MISINFORMED SPECULATORS AND MISPRICING IN THE HOUSING MARKET

ALEX CHINCO AND CHRIS MAYER

Abstract. This paper examines the contribution of out-of-town second-house buyers to mispricing in the housing market. We show that demand from out-of-town second-house buyers during the mid 2000s predicted not only house-price appreciation rates but also implied-to-actual-rent-ratio appreciation rates, a proxy for mispricing. We then apply a novel identification strategy to address the issue of reverse causality. Finally, we give evidence that out-of-town second-house buyers behaved like misinformed speculators, earning lower capital gains (misinformed) and consuming smaller dividends (speculators).

JEL Classification. R12, R32, G02, G12

Keywords. Out-of-Town Second-House Buyers, Housing Boom, Reverse Causality

Date: March 6, 2015.

Alex Chinco: University of Illinois Urbana-Champaign College of Business. alexchinco@gmail.com.

Chris Mayer: Columbia Business School and NBER; Visiting Scholar, FRBNY. cm310@columbia.edu.

Vivek Sampathkumar, Daniel Hubbard, Laura Vincent, James Witkin, and Vera Chau provided dedicated research assistance. We thank Viral Acharya, Xavier Gabaix, Ed Glaeser, Robin Greenwood, Gur Huberman, Aurel Hizmo, Andrew Paciorek, Tomasz Piskorski, Alexi Savov, Chester Spatt, David Sraer, Jeremy Stein, Johannes Stroebel, Stijn Van Nieuwerburgh, Reed Walker, and Jeff Wurgler as well as seminar participants at Harvard Business School, NYU Stern, Columbia Business School, the University of Chicago, the NBER Behavioral Finance/Housing Bubbles joint session, the NBER Conference Project on Housing and the Financial Crisis, the Federal Reserve Bank of Cleveland, the AEA/ASSA Meetings, the NUS-IRES Symposium, and the 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.

Current Version: http://www.alexchinco.com/misinformed-speculators-and-mispricing.pdf

1. Introduction

During the mid 2000s, house-price appreciation rates in places like Phoenix, Las Vegas, and Miami spiked to 30% per year or more, unprecedented numbers for cities with few building restrictions and relatively elastic land supply. Researchers have long puzzled over the apparent inefficiency of housing prices, for example, see Case and Shiller (1989), but the cross-sectional pattern of house-price appreciation rates in the mid 2000s is particularly hard to square with existing models, as Mayer (2010) and Glaeser, Gottlieb, and Gyourko (2010) point out. What explains this pattern? Why did some cities realize unprecedented house-price booms while prices in other cities, like Charlotte and Denver, remained flat?

This paper suggests a simple answer to the puzzle: out-of-town second-house-buyer demand. We use transactions-level deeds records to show that out-of-town second-house buyers behaved like misinformed speculators during the mid 2000s, driving up both house-price appreciation rates and implied-to-actual-rent-ratio appreciation rates, a proxy for mispricing. In the process we apply a novel identification strategy to address the issue of reverse causality—the hypothesis that it was instead something about cities like Phoenix, Las Vegas, and Miami that was changing in a way that both attracted out-of-town second-house buyers and drove up prices. Crucially, this strategy allows us to show that our results are not driven by such an omitted variable while still remaining agnostic about the exact mechanism behind out-of-town second-house-buyer demand.

Predictive Regressions. We begin by classifying every single-family-house purchase in 21 U.S. cities from January 2000 to December 2007 as made by either an owner occupant, a local second-house buyer, or an out-of-town second-house buyer using the property and tax-bill-mailing addresses listed on the deed of sale. A local second-house buyer is someone who buys a second house in Phoenix and has the property-tax bill sent to a different address in Phoenix. An out-of-town second-house buyer is someone who buys a second house in Phoenix but has the property-tax bill sent to an address in another city, such as San Francisco or Los Angeles. Using a panel vector auto-regression (pVAR) specification, we find that a 1%-point increase in the fraction of sales made to out-of-town second-house buyers in a given month is associated with a 1.9%-point increase in the house-price appreciation rate over the next year. By contrast, increases in local second-house buyer demand are not directly associated with increases in house-price appreciation rates.

Next, we proxy for the level of mispricing in each city using the logarithm of the implied-to-actual-rent ratio (IAR). The log IAR represents the excess return over the apartment rental rate of borrowing money, buying a house, living in it for 1 year, and then selling it in a year's time. The higher the log IAR is, the cheaper it is to rent rather than buy, and the higher the level of mispricing on the rent-vs-own margin. Using the same pVAR specification, we find that a 1%-point increase in the fraction of sales made to out-of-town second-house buyers in a given month is associated with a 1.7%-point increase in the IAR appreciation rate over the next year, implying that observable changes in local housing-market conditions on the rent-vs-own margin only explain around $^{(1.9\,-\,1.7)}/_{1.9} \approx 11\%$ of the run up in prices following an out-of-town second-house-buyer demand shock.

Reverse Causality. We then ask the natural follow up question: what about unobservable changes? Perhaps this association between out-of-town second-house-buyer demand and subsequent house-price and IAR appreciation rates is just due to reverse causality? According to this hypothesis, out-of-town second-house-buyer demand was not driving house-price and IAR appreciation rates during the mid 2000s; instead, unobserved changes in the fundamental value of owning a second house in cities like Phoenix, Las Vegas, and Miami were both attracting out-of-town second-house buyers and driving up prices. This argument involving reverse causality from Friedman (1953) is commonly referred to as the Friedman Critique.

We apply a novel identification strategy to address this concern. Our key insight is that a sudden increase in the fundamental value of owning a second house in, say, Phoenix would represent a common shock to the investment opportunity set of all potential second-house buyers, no matter where they live. Thus, if econometrically unobserved changes to fundamentals in Phoenix were driving both house-price dynamics and out-of-town second-house-buyer demand in that city, then we would expect to see jumps in Phoenix's house-price and IAR appreciation rates preceded by symmetric increases in out-of-town second-house-buyer demand from across the country. That is, the geographic distribution of where second-house buyers are coming from should not matter.

This is not the pattern we find in the data. Instead, increases in demand from out-of-town second-house buyers from large cities arrive at different times and have a much larger impact on house prices in Phoenix than increases in demand from small cities. This result holds when we include ordered city-pair fixed effects to control for the average out-of-town second-house-buyer demand between city pairs. In other words, it is not just that there are always more New Yorkers buying second houses in Miami. The result is also robust to including month fixed effects for each home MSA (where the out-of-town second-house buyer lives) to control for time-specific shocks to fundamentals in each home MSA. In other words, it is not just that people in San Francisco realized an income shock in the mid 2000s and bought more second houses. Thus, the correlation between out-of-town second-house-buyer demand and price changes appears inconsistent with an explanation based on reverse causality alone.

Illustrative Example. The logic behind this identification strategy is subtle, and it is worth pausing for a moment to examine a concrete example in more detail. Suppose that potential second-house buyers living in Los Angeles and Milwaukee are each considering buying a second-house in Phoenix. Given the differences in each market's size, let's say that there are 1,000 potential second-house buyers living in Milwaukee and 10,000 living in Los Angeles, 10 times as many as in Milwaukee.

Consider the null hypothesis, h_0, that *a common shock to the fundamental value of owning a second house in Phoenix is both attracting out-of-town second-house buyers and driving up prices*. In this scenario, potential second-house buyers in both Los Angeles and Milwaukee observe the same shock, namely, the common shock to fundamentals, and increase their demand at the same time. Thus, a 10%-point shock to fundamentals would push an additional 1,100 buyers to purchase second houses in Phoenix. Likewise, in the absence of a shock, then no new buyers from either city purchase second houses in Phoenix. As a result, Phoenix always sees aggregate out-of-town second-house-buyer demand shocks of either 1,100 or none at all. Under the null hypothesis the relative sizes of Los Angeles and Milwaukee are irrelevant. It does not matter how potential out-of-town second-house buyers are distributed across the country because they are responding to a common Phoenix-specific shock.

Now, consider the alternative hypothesis, h_A, that *shocks to fundamentals in Phoenix are not the only thing attracting out-of-town second-house buyers*. For instance, the Milwaukee Journal Sentinel might run a glowing review of Phoenix as a winter getaway, or Home and Garden Television (HGTV) might play a Phoenix-based episode of Flip That House in Los Angeles. In this scenario, potential second-house buyers in Los Angeles and Milwaukee would not in general see the same shocks, so there is no common force pulling out-of-town second-house buyers from Los Angeles and Milwaukee towards the city at the same time. As a result, Phoenix might see a 1,000 out-of-town second-house-buyer demand shock from Los Angeles in one month and then a 100 out-of-town second-house-buyer demand shock from Milwaukee in the next. If the 10%-point out-of-town second-house-buyer shocks from each city do not arrive at the same time, then the 10%-point shock from the larger city, Los Angeles, will have a larger price impact. It is only under this alternative hypothesis that the geographic distribution of second house-buyers matters.

Importantly, this identification strategy allows us to address the issue of reverse causality while still remaining agnostic about the exact mechanism driving out-of-town second-house-buyer demand. While understanding the mechanism is certainly an interesting and important topic, it requires instrumenting for the random assignment of out-of-town second-house buyers to local housing markets. As is often the case, it is difficult to think of an instrument that is correlated with speculators' demand but uncorrelated with information about the fundamental value of the asset they are purchasing. In spite of this limitation, our strategy still allows us to show that out-of-town second-house-buyer demand is linked with higher house-price and IAR appreciation rates in a way that cannot be explained by local market conditions. It is interesting to know that the correlation between out-of-town second-house-buyer demand and price dynamics is not explained by fundamentals, even if we do not know why out-of-town second-house buyers behave the way they do.

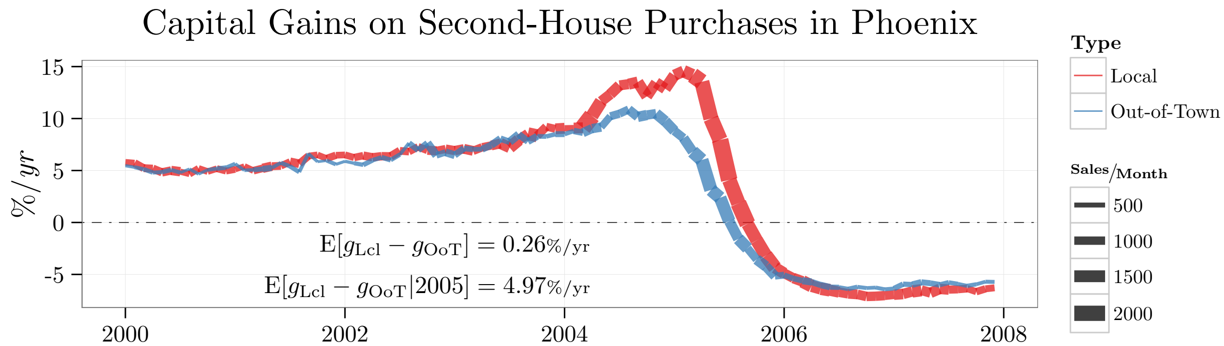


Figure 1. *Line height: Average capital gains on local and out-of-town second-house purchases made in Phoenix. Line width: Number of purchases. Sample: January 2000 to December 2007. Reads: "While local second-house buyers who bought in March 2005 earned nearly an* 15% *per year capital gain on their purchase, out-of-town second-house buyers who bought in the same month earned only a* 10% *per year capital gain on their purchases."*

Misinformed Speculators. We conclude our main analysis by giving evidence that out-of-town second-house buyers behaved like misinformed speculators during the mid 2000s. To begin with, out-of-town second-house buyers realized roughly 5% per year lower capital gains on their purchases than local second-house buyers who bought at the exact same time in the exact same city. What's more, as illustrated in Figure 1, the poor timing of out-of-town second-house buyers appears to be concentrated in purchases near the peak of the market, suggesting that they were less aware of the sharp house-price declines that were to follow, that is, misinformed about local market conditions.

Moreover, out-of-town second-house buyers were more likely to be reliant on such capital gains when computing their returns. In other words, they were more likely to be speculators. Out-of-town landlords are less able to monitor and maintain their property, often paying costly property-management fees of 1 month's rent plus an additional 8% of the rent each year. In addition, at peak the typical out-of-town "second-house" buyer in places like Los Angeles and Jacksonville actually owned 2 or more investment properties, making diversification or consumption from a vacation home unlikely motivations. Finally, out-of-town second-house buyers during the boom were not just wealthy buyers who were less sensitive to capital gains in the purchase of a second house. In fact, out-of-town second-house buyers' primary residences were actually less expensive than the price of the typical house in their home city.

2. Data Description

We compile data on 21 metropolitan statistical areas (MSAs) over the time period from January 2000 to December 2007 using three main sources: transactions-level deeds records from an anonymous data provider, house-price-index (HPI) data from Zillow, and IAR data from Himmelberg, Mayer, and Sinai (2005).

2.1. **Transactions-Level Deeds Records.** We use transactions-level deeds records to classify single-family-house purchases in 21 U.S. cities from January 2000 to December 2007 as

Structure of Transactions-Level Deeds Data

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

Table 1. *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. Reads: "A house at 200 Fremont St. in Las Vegas sold for* \$300k *to someone living at 709 N La Brea Ave. in Los Angeles in May 2006 would be classified as an out-of-town second-house purchase."*

Out-of-Town Second-House Purchases as a % of Sales

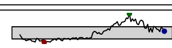
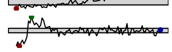
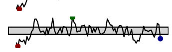
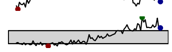
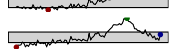
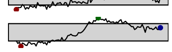
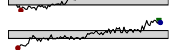
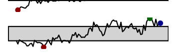
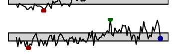
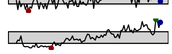
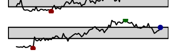
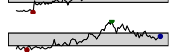
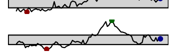
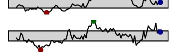
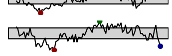
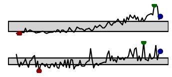
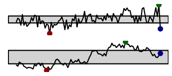
		Mean	Sd	Min	Q25	Q50	Q75	Max
Baltimore		4.8	1.9	2.3	3.2	4.3	6.3	9.7
Charlotte		3.3	2.3	0.5	1.4	2.4	5.9	7.8
Cincinnati		6.3	1.2	2.8	5.8	6.3	6.9	9.5
Cleveland		5.4	0.9	3.0	4.9	5.4	6.0	7.5
Denver		2.2	1.3	0.7	1.1	1.7	3.3	5.5
Jacksonville		5.9	2.6	2.2	3.8	5.0	7.5	12.3
Las Vegas		11.0	3.8	4.7	7.0	12.0	14.5	17.1
Los Angeles		1.2	0.4	0.2	0.9	1.1	1.4	2.2
Miami		4.6	1.5	2.0	3.2	4.4	5.9	7.4
Milwaukee		1.3	0.6	0.2	0.9	1.2	1.6	2.9
Minneapolis		1.4	0.8	0.2	0.7	1.3	2.0	3.4
Orlando		9.9	3.4	3.2	7.5	10.0	12.4	15.7
Philadelphia		2.6	1.3	0.8	1.5	2.6	3.6	5.6
Phoenix		7.7	3.0	3.6	5.5	6.7	9.4	15.5
Riverside		8.3	1.3	5.6	7.4	8.2	9.4	11.4
Sacramento		6.5	1.0	4.3	5.9	6.6	7.4	8.3
San Diego		3.1	1.5	1.4	1.9	2.6	4.0	7.5
San Francisco		2.3	0.4	1.6	2.1	2.3	2.5	3.7
San Jose		1.2	0.5	0.7	1.6	1.8	2.1	3.1
Tampa		7.7	2.5	3.7	5.7	7.3	9.9	12.5
Washington		1.4	0.4	0.6	1.1	1.2	1.6	2.5
Mean		4.6	1.5					

Table 2. *Percent of single-family-house purchases made by out-of-town second-house buyers each month from January 2000 to December 2007. The shaded region in each sparkline plot covers the interquartile range for each MSA and is not a constant scale. Reads: "Over the entire sample period,* 5.9% *of all sales were made to out-of-town second-house buyers each month in Jacksonville; however, the fraction of sales made to out-of-town second-house buyers was more than double this number at its peak in early 2006."*

made by either an owner occupant, a local second-house buyer, or an out-of-town second-house buyer. A deed is a written legal instrument that passes the rights to a particular property, which in our case is a single-family house, from one owner to the next. Deeds records are public in most states due to information disclosure acts and are maintained by the local county, so they contain a complete sales history of any parcel of land, documenting any time a property is sold or a new mortgage is taken out.

