MSA j from Equation (5) into the objective function for an individual trader q from MSA i yields an expression:

$$\begin{aligned} W_{q,i\to j} &= \max_{\vartheta_{q,i\to j}} \mathrm{E}\left[(V_j - \alpha - \beta \cdot X_j)\cdot \vartheta_{q,i\to j} | V_j\right] \\ &= \max_{\vartheta_{q,i\to j}} \mathrm{E}\left[\left(V_j - \alpha - \beta \cdot \sum_{i'=1}^{I}\left(\sum_{q'=1}^{Q_{i'}} \vartheta_{q',i'\to j}\right) - \beta \cdot \varepsilon_j\right)\cdot \vartheta_{q,i\to j}\middle| V_j\right] \end{aligned} \tag{22}$$

Taking the derivative of this optimization program with respect to trader q's demand gives the first order condition:

$$0 = \mathrm{E}\left[\left(V_j - \alpha - \beta \cdot \sum_{i'=1}^{I}\left(\sum_{q'=1}^{Q_{i'}} \vartheta_{q',i'\to j}\right) - \beta \cdot \varepsilon_j\right) - 2\cdot\beta\cdot \vartheta_{q,i\to j}\middle| V_j\right] \tag{23}$$

where we assume $Q_i \approx Q_i - 1$ for simplicity. Evaluating the conditional expectation operator yields:

$$0 = V_j - \alpha - \beta \cdot \sum_{i'=1}^{I}\left(\sum_{q=1}^{Q_{i'}} \vartheta_{q,i'\to j}\right) - 2\cdot\beta\cdot \vartheta_{q,i\to j} \tag{24}$$

We then solve for $\vartheta_{q,i\to j}$ to derive the expression below:

$$\vartheta_{q,i\to j} = -\frac{\alpha + \beta \cdot \sum_{i'=1}^{I}\left(\sum_{q'=1}^{Q_{i'}} \vartheta_{q',i'\to j}\right)}{2\cdot\beta} + \left(\frac{1}{2\cdot\beta}\right)\cdot V_j \tag{25}$$

This expression would be identical for any trader q living in MSA $i \in I$ implying that $\theta_{i\to j} = \theta_{i'\to j}$ for all $i, i' \in \{1, 2, \ldots, I\}$.

□

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:

$$\mathrm{E}\left[P_j|\mathrm{MSA}\right] = \begin{cases} \alpha^* + \beta^* \cdot \sum_{i'=1}^{I} Q_{i'} \cdot \left(\bar{\gamma}^* + \bar{\delta}^* \cdot V_j\right) & \text{if MSA} \neq i \\ \alpha^* + \beta^* \cdot \sum_{i'=1}^{I} Q_{i'} \cdot \left(\bar{\gamma}^* + \bar{\delta}^* \cdot \{V_j + \eta\}\right) & \text{if MSA} = i \end{cases} \tag{26}$$

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:

$$\begin{aligned} \tilde{X}_j^{(i)} &= \sum_{i' \neq i} Q_{i'} \cdot \left(\bar{\gamma}^* + \bar{\delta}^* \cdot V_j\right) + Q_i \cdot \left(\bar{\gamma}^* + \bar{\delta}^* \cdot \{V_j + \eta\}\right) \\ &= \sum_{i'=1}^{I} Q_{i'} \cdot \left(\bar{\gamma}^* + \bar{\delta}^* \cdot V_j\right) + Q_i \cdot \bar{\delta}^* \cdot \eta \end{aligned} \tag{27}$$

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 β^* and $\bar{\delta}^*$ from Proposition 1 yields the desired result. □

References

D. Abreu and M. Brunnermeier. Bubbles and crashes. *Econometrica*, 71, 2003.

R. Barro and J. Ursúa. Macroeconomic crises since 1870. *Brookings Papers on Economic Activity*, 2008.

P. Bayer, C. Geissler, and J. Roberts. Speculators and middlemen: The role of flippers in the housing market. *National Bureau of Economic Research Working Paper Series*, 2011.

F. Black. Noise. *The Journal of Finance*, 41, 1986.

M. Brunnermeier and S. Nagel. Hedge funds and the technology bubble. *The Journal of Finance*, 59, 2004.

K. Case and R. Shiller. The efficiency of the market for Single-Family homes. *The American Economic Review*, 79, 1989.

B. De Long, A. Shleifer, L. Summers, and R. Waldmann. Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance*, 45, 1991.

P. Emrath. Property taxes in the 2000 census. *Housing Economics*, 12, 2002.

F. Ferreira and J. Gyourko. Anatomy of the beginning of the housing boom: U.S. neighborhoods and metropolitan areas, 1993-2009. *National Bureau of Economic Research Working Paper Series*, 2011.