Definition 1 (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.*

Second-House Buyers. For each transaction in our data, we observe not only an address for the property itself but also a mailing address where the county sends the property's tax bill. Table 1 gives a fictitious example of the data for an owner occupant, a local second-house buyer, and an out-of-town second-house buyer. In the mid 2000s, the number of purchases

Local Second-House Purchases as a % of Sales

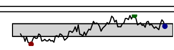
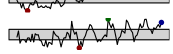
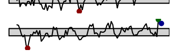
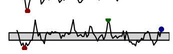
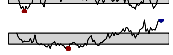
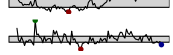
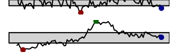
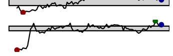
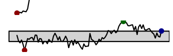
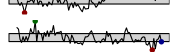
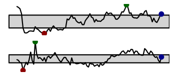
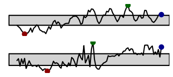
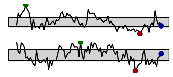
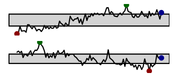
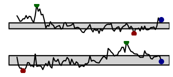
		Mean	Sd	Min	Q25	Q50	Q75	Max
Baltimore		13.1	3.1	7.1	10.3	13.6	15.5	18.9
Charlotte		9.4	1.5	6.7	8.4	9.4	10.5	12.6
Cincinnati		12.3	1.8	6.9	11.1	12.2	13.6	16.7
Cleveland		10.5	1.6	6.8	9.2	10.4	11.4	15.2
Denver		9.9	2.4	6.4	8.0	9.2	11.7	16.2
Jacksonville		17.0	1.7	13.1	15.8	16.9	18.0	24.0
Las Vegas		12.8	3.4	7.3	10.3	12.8	14.7	19.7
Los Angeles		10.3	2.3	3.0	9.9	10.8	11.6	13.5
Miami		14.6	2.0	10.8	13.2	14.3	16.1	18.8
Milwaukee		10.1	2.0	5.5	8.6	9.8	11.6	16.3
Minneapolis		13.3	4.3	5.8	9.0	14.1	16.5	22.6
Orlando		15.9	2.2	10.9	14.1	15.9	17.3	22.7
Philadelphia		16.0	3.1	10.0	13.8	16.4	18.1	22.6
Phoenix		16.2	2.6	11.9	13.9	16.1	18.1	22.1
Riverside		10.4	1.1	8.4	9.6	10.2	11.2	13.3
Sacramento		11.6	1.4	8.7	10.6	11.5	12.7	14.1
San Diego		12.7	2.4	7.8	10.5	13.5	14.6	17.7
San Francisco		10.0	1.5	6.7	8.7	10.0	11.1	14.2
San Jose		8.1	1.9	5.0	6.8	7.7	8.9	15.2
Tampa		17.7	2.6	12.9	15.4	17.3	19.3	24.9
Washington		9.0	1.8	6.5	7.5	8.8	10.2	13.9
Mean		13.8	2.3					

Table 3. *Percent of single-family-house purchases made by local second-house buyers each month from January 2000 to December 2007. The shaded region in each sparkline plot covers the interquartile range for each MSA and is not a constant scale. Reads: "Over the entire sample period,* 16.0% *of all sales were made to local second-house buyers each month in Philadelphia. At peak, this number climbed to* 22.6%*."*

by out-of-town second-house buyers in MSAs like Las Vegas, Miami, and Phoenix grew appreciably. Table 2 gives summary statistics for the number of out-of-town second-house purchases in each MSA i as a fraction of the total number of sales each month. Even at the peak, out-of-town second-house purchases always represented a minority of house purchases. In the most extreme market, Las Vegas, out-of-town second-house buyers purchased 17% of all sales in 2004, up from roughly 7% in the early 2000s. Many of these MSA-specific sparkline plots display a similar hump-shaped pattern in the number of out-of-town second-house purchases, but the scale of this pattern differs dramatically across MSAs. For example, 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.

Definition 2 (Second-House Purchases)**.** *Define $S_{(i,j),t}$ as the annualized number of single-family-house 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. $S_{(i,j),t}$ has units of houses per year.*

Definition 3 (Out-of-Town Second-House Purchases)**.** *Define $S_{j,t}^{\mathrm{OoT}} = \sum_{j \neq i} S_{(i,j),t}$ 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 $j \neq i$. $S_{j,t}^{\mathrm{OoT}}$ has units of houses per year.*

Definition 4 (Local Second-House Purchases)**.** *Define $S_{j,t}^{\mathrm{Lcl}} = S_{(j,j),t}$ as the annualized number of second-house purchases in MSA j at month t where both the mailing and property addresses are located in MSA j. $S_{j,t}^{\mathrm{Lcl}}$ has units of houses per year.*

Local vs. Out-of-Town. Research on second-house buyers typically treats local and out-of-town second-house buyers in the same way. However, as demonstrated in Table 3, purchases by local second-house buyers exhibit a very different time-series pattern than that of out-of-town second-house buyers. The overall share of purchases by local second-house buyers varies much less across markets compared to house-price appreciation. As well, in most cases (Las Vegas is the appreciable exception), the share of local second-house buyers does not exhibit a hump with a peak at or near the peak of house prices.

Other papers have examined credit-report data and found similar patterns in second-house-buyer volume. For example, Haughwout, Lee, Tracy, and van der Klaauw (2011) show that mortgages on second houses represented nearly half of all originations in the four states with the highest house-price appreciation rates during the boom. However, it is important to our study that we use deeds records rather than data on mortgage originations for two reasons. First, Piskorski, Seru, and Witkin (2014) gives evidence that second-house buyers strategically misreported their ownership status to lenders. Second, both local and out-of-town second-house buyers are more likely to do cash purchases, making these buyer types less likely to show up in mortgage-origination data.

Data Source. Finally, while we have cross-referenced our deeds data with the data from Dataquick to verify its accuracy, it is important to point out that we are not using Dataquick data in our main analysis. In fact, it is not possible to perform our analysis using Dataquick data for our sample period because Dataquick often overwrites the old tax-bill-mailing address with the most recent value. To illustrate by example, if we were using Dataquick data, the tax-bill-mailing address for the second row of Table 1 would most likely be back-filled to also read "709 N La Brea Ave, LA" instead of the correct "888 W Bonneville Ave, LV".

2.2. **House-Price Appreciation Rate.** We obtain monthly house-price-index (HPI) data from Zillow at the MSA level. Zillow data are available for a larger number of locations than the S&P/Case-Shiller index data and use 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 the S&P/Case and Shiller indexes during the boom, but show less of a sharp decline in 2007. See http://www.zillow.com/research/

House-Price Appreciation Rates in % per Year

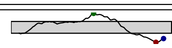
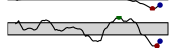
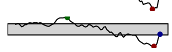
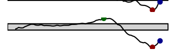
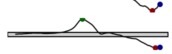
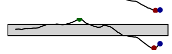
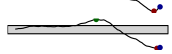
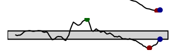
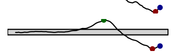
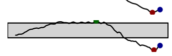
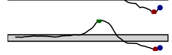
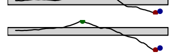
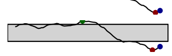
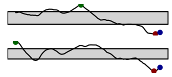
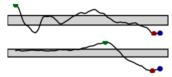
		Mean	Sd	Min	Q25	Q50	Q75	Max
Baltimore		6.5	9.3	−12.3	−1.8	10.1	12.0	21.4
Charlotte		1.0	3.4	−6.5	−1.6	1.3	3.9	6.4
Cincinnati		0.0	2.4	−6.3	−1.1	0.8	1.7	3.2
Cleveland		−1.9	4.2	−11.2	−5.0	−0.3	1.1	4.5
Denver		−0.3	4.4	−9.6	−2.6	−0.2	1.0	11.3
Jacksonville		4.6	9.9	−19.3	2.3	8.2	10.2	17.0
Las Vegas		3.7	18.0	−34.5	−3.7	5.1	7.8	44.3
Los Angeles		6.2	14.9	−28.6	−4.5	10.9	16.6	27.6
Miami		6.5	16.7	−31.1	−1.7	12.9	14.9	27.3
Milwaukee		1.5	5.5	−9.4	−2.2	1.2	4.6	14.2
Minneapolis		1.9	7.0	−13.2	−3.0	4.4	7.1	10.5
Orlando		5.3	15.4	−28.7	0.9	7.3	12.6	32.3
Philadelphia		5.6	6.7	−8.5	0.5	7.7	11.4	12.9
Phoenix		3.5	16.3	−25.4	−7.4	3.7	7.3	39.4
Riverside		6.5	19.1	−40.2	−2.7	11.6	16.8	33.2
Sacramento		5.1	15.7	−26.8	−12.0	13.4	17.0	22.7
San Diego		4.5	14.7	−25.2	−7.6	9.0	15.7	27.9
San Francisco		2.3	12.0	−25.2	−5.5	4.0	11.3	21.9
San Jose		1.3	11.0	−19.2	−5.7	0.3	9.2	25.8
Tampa		4.7	14.3	−26.5	−2.0	9.4	11.5	23.9
Washington		6.4	12.8	−21.7	−6.3	11.1	15.9	20.7
Mean		3.6	11.1					

Table 4. *Annualized house-price appreciation rates from January 2000 to December 2007. The shaded region in the sparkline plots covers the interquartile range for each MSA and is not a constant scale. Reads: "Over the entire sample period, house prices in Las Vegas grew by* 3.69% *per year on average; however, this rate skyrocketed to* 44.3% *per year in 2004."*

zhvi-methodology-6032/ for a more detailed description of the Zillow house-price-index methodology. All of our results are robust to using price data from S&P/Case-Shiller index.

Definition 5 (House-Price Appreciation Rate)**.** *Define the house-price appreciation rate in MSA i in month t as $\Delta \log P_{i,t\to(t+\tau)} = \log P_{i,t+\tau} - \log P_{i,t}$ in units of $1/\tau\text{mo}$.*

Table 4 gives summary statistics for MSA-level house-price appreciation rates in units of percent per year. A number of the markets saw house-price appreciation rates above 20% per year, with house-price appreciation rates exceeding 35% per year in Las Vegas and Phoenix. 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. 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.

2.3. **IAR Appreciation Rate.** In order to examine mispricing, we must pick a pricing model. Beginning with Poterba (1984), many authors have priced real estate by comparing the price of a house to the present value of its stream of rental payments after taking into account the favorable tax treatment given to owner-occupied properties and the associated mortgage payments. This pricing strategy is similar to the dividend discount model for stocks, and we refer to models that price housing along this margin as user-cost models.

User-Cost Model. Yet, unlike in the stock market where analysts have 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. Himmelberg, Mayer, and Sinai (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, with the imputed rent calculated by multiplying the user cost times the price of an owner-occupied house. We use the user-cost-of-housing data from this paper updated through December 2007. Table 6 describes the input variables for the user-cost model.

Definition 6 (User Cost of Housing)**.** *The user cost of housing in MSA i in month t, $U_{i,t\to(t+12)}$, is the fraction of the house price that an owner must pay in order to live in the house for the next year,*

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

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

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 7 (Apartment Rental Rate). *The apartment rental-rate in MSA i in month t, $R_{i,t\to(t+12)}$, is the rent on a 2-bedroom apartment over the next year in units of* $1/\text{yr}$.

Proxy for Mispricing. 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 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 while enduring a constant risk premium of π per year. Himmelberg, Mayer, and Sinai (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

IAR Appreciation Rates in % per Year

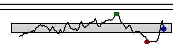
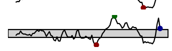
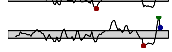
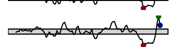
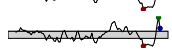
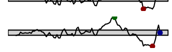
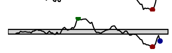
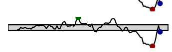
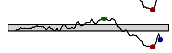
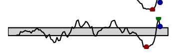
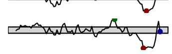
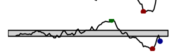
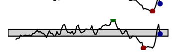
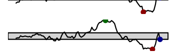
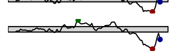
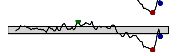
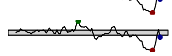
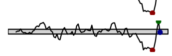
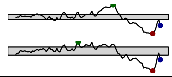
		Mean	Sd	Min	Q25	Q50	Q75	Max
Baltimore		2.0	9.8	−18.5	−4.1	2.4	9.4	24.8
Charlotte		−1.0	8.5	−16.4	−6.4	−1.8	4.1	21.9
Cincinnati		−1.7	5.3	−12.6	−5.3	−1.6	1.8	14.9
Cleveland		−3.4	5.7	−18.2	−6.0	−2.6	−0.5	13.1
Denver		−2.4	7.5	−17.5	−7.2	−1.9	2.4	20.4
Jacksonville		0.8	11.8	−25.9	−4.1	0.5	6.1	31.4
Las Vegas		0.6	16.6	−39.4	−5.4	0.7	6.9	36.9
Los Angeles		0.2	14.0	−39.6	−4.4	3.1	8.5	24.7
Miami		2.7	14.9	−33.1	−0.9	4.7	12.1	24.7
Milwaukee		0.0	6.7	−15.1	−3.4	0.5	4.5	14.3
Minneapolis		−0.1	7.4	−19.6	−2.5	0.9	4.8	13.5
Orlando		1.8	14.7	−32.5	−4.0	1.7	9.3	32.6
Philadelphia		2.4	7.1	−14.5	−0.9	2.4	6.7	17.8
Phoenix		0.2	16.6	−32.3	−8.6	−1.7	7.0	36.6
Riverside		0.4	16.4	−46.3	−3.2	3.7	10.7	27.1
Sacramento		0.9	13.0	−35.2	−4.3	4.8	9.3	20.1
San Diego		−1.4	12.8	−37.2	−4.9	1.6	5.9	25.0
San Francisco		−0.4	13.8	−41.1	−3.5	3.7	7.9	18.2
San Jose		0.2	13.2	−36.1	−2.4	1.8	7.4	23.9
Tampa		1.6	13.5	−30.9	−2.3	3.6	8.5	27.8
Washington		1.6	11.6	−29.1	−3.0	4.6	10.5	18.6
Mean		0.2	11.5					

Table 5. *Annualized IAR appreciation rate from January 2000 to December 2007. The shaded region in the sparkline plots covers the interquartile range for each MSA and is not a constant scale. Reads: "Over the entire sample period, the IAR appreciation rate in Phoenix was* 0.205% *per year on average; however, this rate jumped to* 36.6% *per year in 2004."*

important short-run considerations like easy and cheap leverage in the mid 2000s and time varying risk premia. When the log IAR in a given metropolitan area exceeds zero, owning a house is more expensive than renting relative to the average value over the sample period.