M. Flavin and T. Yamashita. Owner-Occupied housing and the composition of the household portfolio. *The American Economic Review*, 92, 2002.

M. Friedman. *Essays in Positive Economics*. University of Chicago Press, 1953.

E. Glaeser and J. Gyourko. Arbitrage in housing markets. *National Bureau of Economic Research Working Paper Series*, 2007.

E. Glaeser, J. Gottlieb, and J. Gyourko. Can cheap credit explain the housing boom? *National Bureau of Economic Research Working Paper Series*, 2010.

R. Greenwood and S. Nagel. Inexperienced investors and bubbles. *Journal of Financial Economics*, 93, 2009.

J. Harding, T. Miceli, and C. Sirmans. Do owners take better care of their housing than renters? *Real Estate Economics*, 28, 2000.

A. Haughwout, D. Lee, J. Tracy, and W. van der Klaauw. Real estate investors, the leverage cycle, and the housing market crisis. *Federal Reserve Bank of New York Staff Reports*, 514, 2011.

C. Himmelberg, C. Mayer, and T. Sinai. Assessing high house prices: Bubbles, fundamentals and misperceptions. *The Journal of Economic Perspectives*, 19, 2005.

G. Hubbard and C. Mayer. The mortgage market meltdown and house prices. *The B.E. Journal of Economic Analysis and Policy*, 9, 2009.

B. Javorcik. Does foreign direct investment increase the productivity of domestic firms? in search of spillovers through backward linkages. *American Economic Review*, 94, 2004.

A. Kyle. Continuous auctions and insider trading. *Econometrica*, 53, 1985.

S. Levitt and C. Syverson. Market distortions when agents are better informed: The value of information in real estate transactions. *Review of Economics and Statistics*, 90, 2008.

W. Li and Z. Gao. Real estate investors and the boom and bust of the us housing market. *SSRN*, 2012.

H. Lustig and S. Van Nieuwerburgh. Housing collateral, consumption insurance and risk premia: an empirical perspective. *The Journal of Finance*, 60, 2005.

C. Mayer. Housing bubbles: A survey. *Annual Review of Economics*, 3, 2010.

S. Nickell. Biases in dynamic models with fixed effects. *Econometrica*, 49, 1981.

E. Ofek and M. Richardson. Dotcom mania: The rise and fall of internet stock prices. *The Journal of Finance*, 58, 2003.

Office for National Statistics. Community housing survey, 2007. URL `http://www.communities.gov.uk/housing/publications/reports/`.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

M. Petersen. Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22, 2009.

J. Poterba. Tax subsidies to Owner-Occupied housing: An Asset-Market approach. *The Quarterly Journal of Economics*, 99, 1984.

L. Sagalyn. *Cases in Real Estate Finance and Investment Strategy*. Urban Land Institute, 1999.

J. Scheinkman and W. Xiong. Overconfidence and speculative bubbles. *The Journal of Political Economy*, 111, 2003.

A. Shleifer and R. Vishny. The limits of arbitrage. *The Journal of Finance*, 52, 1997.

L. Zingales. What determines the value of corporate votes. *The Quarterly Journal of Economics*, 110, 1995.

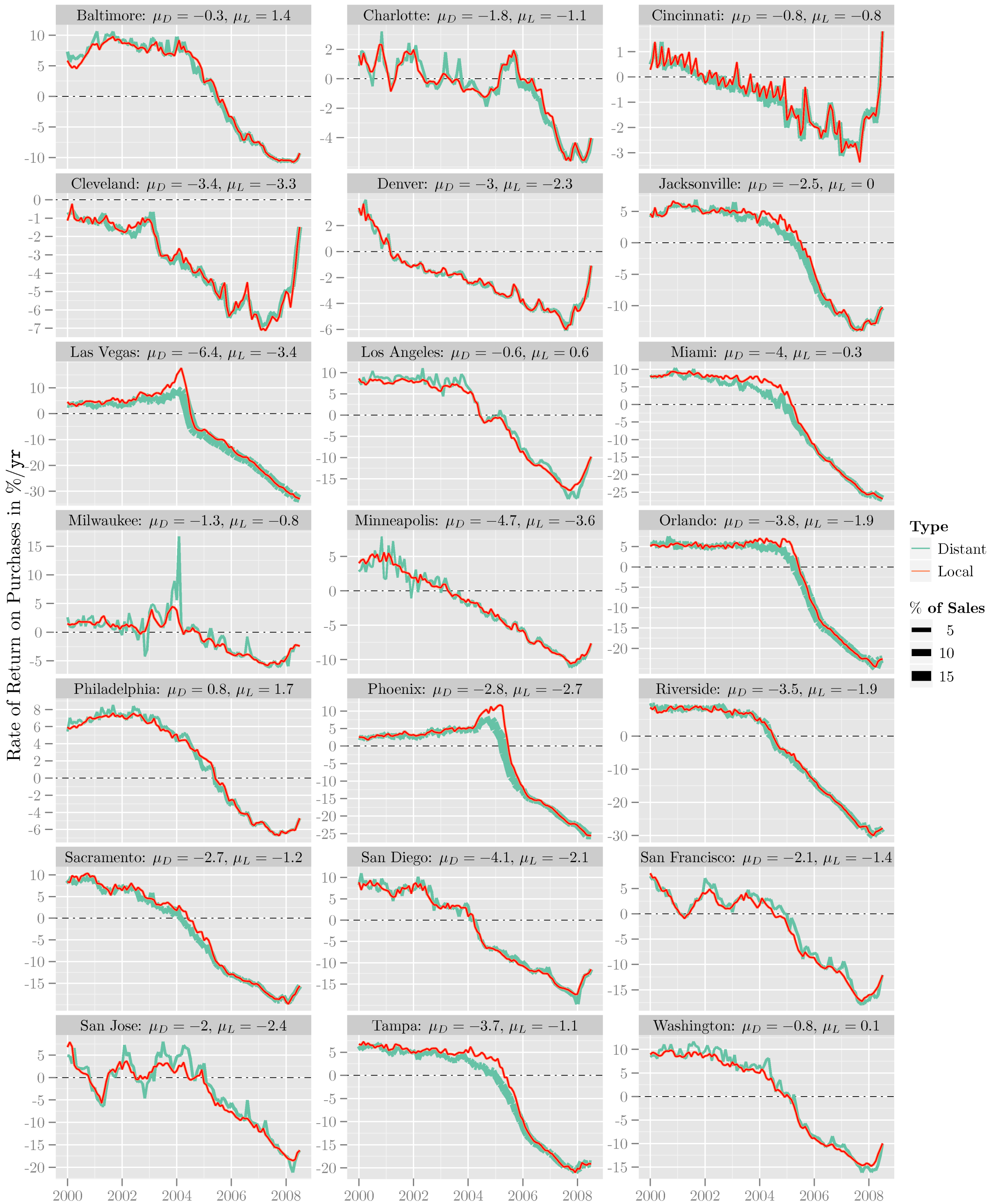


Figure 1. *The capital gain on single family house purchases made by local and distant speculators from Jan.* 2000 *to Dec.* 2007 *in units of* %/yr *using ZIP code by month level house price index data from* Zillow. *The width of each line is scaled by the number of purchases by each buyer type as a fraction of all sales in units of* %. μ_D *and* μ_L *are the mean capital gains 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 *capital gain on average; whereas, local speculators earned a* 17%/yr *capital gain on average in Mar.* 2004. *The average capital gain 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.*"*