Definition 8 (Implied-to-Actual-Rent Ratio (IAR) Appreciation Rate)**.** *The IAR in MSA i in month t, $Z_{i,t}$, is the ratio of the cost of borrowing money, purchasing a house, and then selling it in 1 year to the rent a comparable property for the same amount of time,*

$$Z_{i,t} = \frac{1}{\bar{Z}_i} \cdot \left(\frac{P_{i,t} \cdot U_{i,t \to (t+12)}}{R_{i,t \to (t+12)}} \right) \quad \text{with} \quad \bar{Z}_i = \frac{1}{T'} \cdot \sum_{t=1}^{T'} \left(\frac{P_{i,t} \cdot U_{i,t \to (t+12)}}{R_{i,t \to (t+12)}} \right). \tag{2}$$

The IAR is scaled so that $Z_{i,t} = 1$ is the average value from January 1980 to December 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 $1/\tau\text{mo}$.

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 5 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 \$1,000 per month would have to pay \$1,658 per month in mortgage payments and other costs in order to buy an equivalent house and live in it from January 2004 to December 2004. By comparison, in Denver, this ratio was 1.267

Input Variables to User-Cost Model

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, Miceli, and Sirmans (2000)	Housing-capital depreciation rate.
$\mathrm{E}[\Delta \log P_{i,t+12}]$	Gyourko, Mayer, and Sinai (2013), the US Census, the Livingston Survey, and Zillow	Expected house-price appreciation rate equals either the historical long-term real growth rates by MSA plus expected inflation (main calculations) or the weighted-average growth rate by MSA over the past 5 years (robustness).
π	Flavin and Yamashita (2002)	Risk premium associated with owning a house.

Table 6. *Input variables used to compute the user cost of housing in Himmelberg, Mayer, and Sinai (2005). All variables have units of $1/\text{yr}$ except for the federal marginal tax rate, $\kappa_{i,t}$, which is a fraction.*

between 2004 and 2006, so a tenant would have paid about $1,267 per month to purchase a house that rented for $1,000 per month and live in it from January 2004 to December 2004. While houses in Denver were still priced at a small premium relative to renting at the peak of the boom, they appeared much less overpriced than houses in Phoenix at the same time.

User-Cost Critiques. Researchers have critiqued the user-cost model 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.

Nevertheless, the comparison of house prices to variables like employment and income has no firm theoretical foundation and fails to account for changes in economic fundamentals like interest rates and land supply. For example, 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. By analogy, note that in equity markets there are circumstances where shares of the same stock confer different dividend streams and a user-cost-like model is warranted. That is, shares with voting rights command a higher price-to-dividend ratio than shares of the same stock without voting rights as documented in Zingales (1995).

More quantitatively, Hubbard and Mayer (2009) estimates a log-linearized model,

$$\log P_{i,t} = \widehat{\alpha}_i + \widehat{\kappa}_t + \widehat{\beta} \cdot \log R_{i,t\to(t+12)} + \widehat{\gamma} \cdot \log U_{i,t\to(t+12)} + \varepsilon_{i,t}, \tag{3}$$

over the time interval from January 1980 to December 2007 and finds coefficients of $\widehat{\gamma} = 0.93$ and $\widehat{\beta} = -0.75$, values that are quite close to the theoretical $\gamma = 1.0$ and $\beta = -1.0$ predicted by the static user-cost model. Thus, even though it has many imperfections, the user-cost model appears to provide a simple benchmark for what housing prices might be in a long-term equilibrium. 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.

3. Predictive Regressions

Having outlined the data, we now show that increases in out-of-town second-house-buyer demand predict increases in future house-price and IAR appreciation rates; by contrast, increases in local second-house buyer demand are not directly associated with increases in either of these measures.

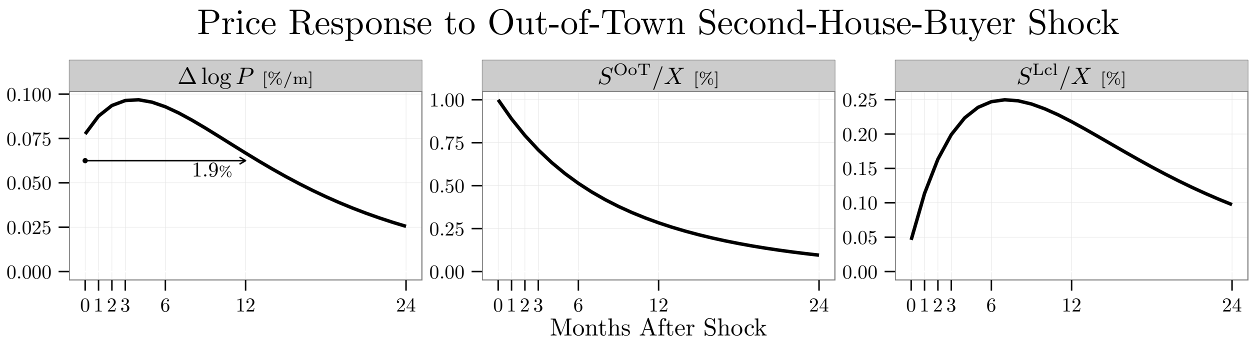


Figure 2. *Response of house-price appreciation rates to a* 1%*-point increase in the fraction of sales made to out-of-town second-house buyers in Las Vegas. Reads: "In the* 12 *months following a* 1%*-point increase in out-of-town second-house-buyer demand, the house-price appreciation rate in Las Vegas will be roughly* 1.9%*-points higher."*

3.1. **Regression Results.** We begin by estimating a panel vector auto-regression (pVAR) characterizing the relationship between the house-price appreciation rate in an MSA and the fraction of all purchases made by local and out-of-town second-house buyers,

$$\mathbf{Y}_{i,t} = \left[\Delta \log P_{i,(t-1)\to t} \quad \frac{S^{\mathrm{OoT}}_{i,t}}{X_{i,t}} \quad \frac{S^{\mathrm{Lcl}}_{i,t}}{X_{i,t}} \right]^{\top}, \tag{4}$$

The fraction of sales made by owner occupants is the omitted category. Specifically, we analyze Equation (5) below,

$$\mathbf{E}_{i,t} = (\mathbf{I} - \boldsymbol{\Theta}\mathcal{L})\,(\mathbf{Y}_{i,t} - \mathbf{A}_i - \mathbf{K}_t)\,, \tag{5}$$

In this representation, $\mathbf{I}$ denotes a 3×3 identity matrix, $\boldsymbol{\Theta}$ denotes the 3×3 transition matrix, $\mathcal{L}$ 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 $\mathbf{E}_{i,t}$ denotes a 3×1 vector of error terms.

Price Response. Panel (a) of Table 7 reveals that a 1%-point increase in the fraction of purchases made by out-of-town second-house buyers in an MSA i in month t is associated with a 0.02%-point increase in the house-price appreciation rate in the following month. 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 Las Vegas's house-price appreciation rate in response to a 1%-point increase in the fraction of purchases made by out-of-town second-house buyers. We plot this impulse-response function in Figure 2. Because shocks to the fraction of out-of-town second-house buyers in Las Vegas are correlated with shocks to both the fraction of local second-house buyers and house-price appreciation rates in Las Vegas, we consider an orthogonalized impulse-response function, $\mathrm{Imp}_i(h) = \boldsymbol{\Theta}^h \mathbf{C}_i$, where $\mathbf{C}_i$ denotes the Cholesky decomposition of the variance-covariance matrix for MSA i, $\mathrm{Var}(\mathbf{E}_{i,t}) = \mathbf{C}_i \mathbf{C}_i^{\top}$. This analysis reveals that a 1%-point increase in the fraction of purchases made by out-of-town second-house buyers in Las Vegas is associated with a 1.9%-point increase in Las Vegas's house-price appreciation rate over the next year.

IAR Response. Out-of-town second-house-buyer demand predicts house-price appreciation rates, but large price movements do not necessarily indicate mispricing. Instead, these price movements could be due to observable fluctuations in housing-market fundamentals on the rent-vs-own margin. Perhaps owning and renting in Las Vegas were becoming more costly in the mid 2000s? To address this concern, we also estimate the exact same pVAR specification using the IAR appreciation rate rather than the house-price appreciation rate,

$$\mathbf{Y}_{i,t} = \left[\Delta \log Z_{i,(t-1)\to t} \quad \frac{S_{i,t}^{\text{OoT}}}{X_{i,t}} \quad \frac{S_{i,t}^{\text{Lcl}}}{X_{i,t}}\right]^{\top}. \tag{6}$$

Comparing Panel (a) in Tables 7 and 8 reveals that the fraction of purchases made by out-of-town second-house buyers is also a strong predictor of future IAR appreciation rates. That is, a 1%-point increase in the fraction of purchases made by out-of-town second-house buyers in a given month is associated with a 0.07%-point increase in the IAR appreciation rate in the subsequent month. This evidence suggests that out-of-town second-house-buyer demand shocks do not just affect prices, rather, they also appreciably distort the rent-vs-own calculus of people living in the target MSA. Just as before, we find very little evidence that increases in the fraction of purchases by local second-house buyers is directly linked to

pVAR Describing House-Price Appreciation Rates

(a) Dep. Var.: House-Price Appreciation Rate

	Estimate	Std. Error			
Lagged Price-Apprec. Rate	0.87	0.02	0.02	0.03	0.03
Lagged Percent Out-of-Town	0.02	0.01	0.01	0.01	0.01
Lagged Percent Local	−0.01	0.00	0.01	0.01	0.01
		$\emptyset$	t	i	t,i

(b) Dep. Var.: Out-of-Town Second-House-Buyer Percent

	Estimate	Std. Error			
Lagged Price-Apprec. Rate	0.08	0.02	0.03	0.02	0.03
Lagged Percent Out-of-Town	0.88	0.01	0.02	0.01	0.02
Lagged Percent Local	0.01	0.01	0.01	0.01	0.01
		$\emptyset$	t	i	t,i

(c) Dep. Var.: Local Second-House-Buyer Percent

	Estimate	Std. Error			
Lagged Price-Apprec. Rate	0.14	0.04	0.05	0.02	0.04
Lagged Percent Out-of-Town	0.07	0.02	0.03	0.02	0.03
Lagged Percent Local	0.82	0.01	0.01	0.03	0.03
		$\emptyset$	t	i	t,i

Table 7. *Parameter values of the transition matrix* $\mathbf{\Theta}$ *in Equation* (5)*.* $N = 1{,}995$ *monthly observations on* 21 *MSAs from January 2000 to December 2007. Reads: "A* 1%*-point increase in the fraction of all purchases made by out-of-town second-house buyers in April is associated with a* 0.02%*-point increase in the house-price appreciation rate in May."*

mispricing, with Table 8 showing a tightly estimated zero effect.

Figure 3 shows the analogous impulse-response function for a 1%-point increase in the fraction of all purchases made by out-of-town second-house buyers on the IAR appreciation rate in Las Vegas. We find that the cumulative effect of this 1%-point shock over the next year is a 1.7%-point increase in the IAR appreciation rate. Put differently, only about $^{(1.9-1.7)}/_{1.9} \approx 11\%$ of the 1.9% increase in house-price appreciation rates following an out-of-town second-house-buyer shock can be accounted for by observable changes in housing-market fundamentals on the rent-vs-own margin.

3.2. **Interpretation and Discussion.** Having seen the regression result, let's now turn our attention to their interpretation.

Standard Errors. In Tables 7 and 8 we report multiple standard-error clustering schemes: the usual OLS standard errors, standard errors clustered by month, standard errors clustered by MSA, and standard errors clustered by both month and MSA as in Thompson (2011). We do this for two main reasons. First, it is not always best to just report the "most robust" standard errors. There is a bias-vs-variance tradeoff. Clustering along more dimensions

pVAR Describing IAR Appreciation Rates

(a) Dep. Var.: IAR Appreciation Rate

	Estimate	Std. Error			
Lagged IAR Apprec. Rate	0.44	0.04	0.08	0.09	0.11
Lagged Percent Out-of-Town	0.07	0.01	0.02	0.03	0.03
Lagged Percent Local	0.00	0.01	0.01	0.01	0.01
		$\emptyset$	t	i	t,i

(b) Dep. Var.: Out-of-Town Second-House-Buyer Percent

	Estimate	Std. Error			
Lagged IAR Apprec. Rate	0.05	0.02	0.02	0.02	0.02
Lagged Percent Out-of-Town	0.89	0.01	0.02	0.01	0.02
Lagged Percent Local	0.01	0.01	0.01	0.01	0.01
		$\emptyset$	t	i	t,i

(c) Dep. Var.: Local Second-House-Buyer Percent

	Estimate	Std. Error			
Lagged IAR Apprec. Rate	0.09	0.03	0.03	0.03	0.03
Lagged Percent Out-of-Town	0.07	0.02	0.02	0.02	0.03
Lagged Percent Local	0.82	0.01	0.01	0.03	0.02
		$\emptyset$	t	i	t,i

Table 8. *Parameter values of the transition matrix* $\mathbf{\Theta}$ *in Equation* (5). $N = 1{,}995$ *monthly observations on* 21 *MSAs from January 2000 to December 2007 with the IAR appreciation rate rather than the house-price appreciation rate. Reads: "A* 1%*-point increase in the fraction of all purchases made by out-of-town second-house buyers in April is associated with a* 0.073%*-point increase in the IAR appreciation rate in May."*

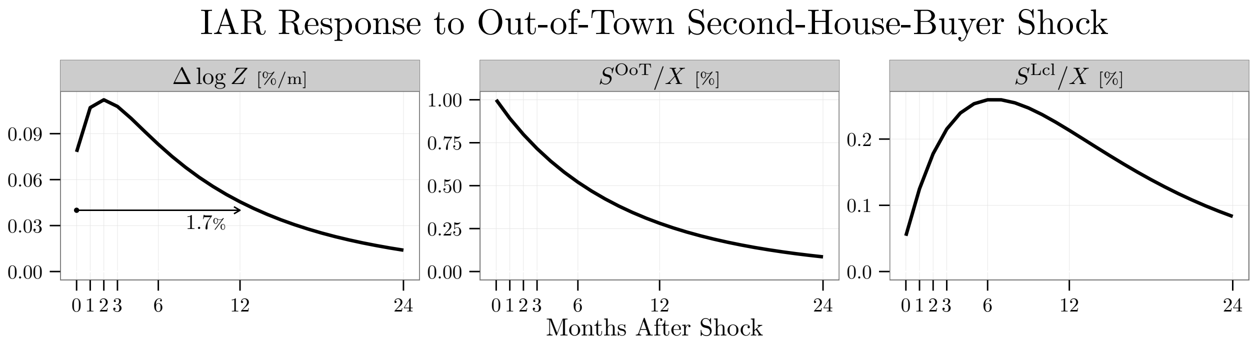


Figure 3. *Response of IAR appreciation rates to a* 1%*-point increase in the fraction of sales made to out-of-town second-house buyers in Las Vegas. Reads: "In the* 12 *months following a* 1%*-point increase in out-of-town second-house-buyer demand, the IAR appreciation rate in Las Vegas will be roughly* 1.7%*-points higher."*

accounts for more complicated dependencies in the standard errors but also increases the bias of the standard-error estimates.

Second, and more importantly, because our pVAR specification is a reduced-form technique for summarizing a more complicated high-dimensional system, the multiple standard errors also serve as useful diagnostic tools. For example, Petersen (2009) points out that it would be worrying if clustering by month dramatically increased the standard error on, say, the auto-regressive coefficient for house-price appreciation rates in the first row of Table 7 because this would indicate that there was a lot of month to month variation in the strength of this relationship. The standard errors above are relatively constant across clustering schemes suggesting that the auto-regressive relationship between the variables, $\mathbf{\Theta}$, is relatively constant over time and across MSAs even though the variables themselves are not.

Coefficient Bias. As is the case whenever you estimate an auto-regressive model with an unknown mean, it is important to address the potential bias of the estimates due to finite-sample effects. Discussion of this bias dates back to at least Marriott and Pope (1945) and Kendall (1945), and Nickell (1981) shows that including group fixed effects in a panel setting can exacerbate the problem. Indeed, monte-Carlo simulations confirm this suspicion. We find that the auto-regressive coefficients are likely biased downward by roughly 0.05 when each panel has $T = 96$ monthly observations, that is, the estimates in the first row of Panel (a), the second row of Panel (b), and the third row of Panel (c) in Tables 7 and 8.

Nevertheless, the very same simulations reveal no significant biases in the estimates for the other right-hand-side variables, like the effect of out-of-town second-house-buyer demand on future house-price appreciation rates or the lack of a direct effect of local second-house-buyer demand on future house-price appreciation rates, the estimates from the second and third rows in Panel (a) of Table 7. In short, the right-hand-side variables are likely a bit more persistent than the pVAR estimates in Tables 7 and 8 suggest, but each variable's effects on the others are accurately estimated.

pVAR vs. pVARMA Specification. Because of the moderate length of each panel, it is difficult to estimate a pVARMA specification in place of the more spartan pVAR(1) specification. The pVAR(1) specification captures the first-order associations between out-of-town second-house-buyer demand and subsequent house-price appreciation rates but not more complicated momentum and reversal effects. For example, it is beyond the scope of our econometric specification to estimate an effect like: an increase in out-of-town second-house-buyer demand in April is associated with higher house-price appreciation rates in May but lower house-price appreciation rates in August.

Alternative User-Cost Calculation. Expectations of future house-price appreciation enter as an input to the user-cost model. In our main specification, we use the long-run house-price appreciation rate for this input. But, perhaps home buyers have different ideas about what the future house-price appreciation rate will be? For instance, they might be using a simpler moving-average model for house prices. To address this concern, we recompute each MSA's

pVAR Describing IAR Appreciation Rates Using
5-year Moving Avg. for Expected House-Price Growth

(a) Dep. Var.: IAR Appreciation Rate

	Estimate	Std. Error			
Lagged IAR Apprec. Rate	0.63	0.03	0.03	0.05	0.05
Lagged Percent Out-of-Town	0.03	0.01	0.01	0.01	0.01
Lagged Percent Local	0.00	0.01	0.01	0.01	0.01
		$\emptyset$	t	i	t,i

(b) Dep. Var.: Out-of-Town Second-House-Buyer Percent

	Estimate	Std. Error			
Lagged IAR Apprec. Rate	0.09	0.03	0.03	0.03	0.03
Lagged Percent Out-of-Town	0.89	0.01	0.02	0.01	0.02
Lagged Percent Local	0.01	0.01	0.01	0.01	0.01
		$\emptyset$	t	i	t,i

(c) Dep. Var.: Local Second-House-Buyer Percent

	Estimate	Std. Error			
Lagged IAR Apprec. Rate	0.12	0.05	0.05	0.04	0.05
Lagged Percent Out-of-Town	0.08	0.02	0.02	0.02	0.03
Lagged Percent Local	0.82	0.01	0.01	0.03	0.02
		$\emptyset$	t	i	t,i

Table 9. *Parameter values of the transition matrix* $\mathbf{\Theta}$ *in Equation* (5) *when the IAR appreciation rate that is calculated using the* 5*-year moving average of the house-price appreciation rate for the expected house-price appreciation rate.* $N = 1{,}995$ *monthly observations on* 21 *MSAs from January 2000 to December 2007. Reads: "A* 1%*-point increase in the fraction of all purchases made by out-of-town second-house buyers in April is associated with a* 0.03%*-point increase in the IAR appreciation rate in May."*

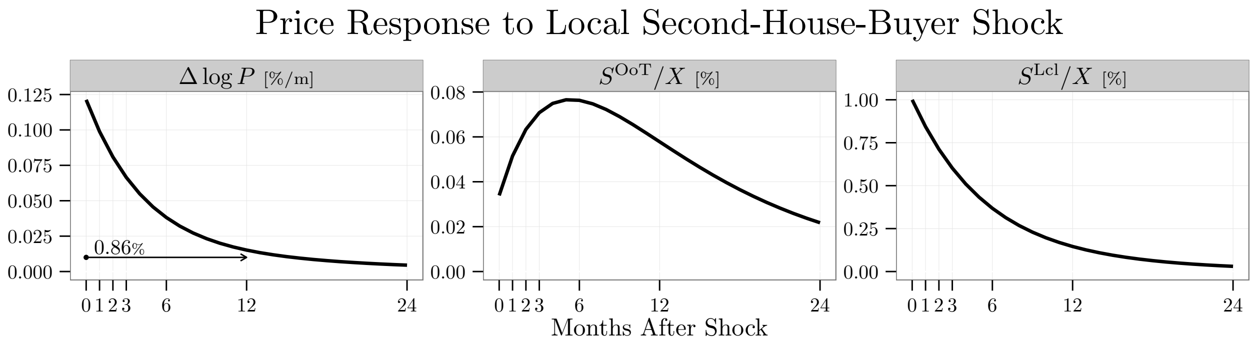


Figure 4. *Response of house-price appreciation rates to a* 1%*-point increase in the fraction of all sales made to local second-house buyers in Las Vegas. Reads: "In the* 12 *months following a* 1%*-point increase in local second-house buyer demand, house-price appreciation rates in Las Vegas are roughly* 0.86%*-points higher."*

implied-to-actual-rent ratio using a five-year rolling average for home buyers' expectation of future house-price appreciation rates. Table 9 shows the estimation results for this new specification, which indicate that our results are not dependent upon the precise way that the expected price appreciation is computed in the IAR appreciation rate calculation.

Local Second-House Buyers. While the fraction of purchases made by out-of-town second-house buyers predicts house-price appreciation rates and IAR appreciation rates over the next year, the fraction of purchases made by local second-house buyers has no direct effect on these variables as shown in the third row of Panel (a) in Tables 7 and 8 respectively. Thus, it is clear that not all second-house buyers were created equally. This finding connects the current paper with a body of existing work looking at the relationship between distance and investment performance. For example, within the context of the real-estate market Kurlat and Stroebel (2014) give a theoretical model of markets with many heterogeneous assets and differentially informed agents. More broadly, Cohen, Frazzini, and Malloy (2010) study the effect of a sell-side equity analyst's social network on her ability to gather information. They find that analysts who are more distant from a firm in terms of having gone to school with fewer of the senior management team underperform their peers. Lu, Swan, and Westerholm (2015) show that distant portfolio managers earn lower returns using Finnish data.

This result might at first seem to be at odds with a recent string of real-estate finance papers studying the impact of local second-house buyers. For example, Bayer, Geissler, and Roberts (2011) analyze the behavior of local second-house buyers in Los Angeles and find that local second-house buyer demand is strongly associated with neighborhood-level price instability. Similarly, Li and Gao (2012) give evidence that local second-house buyers are more likely to be present in cities with high house-price appreciation rates. But, there is no conflict. We estimate the impulse response to a 1%-point increase in local second-house-buyer demand and find that house-price appreciation rates usually rise by around 0.86% points in the year after such a shock.

Even in our data, we find that local second-house-buyer demand and house-price appreciation rates are positively linked. Here is the important part, though. The positive association between local second-house-buyer demand and subsequent house-price appreciation rates comes almost entirely from the fact that local second-house-buyer shocks and out-of-town second-house-buyer shocks tend to arrive at the same time. There is no direct effect of local second-house-buyer demand on subsequent house-price appreciation rates. Thus, the results in these papers are complimentary to ours in that they document the large growth in second-house purchases in the highest appreciating cities; however, they also highlight the need to address concerns about reverse causality.

4. Reverse Causality

We just saw that out-of-town second-house-buyer demand is positively *correlated* with house-price and IAR appreciation rates. We now investigate the obvious follow-up question: are these positive correlations simply due to reverse causality? Perhaps something about housing-market fundamentals in cities like Phoenix, Las Vegas, and Miami was changing in such a way that both attracted out-of-town second-house buyers and drove up prices? We introduce a novel identification strategy to address this issue and find that the effects of out-of-town second-house-buyer demand cannot be explained by such an omitted variable.

4.1. Empirical Prediction.

What is the prediction that differentiates a world governed by reverse causality from a world where out-of-town second-house buyers have a causal impact?

Illustrative Example. To keep the discussion concrete, let's return to the illustrative example from Section 1 where there were 1,000 potential second-house buyers living in Milwaukee and 10,000 living in Los Angeles, 10 times as many as in Milwaukee. Appendix A gives a more fully developed economic model.

Recall that, if reverse causality is at work, h_0, then *a common shock to the fundamental value of owning a second house in Phoenix is both attracting out-of-town second-house buyers and driving up prices.* Under this null hypothesis, potential second-house buyers in both Los Angeles and Milwaukee would see the same shock, namely, the shock to fundamentals, and always increase their demand at the exact same time. As a result, if shocks to housing-market fundamentals in Phoenix always warranted a 10%-point increase in demand, then Phoenix would always see aggregate out-of-town second-house-buyer shocks of either 1,100 buyers or none at all. So, and this is the key observation, under the null hypothesis the relative sizes of Los Angeles and Milwaukee are irrelevant. Because out-of-town second-house buyers would be responding to the same, common, Phoenix-specific shock, they would always show up at the exact same time, and would not matter how they were distributed across the country.

By contrast, according to the alternative hypothesis, h_A, *shocks to fundamentals in Phoenix are not the only thing attracting out-of-town second-house buyers.* In this scenario, potential second-house buyers in Los Angeles and Milwaukee would not in general see the same shocks, so Phoenix could see a 1,000-buyer demand shock from Los Angeles in one month and then a 100-buyer demand shock from Milwaukee in the next. It is only under the alternative hypothesis that the 10%-point increase in demand from the bigger city would have a larger price impact. It is only under this alternative hypothesis that the geographic distribution of potential second-house buyers would matter.

We use this thought experiment to guide our econometric specification. We have to test whether or not 10%-point increases in out-of-town second-house-buyer demand from large and small cities have the same price impact in the target city. If they do, then the data is consistent with reverse causality. If they do not, then the effects of out-of-town second-house-buyer demand cannot be entirely explained by such a story. This is the key prediction.

Econometric Implementation. To test this prediction, we introduce a pair of new variables.

Definition 9 (Number of Out-of-Town Second-House Buyers)**.** *Let Q_i denote the average number of out-of-town second-house purchases made by buyers living in MSA i each month over the period from January 2000 to December 2007 so that $T = 96$:*

$$Q_i = \frac{1}{96} \cdot \sum_{t=1}^{96} \left(\sum_j S_{(i,j),t} \right). \tag{7}$$

Definition 10 (Out-of-Town Second-House-Buyer Share)**.** *Let $\theta_{(i,j),t}$ denote the fraction of second-house buyers in MSA i that purchase a second house in MSA j at time t,*

$$\theta_{(i,j),t} = {}^{S_{(i,j),t}}/_{Q_i}, \tag{8}$$

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

We then proceed in two steps. First, we estimate Equation (9),

$$\Delta \log P_{j,t\to(t+1)} = \alpha_{(i,j)} + \kappa_t + \beta \cdot \Delta \log P_{j,(t-1)\to t} + \gamma \cdot \theta_{(i,j),t} + \varepsilon_{(i,j),t}, \tag{9}$$

which characterizes the relationship between an MSA j's house-price appreciation rate and the fraction of out-of-town second-house buyers from each other MSA $i \neq j$ that choose to purchase a house there. Because we know that out-of-town second-house-buyer demand is positively associated with future house-price appreciation rates, we expect the coefficient γ to be positive. Put differently, under both the null hypothesis, h_0, and the alternative hypothesis, h_A, house-price appreciation rates in Phoenix rise when a larger share of the potential out-of-town second-house buyers in Los Angeles and Milwaukee decide to purchase in Phoenix. The question is: do they rise by the same amount?

To answer this question, we then estimate Equation (10) below,

$$\begin{aligned}\Delta \log P_{j,t\to(t+1)} &= \alpha_{(i,j)} + \kappa_t + \beta \cdot \Delta \log P_{j,(t-1)\to t} + \gamma \cdot \theta_{(i,j),t} \\ &\quad + \delta_{\text{Med}} \cdot 1_i^{\text{Med}} \cdot \theta_{(i,j),t} + \delta_{\text{Lrg}} \cdot 1_i^{\text{Lrg}} \cdot \theta_{(i,j),t} + \varepsilon_{(i,j),t}.\end{aligned} \tag{10}$$

This regression interacts the fraction of all out-of-town second-house buyers from each MSA i that buy second-houses in MSA j, $\theta_{(i,j),t}$, with the size of the home market, MSA i. Specifically, we define the three indicator variables which divide the 21 MSAs in our sample into terciles based on the number of second-house buyers. 1_i^{Sml} is one if MSA i is one of the 7 MSAs with the fewest potential out-of-town second-house buyers and zero otherwise, 1_i^{Med} is one if MSA i is one of the 7 MSAs which have a moderate number of potential out-of-town second-house buyers and zero otherwise, and 1_i^{Lrg} is one if MSA i is one of the 7 MSAs with the most potential out-of-town second-house buyers and zero otherwise.