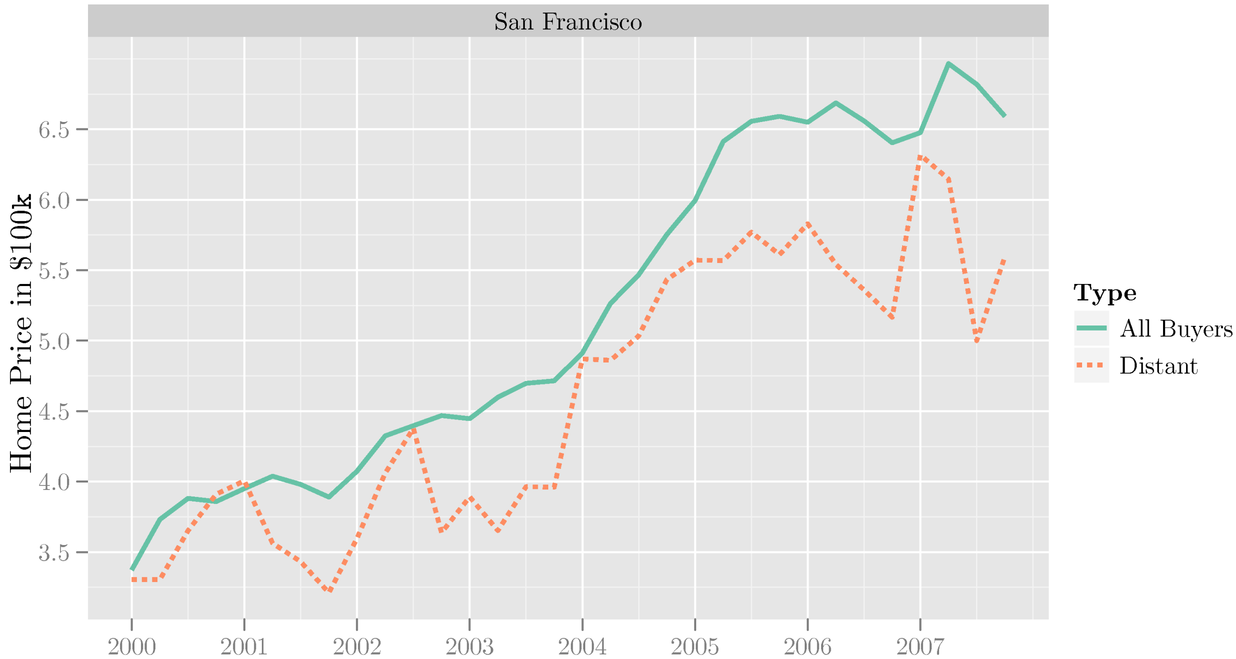


Figure 2. *Median primary residence house price for the populations of distant speculators and of all buyers living in San Francisco in units of* \$100k *over the time period from Jan.* 2000 *to Dec.* 2007. *The price of the primary residences of out of town second house buyers living in San Francisco is computed by scaling up the most recent sale price by the* Zillow *ZIP code level price index. Reads: "In Jan.* 2005, *the median value of all single family houses purchased in San Francisco was* \$600k. *By contrast, the median value of primary residences of out of town second house buyers who live in San Francisco and bought a second house in another MSA in Jan.* 2005 *was only* \$555k.*"*