If the null hypothesis h_0 that a common shock to the fundamental value of owning a second house in Phoenix is both attracting out-of-town second-house buyers and driving up prices is true, then we should find $\delta_{\text{Med}} = \delta_{\text{Lrg}} = 0$. That is, a 1%-point increase in the demand per potential second-house buyer living in Los Angeles (a large market) for second houses in Phoenix should be equally predictive of an increase in house-price appreciation rates as a 1%-point increase in the demand per potential second-house buyer living in Milwaukee (a small market). Under the null hypothesis, after all, we tend to see these demand shock at the same time. By contrast, if the alternative hypothesis h_A that shocks to fundamentals in

Reverse-Causality Specification Using House-Price Apprec. Rates

(a) Dep. Var.: House-Price Appreciation Rate

	Estimate	Std. Error			
Lagged House-Price Appreciation Rate	0.85	0.00	0.02	0.01	0.02
Out-of-Town Second-House-Buyer Share	0.20	0.02	0.05	0.03	0.05
		$\emptyset$	t	(i,j)	$t,(i,j)$

(b) Dep. Var.: House-Price Appreciation Rate

	Estimate	Std. Error			
Lagged House-Price Appreciation Rate	0.85	0.00	0.02	0.01	0.02
Out-of-Town Second-House-Buyer Share	0.05	0.04	0.05	0.04	0.05
Medium MSA × Out-of-Town Buyer Share	0.18	0.06	0.07	0.07	0.08
Large MSA × Out-of-Town Buyer Share	0.31	0.06	0.08	0.07	0.09
		$\emptyset$	t	(i,j)	$t,(i,j)$

Table 10. *Coefficient estimates from Equations (9) and (10). $N = 39{,}900$ monthly observations from February 2000 to December 2007 on the $21 \times 20 = 420$ ordered MSA pairs. Reads: "A 10%-point increase in the fraction of out-of-town second-house buyers purchasing into Phoenix is associated with a 2.0%-point increase in house-price appreciation rates. This impact jumps to 3.6% points if the out-of-town second-house buyers come from Los Angeles, a large MSA, and falls to 0.5% points if they come from Milwaukee, a small MSA."*

Phoenix are not the only thing attracting out-of-town second-house buyers is true, then we should find $\delta_{\mathrm{Lrg}} > \delta_{\mathrm{Med}} > 0$, that the geographic distribution matters.

4.2. **Estimation Results.** We estimate Equations (9) and (10) using a panel dataset at a monthly frequency from February 2000 to December 2007 on the $21 \times 20 = 420$ ordered MSA pairs with all $i = j$ pairs removed, that is, with all local second-house-buyer flows removed. Observations from January 2000 are removed due to the missing 1 month lagged values yielding a balanced panel of 39,900 observations. The data strongly reject the null hypothesis that a common shock to the fundamental value of owning a second house in Phoenix is both attracting out-of-town second-house buyers and driving up prices. The geographic distribution of potential second-house buyers matters in both a statistically measurable as well as an economically meaningful way.

Positive Correlation. Panel (a) in both Table 10 and Table 11 reports the estimated coefficients and standard errors from Equation (9) using both house-price and IAR appreciation rates as the dependent variable. These panels indicate that an increase in the fraction of potential out-of-town second-house buyers investing in an MSA has a positive and statistically significant effect on both measures. The point estimate in Table 10 implies that a 10%-point increase in the fraction of out-of-town second-house buyers purchasing a second house into, say, Las Vegas predicts a 2.0%-point increase in the monthly house-price appreciation rate in Las Vegas. Similarly, Panel (a) in Table 11 says that a 10%-point increase

Reverse-Causality Specification Using IAR Apprec. Rates

(a) Dep. Var.: IAR Appreciation Rate

	Estimate	Std. Error			
Lagged IAR Appreciation Rate	0.45	0.00	0.08	0.02	0.08
Out-of-Town Second-House-Buyer Share	0.81	0.05	0.14	0.09	0.16
		$\emptyset$	t	(i,j)	$t,(i,j)$

(b) Dep. Var.: IAR Appreciation Rate

	Estimate	Std. Error			
Lagged IAR Appreciation Rate	0.45	0.00	0.08	0.02	0.08
Out-of-Town Second-House-Buyer Share	0.41	0.08	0.10	0.09	0.12
Medium MSA $\times$ Out-of-Town Buyer Share	0.48	0.12	0.13	0.18	0.19
Large MSA $\times$ Out-of-Town Buyer Share	0.76	0.12	0.15	0.22	0.25
		$\emptyset$	t	(i,j)	$t,(i,j)$

Table 11. *Coefficient estimates from Equations* (9) *and* (10) *using IAR appreciation rates rather than house-price appreciation rates as the dependent variable.* $N = 39{,}900$ *monthly observations from February 2000 to December 2007 on the* $21 \times 20 = 420$ *ordered MSA pairs. Reads: "A* 10%*-point increase in the fraction of out-of-town second-house buyers purchasing into Phoenix is associated with a* 8.1%*-point increase in IAR appreciation rate. This impact jumps to* 11.7% *points if the out-of-town second-house buyers come from Los Angeles, a large MSA, and falls to* 4.1% *points if they come from Milwaukee, a small MSA."*

in the fraction of out-of-town second-house buyers purchasing a second house in Las Vegas predicts a 8.1%-point increase in the IAR appreciation rate. In other words, when more out-of-town second-house buyers pile into a city, renting begins to look like a more and more attractive option. This is exactly the positive correlation that we should find under both the null hypothesis, h_0, and the alternative hypothesis, h_A.

Geography Matters. Next, if we focus on Panel (b) in both Tables 10 and 11, then we see that the coefficient δ_{Lrg} is statistically different from zero. Comparing the second and fourth rows of Panel (b) in Table 10 reveals that an out-of-town second-house-buyer demand shock from Los Angeles, a large city, has a more than 7 times larger effect on house-price appreciation rates in Phoenix than the same shock coming from Milwaukee, a small city: $7.2 \approx {}^{(0.31\,+\,0.05)}/_{0.05}$. Similarly, comparing the second and fourth rows of Panel (b) in Table 11 shows that an out-of-town second-house-buyer demand shock from Los Angeles, a large city, has an effect on IAR appreciation rates in Phoenix that is almost 3 times as large as the effect of same shock coming from Milwaukee, a small city: $2.9 \approx {}^{(0.76\,+\,0.41)}/_{0.41}$. The ordering of the interaction terms is consistent with the alternative hypothesis, h_A, that demand from distant speculators causes house price and IAR appreciation rates to increase. In all specifications $\delta_{\text{Lrg}} \geq \delta_{\text{Med}} \geq 0$. In all of the specifications the geographic distribution of where out-of-town second-house buyers are coming from matters. What's more, this variation is both statistically significant and economically meaningful, increasing the price impact by 7 fold.

Not Just Momentum. In order to really clarify what we are learning by finding that the geographic distribution of where out-of-town second-house buyers are coming from matters, that is, that $\delta_{\text{Lrg}} \geq \delta_{\text{Med}} \geq 0$ in Tables 10 and 11, it is illustrative to examine how this finding relates to existing results on price momentum in the housing market. After all, house-price appreciation rates are certainly persistent, and Piazzesi and Schneider (2009) use the Michigan Survey of Consumers to show that some home buyers did trade on momentum in the "late" phase of the housing boom, 2004-2005. But, because the geographic distribution of where out-of-town second-house-buyer demand is coming from matters, our results cannot be entirely explained by such a story. If momentum effects were actually the thing attracting out-of-town second-house buyers, then higher-than-usual house-price appreciation rates in Phoenix would be a common shock to the investment opportunity set of potential second-house buyers everywhere, just like a shock to market fundamentals. As a result, we should see a common response. We do not. So, while some out-of-town second-house buyers might be momentum traders, this behavior on its own cannot explain our main results.

4.3. **Alternative Stories.** Of course, there could be other non-city-specific reasons that out-of-town second-house-buyer demand shocks from larger cities might have a larger impact on house-price and IAR appreciation rates than the same shock from smaller cities. Let's now investigate some of these alternative stories in more detail.

Ordered-City-Pair Fixed Effects. We start with the most obvious concern: base rates matter. For instance, second-house buyers living in San Francisco are more likely to invest in Phoenix than Miami and second-house buyers living in New York are more likely to invest in Miami than Phoenix. What's more, second-house buyers living in New York are more likely to invest in Miami than second-house buyers living in Miami are to invest in New York. To account for this potential confound, we include ordered-city-pair fixed effects,

$$\alpha_{(i,j)} = \bar{\alpha}_j - \gamma \cdot \mathrm{E}[\theta_{(i,j),t}], \tag{11}$$

in our main regression specification from Equations (9) and (10).

These ordered-city-pair fixed effects account for the mean house-price or IAR appreciation rate $\bar{\alpha}_j$ in each target MSA j over the sample period. In addition, they adjust the predicted house-price or IAR appreciation rate in the target MSA j for the average rate at which second-house buyers living in MSA i invest in MSA j. For instance, $\gamma \cdot \mathrm{E}[\theta_{(\mathrm{SFO},j),t}]$ differentially controls for the tendency of potential out-of-town second-house buyers living in San Francisco to purchase more houses in Phoenix than in Milwaukee:

$$\mathrm{E}[\theta_{(\mathrm{SFO},\mathrm{PHX}),t}] \neq \mathrm{E}[\theta_{(\mathrm{PHX},\mathrm{SFO}),t}] \neq \mathrm{E}[\theta_{(\mathrm{SFO},\mathrm{MIL}),t}]. \tag{12}$$

Thus, in our main specifications, we really test whether or not a 10%-point increase in out-of-town second-house-buyer demand from Los Angeles into Phoenix relative to the average Los Angeles to Phoenix demand has a large price impact than a 10%-point increase in demand from Milwaukee into Phoenix relative to the average Milwaukee to Phoenix demand. The ordered-city-pair fixed effects mean that the results from Tables 10 and 11 are robust to stories involving constant differences in the rate at which second-house buyers in one city invest in another. Put differently, any confounding explanation has to be time specific.

Home-City-by-Month Fixed Effects. One natural time-specific explanation that you might think about is an income shock. For example, potential buyers living in New York might receive an income shock during the mid 2000s that leads them to have higher demand for investing in second homes relative to potential second-house buyers in other markets. If larger cities were more likely to realize such income shocks, then perhaps out-of-town second-house buyer demand is not actually causing house-price and IAR appreciation rates to rise? Maybe it is just the distribution of income shocks?

To address this concern, we re-run the regressions from Equation (10) with home-city-by-month and target-city fixed effects,

$$\begin{aligned}\Delta \log P_{j,t\to(t+1)} = \alpha_j + \kappa_{(i,t)} + \beta \cdot \Delta \log P_{j,(t-1)\to t} + \gamma \cdot \theta_{(i,j),t} \\ + \ \delta_{\mathrm{Med}} \cdot 1_i^{\mathrm{Med}} \cdot \theta_{(i,j),t} + \delta_{\mathrm{Lrg}} \cdot 1_i^{\mathrm{Lrg}} \cdot \theta_{(i,j),t} + \varepsilon_{(i,j),t},\end{aligned} \tag{13}$$

where, for example, the $\kappa_{(i,t)}$ terms capture the time-varying effect of an income shock to potential second-house buyers in New York on house-price and IAR appreciation rates in other cities. Note that this regression increases the number of fixed-effect coefficients that need to be estimated by four fold, from $(21 \times 20) + 96 = 516$ to $(21 \times 96) + 20 = 2{,}036$. These additional coefficients lower the specification's power and raise its standard errors.

Nevertheless, we find in Table 12 that our previous results are qualitatively unaffected by including home-city-by-month fixed effects. After controlling for the average level of out-of-town second-house-buyer activity each month conducted by traders living in both Los Angeles and Milwaukee respectively, it is still true that a 10%-point increase in demand from Los Angeles, a large MSA, has a larger impact on house-price and IAR appreciation rates in Phoenix than a 10%-point increase in demand from Milwaukee, a small MSA.

What Is Left? It is still possible that there is some confounding effect which both attracts out-of-town second-house buyers and drives up subsequent house-price and IAR appreciation rates, but such a confound would have to mirror the ordered-city-pair by time pattern of out-of-town second-house-buyer demand that we find in the data. Because we find that the geographic distribution of where out-of-town second-house buyers are coming from matters, we know that any confounding effect cannot be target-city specific. Because we include ordered-city-pair fixed effects in our main specification, we know that any confound cannot be specific to the average rate at which out-of-town second-house buyers in each city invest in other cities. Finally, because our results are robust to including home-city-by-time fixed effects, we know that any confound cannot be a result of shocks to large MSAs which push out-of-town second-house buyers there to increase their demand. In short, any confounding explanation has to be a networked phenomenon. It has to mirror the size and timing of the out-of-town second-house-buyer demand between cities.

Of course, it could be the case that information about house-market fundamentals in, say, Phoenix was spreading to potential out-of-town second-house buyers living in other cities in the precisely the way you would need to generate the staggered out-of-town second-house-buyer demand shocks we see in the data. For example, maybe the Los Angeles Times reported information about housing-market fundamentals in Phoenix one month, inducing Los Angeles buyers to step in, and then in the next month the Milwaukee Journal Sentinel reported the same news while the Los Angeles Times was filled with other news and did not devote as much space to the Phoenix housing market. But, the punchline of our analysis is that no simpler explanation will do. Any alternative story has to explain why changes in Phoenix's housing-market fundamentals induce out-of-town second-house buyers from Los Angeles decide to invest at different times than Milwaukee.

Constructive Result. More to the point, the critique that some other variable might exactly mirroring the key right-hand-side variable is not specific to our approach. Suppose we had discovered an instrument which "randomly" assigned out-of-town second-house buyers to different housing markets. We could then compute the average effect of an MSA getting treated with additional out-of-town second-house buyers using more standard instrumental-variable methods. But, even here, the key concern would be whether or not some omitted variable was driving both the instrument and price dynamics.

While we do not characterize the mechanism underlying out-of-town second-house-buyer demand, by exploiting our data's networked structure, we can nevertheless give tight bounds on what that mechanism has to look like. Future econometricians interested in understanding why out-of-town second-house buyers in Los Angeles invested in Phoenix when they did need to look for instruments that shock ordered-city-pair flows at different times, not just shocks to Phoenix's housing-market fundamentals or shocks to out-of-town second-house buyers in Los Angeles' wealth. Similarly, future theorists trying to model out-of-town second-house-buyer demand in more detail need to write down a networked model since city-specific shocks cannot explain our results. Rather than pinning down the suspected mechanism, our approach offers a police sketch to further the investigation.

5. Misinformed Speculators

The last two sections have shown that out-of-town second-house-buyer demand predicts house-price and IAR appreciation rates and that this correlation cannot be explained by

Reverse-Causality Specification with (t,i) and j Fixed Effects

(a) Dep. Var.: House-Price Appreciation Rate

	Estimate	Std. Error			
Lagged House-Price Appreciation Rate	0.86	0.00	0.00	0.03	0.03
Out-of-Town Second-House-Buyer Share	0.03	0.03	0.03	0.03	0.03
Medium MSA × Out-of-Town Buyer Share	0.09	0.04	0.04	0.04	0.04
Large MSA × Out-of-Town Buyer Share	0.10	0.04	0.04	0.05	0.05
		$\emptyset$	(t,i)	j	$(t,i),j$

(b) Dep. Var.: IAR Appreciation Rate

	Estimate	Std. Error			
Lagged IAR Appreciation Rate	0.48	0.00	0.02	0.10	0.10
Out-of-Town Second-House-Buyer Share	0.23	0.06	0.07	0.06	0.07
Medium MSA × Out-of-Town Buyer Share	0.23	0.09	0.10	0.10	0.11
Large MSA × Out-of-Town Buyer Share	0.23	0.08	0.08	0.15	0.15
		$\emptyset$	(t,i)	j	$(t,i),j$

Table 12. *Coefficient estimates from Equation* (10) *with home-city-by-month and target-city fixed effects,* (t,i) *and* j*, using both house-price and IAR appreciation rates as the dependent variable.* $N = 39{,}900$ *monthly observations from February 2000 to December 2007 on the* $21 \times 20 = 420$ *ordered MSA pairs. Reads: "After controlling for the average level of out-of-town second-house-buyer activity each month conducted by traders living in both Los Angeles and Milwaukee respectively, it is still true that a* 10%*-point increase in demand from Los Angeles, a large MSA, has a larger impact on house-price and IAR appreciation rates in Phoenix than a* 10%*-point increase in demand from Milwaukee, a small MSA."*

shocks to fundamentals in the target city. That is, the price impact of out-of-town second-house-buyer demand cannot be explained by reverse causality. We now give some supporting evidence that these out-of-town second-house buyers behaved liked misinformed speculators, earning lower capital gains on their purchases (misinformed) and consuming less of the dividends (speculators)

5.1. **Misinformed.** In many ways it should not be surprising that out-of-town second-house buyers appear uninformed relative to local second-house buyers and owner occupants.

Existing Evidence. By definition, out-of-town second-house buyers live farther away from their purchases than local second-house buyers. Thus, these traders are unlikely to "know the neighborhood" as well as local buyers. In addition, out-of-town second-house buyers face a difficult principal-agent problem when dealing with local real-estate agents who are paid on commission. Levitt and Syverson (2008), for example, find that real-estate agents have substantial discretion in the timing and pricing of sales, with brokers receiving about 3.7%-points more than other owner occupants when selling their own houses. Out-of-town second-house buyers with higher monitoring costs likely face an even larger distortion.

Lower Capital Gains. As more direct evidence, we show that out-of-town second-house buyers had worse exit timing than local second-house buyers. Figure 5 shows the average realized capital gains on single-family-house purchases by local and out-of-town second-house buyers in each MSA during our sample period. 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 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 April 2013 and t. To extend the sample from December 2007 where our data end to April 2013, we append Dataquick data to each property that hasn't sold by December 2007 and assign the house-price appreciation rate through April 2013 to observations that are right censored.

Figure 5 shows that, in key markets such as Las Vegas, Phoenix, and Miami, out-of-town second-house buyers earned significantly lower capital gains on their investments relative to local second-house buyers. For instance, out-of-town second-house buyers investing in Las Vegas in March 2004 earned an 11% per year capital gain on average; by contrast, local second-house buyers purchasing in the same month earned a 22% per year capital gain. In addition, the average capital gain on out-of-town second-house-buyer purchases decreased from 11% per year to -10% per year as out-of-town second-house-buyer demand rose from 5% of all sales in March 2004 to 13% of all sales in January 2007.

What's more, while out-of-town second-house buyers realized 2.42% per year lower capital gains than local second-house buyers in Las Vegas during the entire sample period, this gap is largest for buyers who bought near the peak of the boom. These patterns exist only for "boom" markets and are absent in other markets such as San Francisco and Cleveland

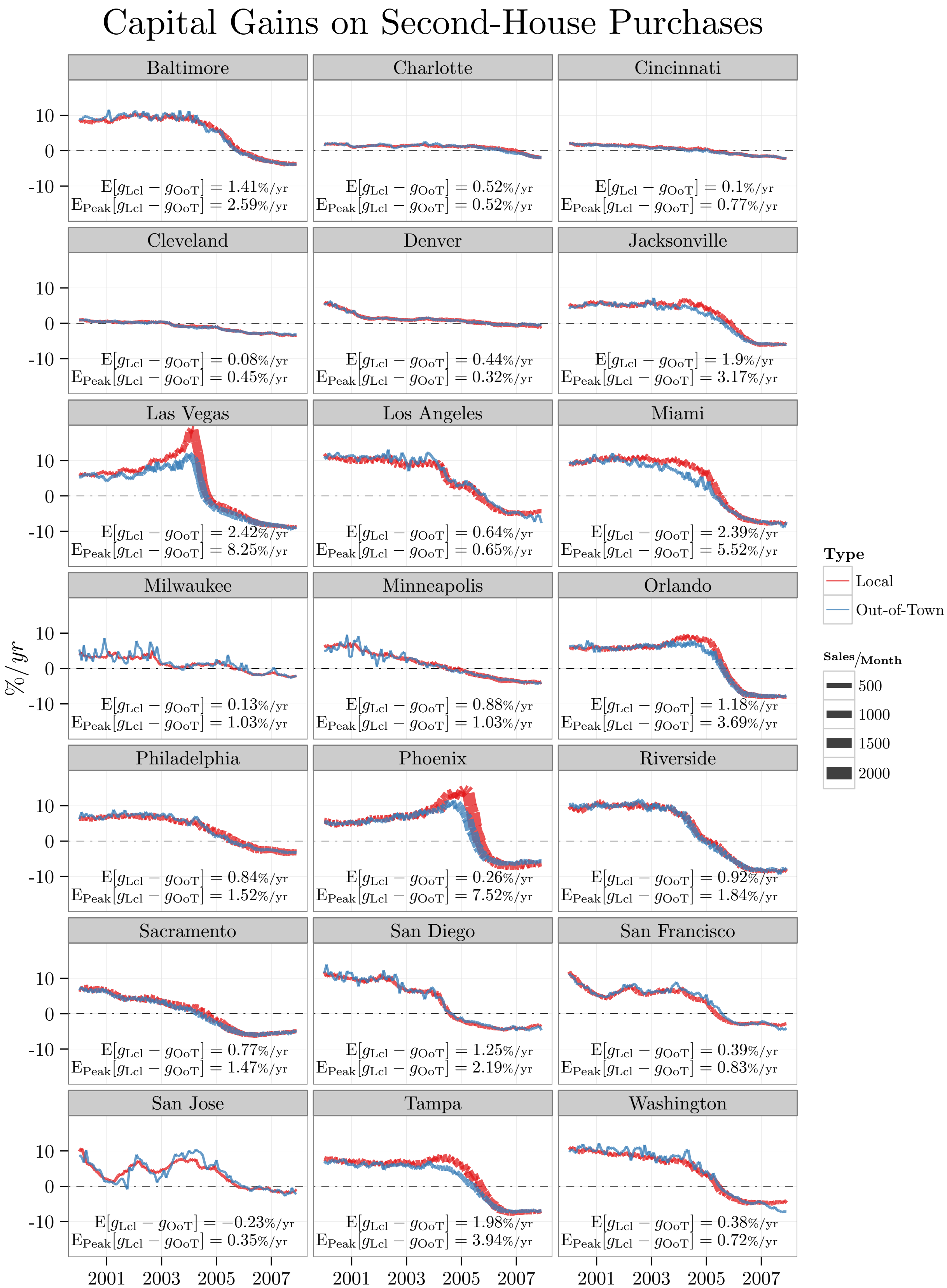


Figure 5. *g_{OoT} and g_{Local} are the realized capital gains for out-of-town and local second-house buyers in units of %/yr. Reads: "Out-of-town second-house buyers investing in Las Vegas in March 2004 realized an 11% per year capital gain; whereas, local second-house buyers realized a 22% per year capital gain."*

which traditionally have either very cyclical or very flat house-price appreciation rates. Of particular interest for this study of mispricing over the cycle, most of the differences between the two groups of buyers are concentrated around the peak of the boom. This finding mirrors the results in Brunnermeier and Nagel (2004) who report that the capital gains of hedge funds (informed/local) in tech stocks during the rise and fall of the Dot-com boom far exceeded the capital gains of the average investor (uninformed/distant). There is a striking similarity between Figure 1 in the current paper and Figure 6 in Brunnermeier and Nagel (2004).

You might be concerned that we only have data through April 2013, so maybe out-of-town second-house buyers were not able to sell right before the market turned, but they were better able to weather the storm and sell their properties as each housing market calmed down? This does not appear to be the case. The discrepancy between the capital gains earned by local and out-of-town second-house buyers increases as we lengthen our sample period. For instance, if we end our sample in December 2007, then the capital gains of local second-house buyers who invested in Las Vegas in March 2004 were only 9%-points per year higher than those of out-of-town second-house buyers.

Regression Analysis. Since both out-of-town and local second-house buyers bought their houses at the same time in Figure 5, the differences in capital gains earned by each group of traders must stem from differences in their respective exit timings. The figure suggests that local second-house buyers in markets such as Las Vegas, Phoenix, and Miami were better able to time the market downturn than out-of-town second-house buyers. To quantify this intuition, we estimate the regression specification in Equation (14):

$$\begin{aligned}
F_{n,i,t-6} = \alpha_i &+ \widehat{\alpha}_i \cdot 1_n^{\text{OoT}} + \kappa_t + \widehat{\kappa}_t \cdot 1_n^{\text{OoT}} + \widehat{\kappa} \cdot 1_n^{\text{OoT}} \\
&+ \beta \cdot \Delta \log P_{i,t\to(t+12)} + \widehat{\beta} \cdot \Delta \log P_{i,t\to(t+12)} \cdot 1_n^{\text{OoT}} \\
&\quad + \gamma \cdot 1_{i,t}^{\text{PostPeak}} + \widehat{\gamma} \cdot 1_{i,t}^{\text{PostPeak}} \cdot 1_n^{\text{OoT}} \\
&\qquad + \delta \cdot \Delta \log P_{i,t\to(t+12)} \cdot 1_{i,t}^{\text{PostPeak}} + \widehat{\delta} \cdot \Delta \log P_{i,t\to(t+12)} \cdot 1_{i,t}^{\text{PostPeak}} \cdot 1_n^{\text{OoT}} \\
&\qquad\quad + \varepsilon_{n,i,t}.
\end{aligned} \tag{14}$$

This regression relates the probability that a second-house buyer "flips" their house within 6 months to (a) the buyer's type (local vs. out-of-town), (b) whether house prices have hit their peak, (c) the extent to which house prices will be rising or falling in the upcoming year, and (d) the interaction of these terms.

If local second-house buyers were better informed about future house-price appreciation rates, then this knowledge should be revealed in the timing of their resale decision. These buyers should be more likely to exit the each market immediately before house-price appreciation rates begin 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%-point per year 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% per year to ensure we are not identifying small local peaks.

Table 13 shows 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 purchase within the 6 months immediately following the peak in local house price appreciation rates, out-of-town second-house buyers are only $4.3 - 3.1 = 1.2\%$ more likely to flip their purchase during this key interval. What's more, a t-test reveals that the point estimate for out-of-town second-house buyers is not statistically different from zero, suggesting that the likelihood of flipping is nearly unchanged for out-of-town second-house buyers immediately after the peak. Table 13 suggests that out-of-town second-house buyers are not using insights about future house-price appreciation rates to strategically exit their investments. Out-of-town second-house buyers appear much less clairvoyant than local second-house buyers.

Exit Timing of Second-House Buyers

Dep. Var.: House resells within 6 Months					
	Estimate	Std. Error			
Out-of-town Second-House Buyer	-0.05	0.02	0.02	0.02	0.02
Future House-Price Appreciation Rate	0.12	0.01	0.03	0.03	0.04
Post Peak Resale	0.04	0.01	0.02	0.02	0.02
Post Peak × Future Apprec. Rate	-0.15	0.05	0.06	0.04	0.07
Out-of-town × Future Apprec. Rate	-0.13	0.02	0.02	0.03	0.04
Out-of-town × Post Peak	-0.03	0.02	0.02	0.02	0.02
Out-of-town × Post Peak × Future Apprec. Rate	0.11	0.07	0.07	0.09	0.11
		$\emptyset$	t	i	$t{,}i$

Table 13. *Estimated coefficients from Equation (14). Resale within* 6 *months: One if purchase in month* $t-6$ *in MSA* i *resells during the interval* $(t-6{,}t]$. *Future house-price appreciation rate: MSA-level house-price appreciation rate from* $t \to (t+12)$ *in units of percent per year. Post peak: One if house-price appreciation rate in MSA peaked in months* $(t-6{,}t]$ *and peak reached* 20% *per year or more.* $N = 1{,}390{,}118$ *monthly observations from July 2000 to December 2007 on all single-family-house sales to local and out-of-town second-house buyers the* 21 *MSAs. Reads: "While local second-house buyers are* 4.3% *more likely to flip their purchase within* 6 *months immediately following the peak, out-of-town second-house buyers are only* $4.3 - 3.1 = 1.2\%$ *more likely to flip their purchase during this key interval, and this point estimate that is not statistically distinguishable from zero."*

5.2. **Speculators.** Out-of-town second-house buyers earn lower capital gains than local second-house buyers, but returns are composed of both capital gains and dividends. We now give evidence that out-of-town second-house buyers are less able to consume the dividends generated by their purchases. It is in this sense that we use the term "speculators."

Motivation for Buying. There are a number of reasons someone might want to buy a second house in a different city: a buyer might want to rent the property out as an additional source of income, live in the house for part of the year, or renovate the house and sell it for a profit at sometime in the future. According to a 2005 survey conducted by the National Association of Realtors,[1] 31% of second-house buyers in 2004 planned on using it as a vacation home while the remaining 69% planned on using it as a rental property. 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.

Out-of-Town Landlords. First, out-of-town second-house buyers who wish to rent out their purchase face higher management and maintenance costs since 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 month's rent plus an additional 8% of the annual rent 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. To put these figures in perspective, imagine you decided to purchase a \$200k house in Las Vegas (the median price in 2004) with a 20% downpayment (the median downpayment for out-of-town second house buyers in 2004) in order to rent it out for \$1000 per month (the median rental price for 2 bedroom houses in Las Vegas in 2004). Over the course of the next 2 years, the monthly property-management fees would amount to 7% of your initial \$40k investment. This is a non-trivial drag on your returns, and even assumes that you would be able to immediately rent out your second-house purchase and keep a tenant for all 24 months. To make matters worse, any second-house buyer (both local and out-of-town) wishing to rent out their property faces the prospect of higher physical-depreciation costs tenants treat houses relatively poorly compared to owner occupants as noted in Harding, Miceli, and Sirmans (2000).