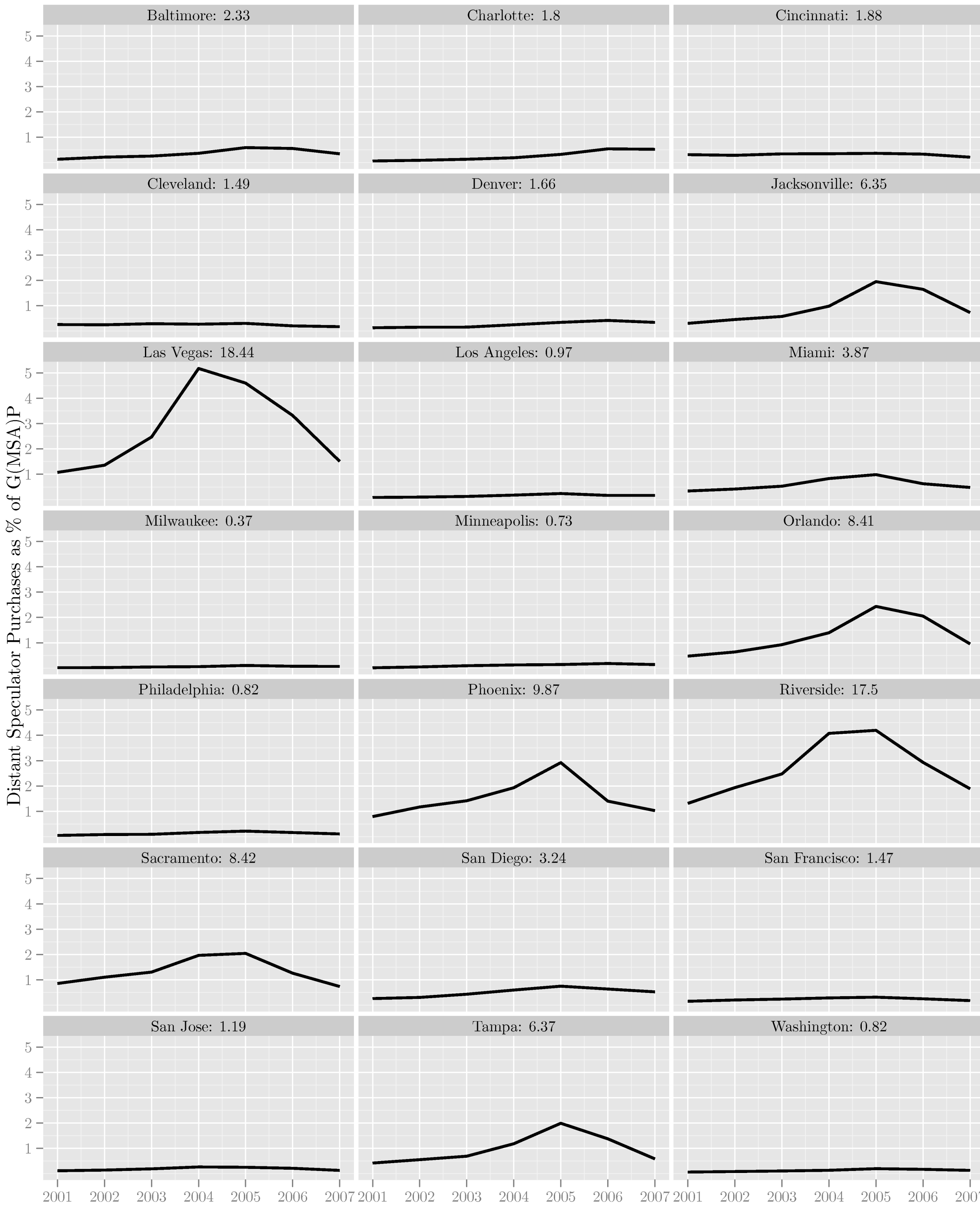


Figure 3. *Sum of the sales prices of single family houses sold to distant speculators as a fraction of total G(MSA)P in each MSA in units of % from* 2000 *to* 2007*. We compute G(MSA)P using data from the BEA as the product of the per capita income in each MSA times the population. The number at the top of each panel represents the sum of the G(MSA)P shares in each MSA from* 2002 *to* 2007*. Reads: "The sum of the sales prices in Las Vegas exceeded* 5% *of the G(MSA)P for the entire MSA in* 2004*."*

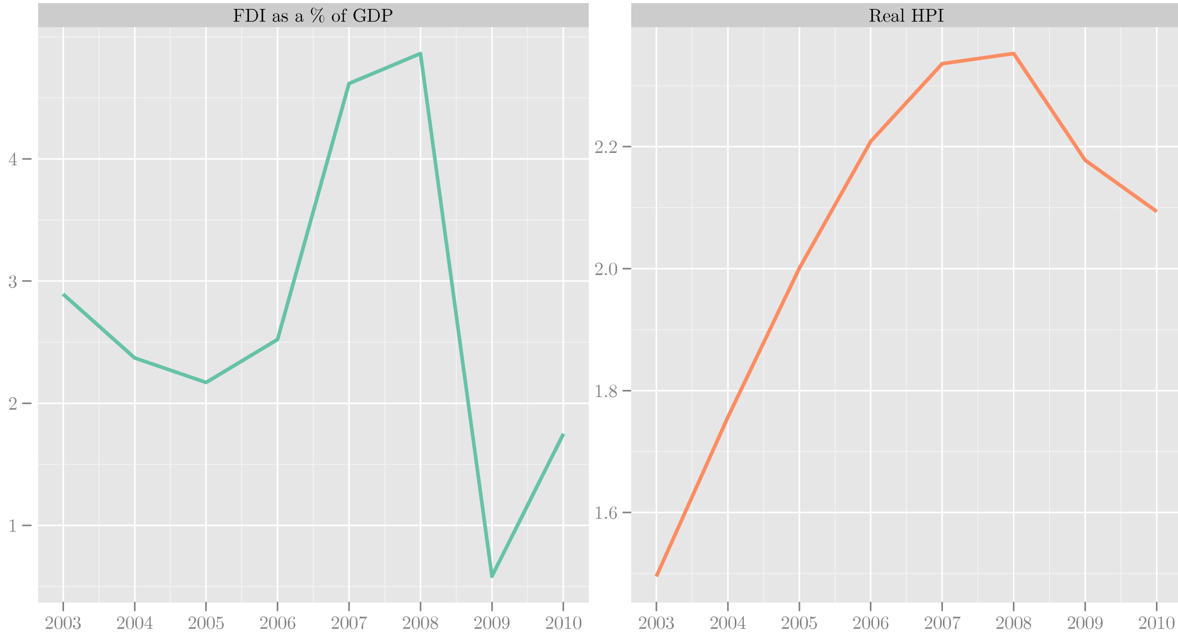


Figure 4. *Left Panel: Net foreign direct investment (henceforth, FDI) in Spain from the World Bank as a percent of Spain's GDP from* 2003 *to* 2010*. Reads: "Net FDI inflows into Spain amounted to a little less than* 5% *of Spain's GDP in* 2008*." Right Panel: Real HPI index level in Spain over this same time period. Reads: "The real HPI index level rose by just over* 230% *from a base of* 1 *in* 2000*."*