Part-Time Residents. It is even more obvious that part-time residents do not consume all of the dividends from their purchase since they only live in the house part of the year. You might be concerned that out-of-town second-house buyers are rich and simply getting lots of utility from owning a vacation house in Phoenix or a weekend getaway in Miami, but this does not appear to be the case. We examine the price of the house that is the primary residence of out-of-town second-house buyers in the highest income cities including San Francisco, San Jose, and New York. Figure 6 shows that, in January 2005, the median

[1] Investment and Vacation Home Buyers Survey. *National Association of Realtors*, Mar. 6 2005.

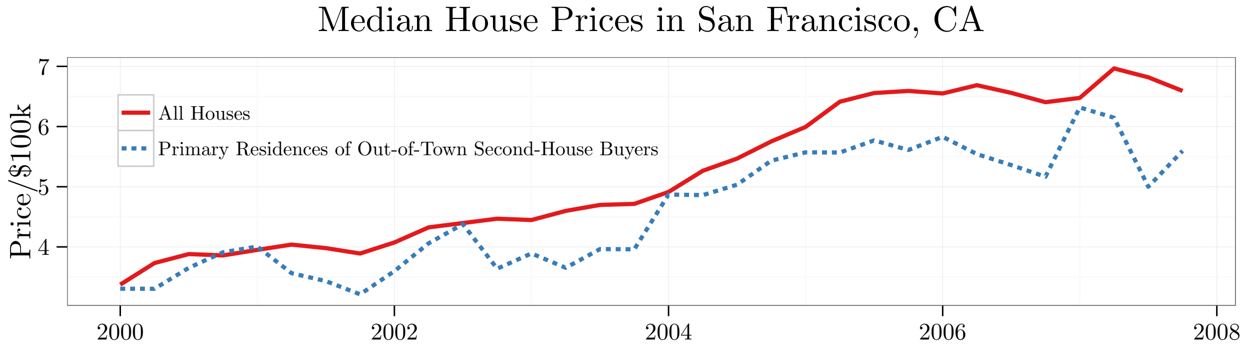


Figure 6. *Median price of the primary residences of out-of-town second-house buyers and of all buyers living in San Francisco in units of* \$100k *over the time period from January 2000 to December 2007. The price of the primary residences of out-of-town second house buyers living in San Francisco is computed by scaling up the most recent sale price by price index. Reads: "In January 2005, the median value of all single-family-houses purchased in San Francisco was* \$600k*. By contrast, the median value of the primary residences of out-of-town second-house buyers who lived in San Francisco and bought a second house in another MSA in January 2005 was only* \$555k*."*

value of all single-family-houses purchased in San Francisco was \$600k. By contrast, the median value of the primary residences of out-of-town second-house buyers who lived in San Francisco and bought a second house in another MSA in January 2005 was only \$555k. While the value of their primary residence is not a perfect proxy for out-of-town second house buyers' wealth, these results are inconsistent with a story based on wealth effects.

6. Conclusion

We conclude by discussing some alternative applications of our identification strategy and giving a sense of the scale of out-of-town second-house-buyer demand shocks.

Alternative Applications. This paper introduces a new approach to addressing concerns about reverse causality. This approach can be used in other situations where researchers suspect that misinformed speculators might be destabilizing prices but natural instruments for speculative demand are lacking. Consider the recent housing boom in Spain that also appears to have been fueled by purchases of foreign speculators. Office for National Statistics (2007) found that at peak 1.8Mil households in England owned a second house and, among these properties, 87k were in Spain. More recently, there has been a great deal of discussion about the role of foreign buyers in driving up prices in Canadian markets like Toronto and Vancouver.[2] Similar events unfolded 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, inviting a host of foreign investors from countries like Japan. For a detailed example, see Sagalyn (1999), which discuss the purchase of Rockefeller Center by

[2]James Surowiecki. Real Estate Goes Global. *The New Yorker*, May. 26 2014.

Mitsubishi Trust, Co. for more than \$1Bil in the late 1980s.

But, there are numerous other applicable settings outside the confines of real-estate finance. For example, researchers have long puzzled over the "Asian economic miracle" and subsequent crisis. What was the explanation for this boom and bust cycle? On one hand, it could have been, h_0, that *a common shock to the fundamental value of investing in Southeast Asian countries both attracted foreign capital and driving up prices.* In this scenario, potential investors in the US and England see the same shock (the shock to fundamentals), and would always increase their Southeast Asian demand at the exact same time, making the relative sizes of the US and English pools of investable funds irrelevant. It does not matter how foreign capital is distributed across the globe because investors in each country are responding to a common Southeast Asia-specific shock. There is also the alternative hypothesis, h_A, that *shocks to fundamentals in places like Hong Kong and Thailand were not the only thing attracting foreign capital to Southeast Asia during the mid 1990s.* For example, the New York Times and Financial Times might have had differing levels of positive Southeast Asia sentiment. In this scenario, potential foreign investors in the US and England would not in general see the same shocks. It is only under the alternative hypothesis that the geographic distribution of resources matters.

Sense of Scale. How big were out-of-town second-house-buyer demand flows? Figure 7 plots the sum of the sales prices on out-of-town second-house-buyer purchases as a percent of MSA-specific gross domestic product from 2000 to 2007 using data comes from the Bureau of Economic Analysis. These calculations treat all purchases as being new capital coming from outside the MSA, whether financed by debt or equity since the majority of debt financing came via selling residential mortgage-backed securities (RMBS). This figure shows that out-of-town second-house-buyer demand in Las Vegas exceeded 5% of Las Vegas's GDP in 2004. Demand shocks from out-of-town second-house buyers are enormous relative to the economic output of many MSA-level economies.

Alternatively, consider the Spanish experience. In Panel (a) of Figure 8 we plot the foreign direct investment (FDI) in Spain as a percent of Spain's GDP from 2003 to 2010 using data from the World Bank. In Panel (b) we plot the real HPI level in Spain over this same time period. Just like in Las Vegas, we find that FDI as a percent of GDP spikes to just under 5% in 2008 and that the timing of this spike corresponds to the peak of the HPI level. As a point of comparison, 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. Building on this analogy to FDI, we see an opportunity in future work to study the impact of these spillovers from the residential-housing market on local economies.

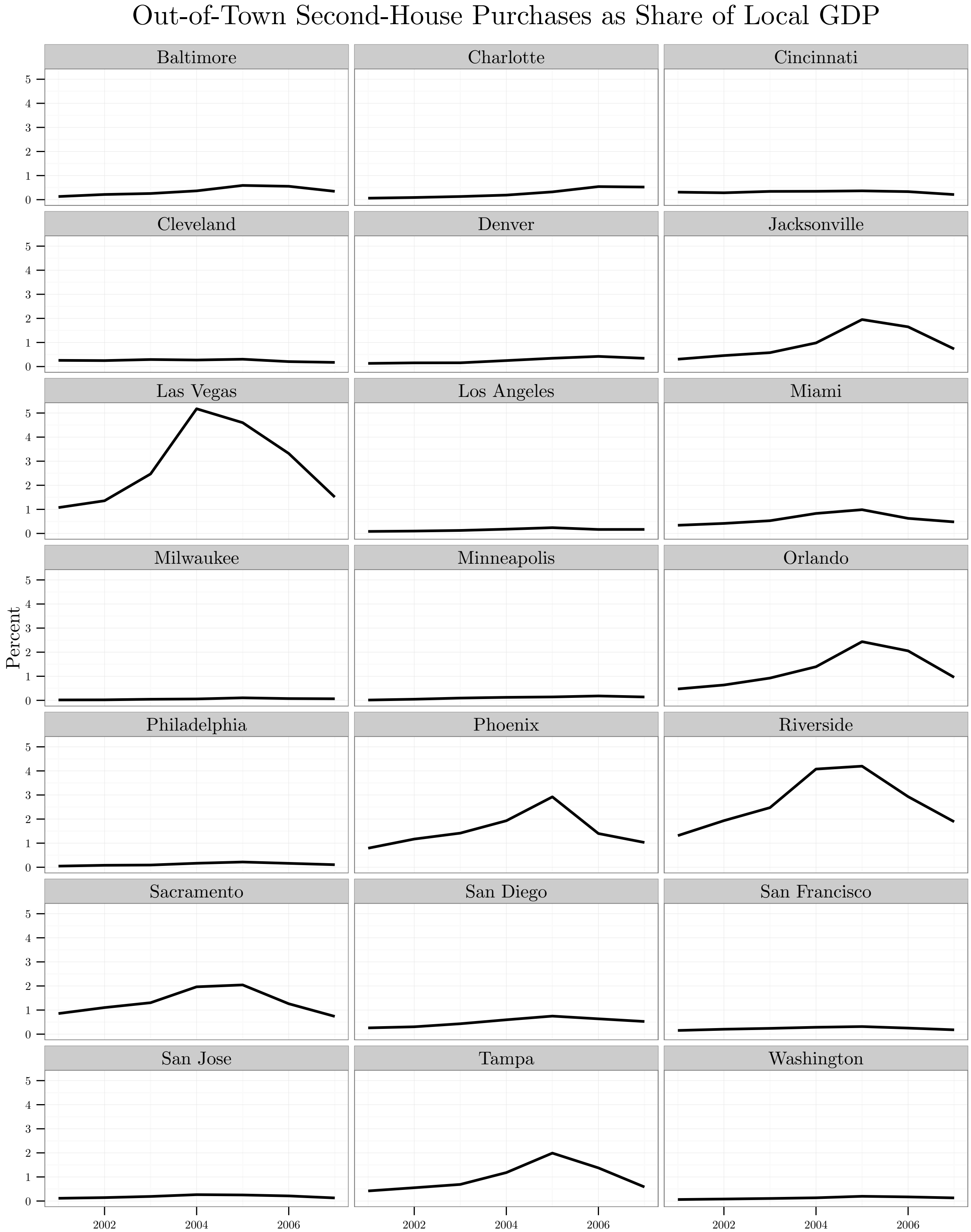


Figure 7. *We compute MSA-specific GDP using data from the BEA as the product of the per capita income in each MSA times the population. Reads: "The sum of the sales prices on out-of-town second-house buyers' purchases in Las Vegas exceeded* 5% *of the GDP for the entire MSA in 2004."*

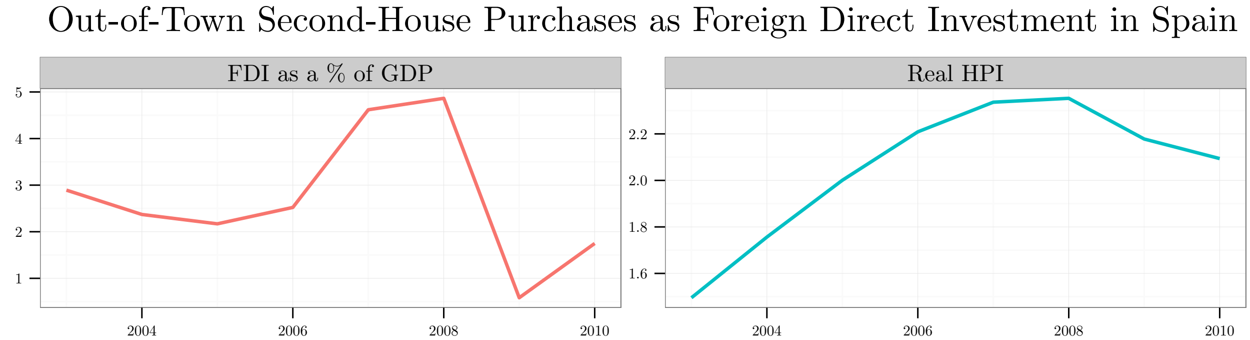


Figure 8. *Left Panel: Net foreign direct investment (FDI) in Spain from the World Bank as a percent of Spain's GDP from 2003 to 2010. Reads: "Net FDI inflows into Spain amounted to a little less than* 5% *of Spain's GDP in 2008." Right Panel: Real HPI index level in Spain over this same time period. Reads: "The real HPI index level rose by just over* 230%*."*

References

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

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

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

Cohen, L., A. Frazzini, and C. Malloy (2010). Sell side school ties. *Journal of Finance 65*.

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

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

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

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

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

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

Gyourko, G., C. Mayer, and T. Sinai (2013). Superstar cities. *The American Economic Journal: Economic Policy 5*.

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

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

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

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

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

Kendall, M. (1945). Note on bias in the estimation of autocorrelations. *Biometrika 41*.

Kurlat, P. and J. Stroebel (2014). Testing for information asymmetries in real estate markets. *SSRN Working Paper*.

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

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

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

Lu, W., P. Swan, and J. Westerholm (2015). Other people's money: Mum and dad investors vs the professionals. *Working Paper*.

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

Marriott, F. and J. Pope (1945). Bias in the estimation of autocorrelations. *Biometrika 41*.

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

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

Office for National Statistics (2007). Community housing survey.

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

Piazzesi, M. and M. Schneider (2009). Momentum traders in the housing market: Survey evidence and a search model. *The American Economic Review 99*.

Piskorski, T., A. Seru, and J. Witkin (2014). Asset quality misrepresentation by financial intermediaries: Evidence from rmbs. *NBER Working Paper*.

Poterba, J. (1984). Tax subsidies to owner-occupied housing: An asset-market approach. *The Quarterly Journal of Economics 99*.

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

Thompson, S. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics 99*.

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

Appendix A. A Simple Model of Speculation

We develop a simple noisy-rational-expectations model of the US residential housing market to clarify our identification strategy and better interpret our empirical results.

A.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 in dollars per house. We model the true value of housing in each MSA i as an iid 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 $\vartheta_{q,(i,j)}$ denote the number of houses in MSA j demanded by the q^{th} trader in MSA i and let $S_{(i,j)}$ 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 $\theta_{(i,j)} = {}^1\!/_{Q_i} \cdot \sum_{q=1}^{Q_i} \vartheta_{q,(i,j)}$ 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:

$$X_j = \sum_{i=1}^{I} S_{(i,j)} + \varepsilon_j = \sum_{i=1}^{I} \left(\sum_{q=1}^{Q_i} \vartheta_{q,(i,j)} \right) + \varepsilon_j = \sum_{i=1}^{I} \big(Q_i \cdot \theta_{(i,j)} \big) + \varepsilon_j \tag{15}$$

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:

$$P_j = \mathrm{E}\left[V_j | X_j\right] = \alpha + \beta \cdot X_j \tag{16}$$

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 that traders use a linear demand rule given by:

$$\vartheta_{q,(i,j)} = \gamma_{q,(i,j)} + \delta_{q,(i,j)} \cdot V_j \tag{17}$$

This is the standard ansatz for Kyle (1985) type models and can easily be verified in equilibrium. The coefficient $\gamma_{q,(i,j)}$ has units of houses per trader and the coefficient $\delta_{q,(i,j)}$ 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:

$$\begin{aligned} W_{q,i} &= \sum_{j=1}^{I} W_{q,(i,j)} \\ W_{q,(i,j)} &= \max_{\vartheta_{q,(i,j)}} \mathrm{E}\left[(V_j - P_j) \cdot \vartheta_{q,(i,j)} \middle| V_j \right] \end{aligned} \tag{18}$$

Definition 11 (Equilibrium). *An equilibrium consists of price parameters* (α^*,β^*) *and demand parameters* $\{(\gamma^*_{q,(i,j)},\delta^*_{q,(i,j)})\}$ *for each trader over every ordered MSA pair such that:*

(1) Given market makers follow the pricing rule in Equation (16), the housing demand schedule $\{\vartheta_{q,(i,j)}\}_{i,j\in I}$ *dictated by the demand rule parameters* $\{(\gamma^*_{q,(i,j)},\delta^*_{q,(i,j)})\}_{i,j\in I}$ *solves each trader's optimization problem in Equation (18).*

(2) Given all traders follow the demand rules specified in Equation (17), the price parameters (α^*,β^*) *satisfy the expectations equality in Equation (16).*

A.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 $\theta_{(i,j)}$ as well as the price impact of each trader's demand decisions $(\gamma_{(i,j)},\delta_{(i,j)})$ 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 number of houses in MSA* j*:*

$$\bar{\theta}_j = \theta_{(1,j)} = \theta_{(2,j)} = \cdots = \theta_{(i,j)} \tag{19}$$

Proof. Substituting both the functional form for the housing price in MSA j from Equation (16) and the functional form for the aggregate demand in MSA j from Equation (15) into the objective function for an individual trader q from MSA i yields an expression:

$$\begin{aligned} W_{q,(i,j)} &= \max_{\vartheta_{q,(i,j)}} \mathrm{E}\left[(V_j - \alpha - \beta\cdot X_j)\cdot\vartheta_{q,(i,j)}\middle| V_j\right] \\ &= \max_{\vartheta_{q,(i,j)}} \mathrm{E}\left[\left(V_j - \alpha - \beta\cdot\sum_{i'=1}^{I}\left(\sum_{q'=1}^{Q_{i'}}\vartheta_{q',(i',j)}\right) - \beta\cdot\varepsilon_j\right)\cdot\vartheta_{q,(i,j)}\middle| V_j\right] \end{aligned} \tag{20}$$

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

$$0 = \mathrm{E}\left[\left(V_j - \alpha - \beta\cdot\sum_{i'=1}^{I}\left(\sum_{q'=1}^{Q_{i'}}\vartheta_{q',(i',j)}\right) - \beta\cdot\varepsilon_j\right) - 2\cdot\beta\cdot\vartheta_{q,(i,j)}\middle| V_j\right] \tag{21}$$

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

$$0 = V_j - \alpha - \beta\cdot\sum_{i'=1}^{I}\left(\sum_{q=1}^{Q_{i'}}\vartheta_{q,(i',j)}\right) - 2\cdot\beta\cdot\vartheta_{q,(i,j)} \tag{22}$$

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

$$\vartheta_{q,(i,j)} = -\frac{\alpha + \beta\cdot\sum_{i'=1}^{I}\left(\sum_{q'=1}^{Q_{i'}}\vartheta_{q',(i',j)}\right)}{2\cdot\beta} + \left(\frac{1}{2\cdot\beta}\right)\cdot V_j \tag{23}$$

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

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:*

$$\tilde{P}_j^{(i)} - P_j = \left(\frac{Q_i}{\sum_{i'=1}^{I} Q_{i'}} \right) \cdot \frac{\eta}{2} \tag{24}$$

Proof. 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:

$$\mathrm{E}\left[P_j | \mathrm{MSA}\right] = \begin{cases} \alpha^* + \beta^* \cdot \sum_{i'=1}^{I} Q_{i'} \cdot \left(\bar{\gamma}^* + \bar{\delta}^* \cdot V_j\right) & \text{if MSA} \neq i \\ \alpha^* + \beta^* \cdot \sum_{i'=1}^{I} Q_{i'} \cdot \left(\bar{\gamma}^* + \bar{\delta}^* \cdot \{V_j + \eta\}\right) & \text{if MSA} = i \end{cases} \tag{25}$$

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:

$$\begin{aligned} \tilde{X}_j^{(i)} &= \sum_{i' \neq i} Q_{i'} \cdot \left(\bar{\gamma}^* + \bar{\delta}^* \cdot V_j\right) + Q_i \cdot \left(\bar{\gamma}^* + \bar{\delta}^* \cdot \{V_j + \eta\}\right) \\ &= \sum_{i'=1}^{I} Q_{i'} \cdot \left(\bar{\gamma}^* + \bar{\delta}^* \cdot V_j\right) + Q_i \cdot \bar{\delta}^* \cdot \eta \end{aligned} \tag{26}$$

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. □

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:

$$\tilde{P}_j^{(\mathrm{MSA})} - P_j = \begin{cases} \left(\frac{10\times 10^6}{55\times 10^6}\right) \cdot \frac{\$5000}{2} = \$454.55 & \text{if MSA} = i' \\ \left(\frac{1\times 10^6}{55\times 10^6}\right) \cdot \frac{\$5000}{2} = \$45.45 & \text{if MSA} = i'' \end{cases} \tag{27}$$

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.

A.3. **Empirical Strategy.** The goal of this simple model is to provide a scaffolding within which to better understand the empirical strategy we employ. First, we identify a group of misinformed 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 $\eta > 0$. We give evidence that out-of-town second-house buyers satisfy this criteria. Second, we show that an increase in demand from out-of-town second-house buyers predicts increases in house-price and IAR appreciation rates. Within the model, this task is tantamount to checking 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 out-of-town second house buyers have above average demand. 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 address the issue of reverse causality. Within the model, this task corresponds to checking whether or not high realized prices in MSA j are due to high realized housing values 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. 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 out-of-town second-house buyers going from MSA i to MSA j is bigger when the total number of out-of-town second-house buyers living in MSA i is larger. These results are 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 distortion.

Appendix B. Early Ranking Period

Perhaps the size of potential out-of-town second-house buyers' belief distortions are random variables that move over time? In such a world, covariance between the home MSA size and the size of potential out-of-town second-house buyers' belief distortion may bias our results. To address this concern, in Table 14 we again re-run our analysis, only this time we instead compute the number of potential out-of-town second-house buyers living in each MSA using the ranking in 2000. Let $\widehat{Q}_i$ denote the number of potential out-of-town second house buyers living in MSA i similarly defined but measured over the period from January

2000 to December 2000 so that $T = 12$:

$$\widehat{Q}_i = \frac{1}{12} \cdot \sum_{t=1}^{12} \left(\sum_{i \neq j} S_{(i,j),t} \right) \tag{28}$$

By contrast, the first definition of the number of potential second-house buyers in each MSA i represents the sample average over the entire period from January 2000 to December 2007.

Since this variable is computed using the entire time series, it is potentially simultaneously determined with investment opportunities in the largest markets for out-of-town second house buyers that appear attractive later in the sample period. For instance, some out-of-town second-house buyers 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. Table 14 confirms that our results do not qualitatively change when we account for this problem.

Appendix C. Preference Heterogeneity

Even if out-of-town second-house buyers are both less informed and less able to consume the dividend from their purchase than local second-house buyers, they would not be misinformed speculators if they had completely different motivations for making their purchases.

Reverse-Causality Specification with Early Ranking Period

(a) Dep. Var.: House-Price Appreciation Rate

	Estimate	Std. Error			
Lagged House-Price Appreciation Rate	0.85	0.00	0.02	0.01	0.02
Out-of-Town Second-House-Buyer Share	0.04	0.02	0.02	0.01	0.02
Medium MSA × Out-of-Town Buyer Share	0.03	0.02	0.01	0.02	0.02
Large MSA × Out-of-Town Buyer Share	0.15	0.03	0.04	0.05	0.05
		$\emptyset$	t	(i,j)	$t,(i,j)$

(b) Dep. Var.: IAR Appreciation Rate

	Estimate	Std. Error			
Lagged IAR Appreciation Rate	0.45	0.00	0.08	0.02	0.08
Out-of-Town Second-House-Buyer Share	0.23	0.03	0.05	0.03	0.05
Medium MSA × Out-of-Town Buyer Share	−0.01	0.04	0.03	0.06	0.06
Large MSA × Out-of-Town Buyer Share	0.42	0.06	0.06	0.14	0.14
		$\emptyset$	t	(i,j)	$t,(i,j)$

Table 14. *Coefficient estimates from Equation (10) where the number of potential second-house buyers in each city, Q_i, is estimated over the period from January 2000 to December 2000 using both house-price and IAR appreciation rates as the dependent variable.* $N =$ 39,900 *monthly observations from February 2000 to December 2007 on the* $21 \times 20 = 420$ *ordered MSA pairs. Reads: "When ranking MSAs according to their potential out-of-town second-house-buyer populations in the year 2000, a* 10%*-point increase in demand from Los Angeles, a large MSA, has a larger impact on house-price and IAR appreciation rates in Phoenix than a* 10%*-point increase in demand from Milwaukee, a small MSA."*

For example, a recent New York Times article on the marriage market in China[3] reported that "70% of single women said they would tie the knot only with a prospective husband who owned a home." Obviously, we would not want to classify Chinese bachelors as misinformed speculators if we found that they were willing to overpay for housing and tended not to live in their purchase until married. This is not what is going on in the US. An out-of-town second-house buyer in our data looks very much like an average home buyer except for the fact that she has decided to treat housing as a financial asset and make a speculative bet.

First, perhaps out-of-town second-house buyers are just soon-to-be retirees looking to buy a house to retire in? For instance, a 2004 Money Magazine article[4] writes that "every 8 seconds, a boomer turns 50... so if you're thinking about retiring in a new home—whether it's your only base or one of two—now's the time to find it. 'There's a concern that the prices of properties most desired by boomers may get out of reach,' says David Hehman, CEO of EscapeHomes.com, a San Francisco realty firm specializing in second homes... Our advice: Buy now, retire later." Do we really want to call such a buyer a misinformed speculator? In a word: yes. There is no rule saying that misinformed speculators can not sport bifocals and wear scratchy sweaters. If you replace the word "properties" with "tech stocks" the article may well have been written 4 years earlier. Baby boomers that bought a second house in Phoenix or Las Vegas with the goal of waiting a few years and then retiring there made a calculated bet on future house prices and certainly did not consume the full dividend from their housing purchase while still employed.

Next, consider the possibility that out-of-town second-house buyers are interested in the diversification benefits of owning a second house in a market where returns are less correlated with other assets in the portfolio. For example, Lustig and Van Nieuwerburgh (2005) gives empirical evidence of a housing capital risk premia due to the covariance of its returns with the returns to the household's human capital. Nevertheless, diversification can not be the main motivation for out-of-town second-house buyers in our sample because out-of-town second-house buyers tended to own several houses in MSAs with highly correlated price dynamics. We calculate that the typical out-of-town second-house buyer who owned an investment property in Los Angeles in January 2006 owned 2.1 houses in addition to his primary residence—he owned his primary residence, the second house in Los Angeles that he bought in January 2006, and 1 additional house that he bought in previous months.

It is not even clear that out-of-town second-house buyers who only bought 1 additional house were interested in diversification. Imagine you are a home owner living in Los Angeles who is considering buying a second house in Phoenix using a \$40k downpayment. If you do not make this purchase, your utility is

$$U_{\text{Don't Invest}} = \mathrm{E}\,[\,\Delta P_{\text{LA}} + D_{\text{LA}}\,] - \frac{\gamma}{2} \cdot \mathrm{Var}\,[\,\Delta P_{\text{LA}} + D_{\text{LA}}\,]\,, \tag{29}$$

where ΔP_{LA} denotes the dollar house-price appreciation in Los Angeles, and D_{LA} denotes the housing dividend paid out by your primary residence in Los Angeles. If you do make the purchase, your utility is

$$\begin{aligned} U_{\text{Invest}} = \mathrm{E}\,[\,(\Delta P_{\text{LA}} + D_{\text{LA}}) + (\{\Delta P_{\text{Phx}} - \kappa\} + \{D_{\text{Phx}} - \lambda\})\,] - \$40\text{k} \\ - \frac{\gamma}{2} \cdot \mathrm{Var}\,[\,(\Delta P_{\text{LA}} + D_{\text{LA}}) + (\{\Delta P_{\text{Phx}} - \kappa\} + \{D_{\text{Phx}} - \lambda\})\,]\,, \end{aligned} \tag{30}$$

[3] Andrew Jacobs. For many Chinese men, no deed means no dates. *The New York Times*, Apr. 14 2011.
[4] Marion Asnes. Buy Now, Retire Later. *Money Magazine*, Jul. 1 2004.

where ΔP_{Phx} denotes the dollar house-price appreciation in Phoenix, κ denotes the wedge between the house-price appreciation rates earned by out-of-town and local second-house buyers, D_{Phx} denotes the housing dividend paid out by the second house in Phoenix, and λ denotes the dividend loss due to being an out-of-town second-house buyer.

On a \$200k house the dollar house price appreciation wedge would be about $\kappa = 0.09 \times \$200\text{k} = \$18\text{k}$ at peak. The dividend drag of being an out-of-town owner would be around $\lambda = \$2\text{k}$ in the first year of ownership. Thus, in order for the second-house purchase in Phoenix to make sense for you, it has to be the case that:

$$\$20\text{k} < -\,\gamma \cdot \rho_{\text{LA,Phx}} \cdot \sigma_{\text{LA}} \cdot \sigma_{\text{Phx}} \tag{31}$$

after setting $\$40\text{k} = \mu_{\text{Phx}} + D_{\text{Phx}} - \frac{\gamma}{2} \cdot \sigma^2_{\text{Phx}}$ since there is an active market in Phoenix.

But, on a \$500k primary residence in Los Angeles, the yearly standard deviation in dollar price appreciation over the time period from 1996 to 2011 is $\sigma_{\text{LA}} = \$60\text{k}$. On a \$200k second house in Phoenix, the yearly standard deviation in dollar price appreciation over the same time period is $\sigma_{\text{Phx}} = \$30\text{k}$. Thus, for a risk-aversion coefficient $\gamma = 2$, there just needs to be a really weak negative correlation of -0.01 between dollar house-price appreciation rates in Los Angeles and Phoenix for the second house in Phoenix to be a good investment. Here is the punchline: the actual correlation is extremely positive at $\rho_{\text{LA,Phx}} = 0.81$! Buying a second house in Phoenix just is not a good hedge against house price fluctuations in Los Angeles.

Diversification benefits surely play some role in determining macro-level house price dynamics. However, the previous paragraphs suggest they are unlikely to be key factors governing the choices of out-of-town second-house buyers. This is unlikely to be what is driving out-of-town second-house buyer demand.