Speculative Fever: Investor Contagion in the Housing Bubble

By Patrick Bayer, Kyle Mangum and James W. Roberts*

Historical anecdotes of new investors being drawn into a booming asset market, only to suffer when the market turns, abound. While the role of investor contagion in asset bubbles has been explored extensively in the theoretical literature, causal empirical evidence on the topic is much rarer. This paper studies the recent boom and bust in the U.S. housing market and establishes that many novice investors entered the market as a direct result of observing investing activity of multiple forms in their own neighborhoods and that "infected" investors performed poorly relative to other investors along several dimensions.
JEL: D40, D84, R30
Keywords: Speculation, Housing Markets, Asset Pricing, Financial Intermediaries, Asset Bubbles, Contagion

Historical accounts of well-known financial boom and bust episodes have drawn attention to several phenomena that appear to signify and contribute to asset bubbles. A common observation is that market participation tends to broaden significantly during a speculative boom, as investors with limited experience or expertise are drawn into the market. In his famous description of the boom and bust in the 1637 Dutch tulip market, for example, Mackay (1841) commented that at its peak, "Nobles, citizens, farmers, mechanics, seamen, footmen, maid-servants, even chimney-sweeps and old clotheswomen, dabbled in tulips."[1]

Such a "speculative fever" is widely viewed as symptomatic of bubble-like episodes and financial crises,[2] and many modern theoretical models of asset bubbles characterize both rational and irrational herd behavior capable of generating exactly the sort of investor contagion described in these historical accounts.[3]

* Bayer: Department of Economics, Duke University, 213 Social Sciences Building, 419 Chapel Drive, Durham, NC 27708-0097, and NBER (email: patrick.bayer@duke.edu); Mangum: Federal Reserve Bank of Philadelphia, 10 Independence Mall, Philadelphia, PA 19106, (email: kmangum104@gmail.com); Roberts: Department of Economics, Duke University, 213 Social Sciences Building, 419 Chapel Drive, Durham, NC 27708-0097, and NBER (email:j.roberts@duke.edu). We thank Jerry Carlino, Chris Cunningham, Kris Gerardi, Steve Ross, Alex Zevelev, and many seminar and conference participants for their useful feedback on earlier versions of this paper. Yicheng Liu and Mingrui Ma provided excellent research assistance. Any errors are our own. The views of this paper are those of the authors and do not necessarily represent the position of FRB Philadelphia or any other part of the Federal Reserve System.

[1]Similarly, in his "anatomy of a typical crisis", Kindelberger (1978) notes that financial market bubbles are frequently characterized by "More and more firms and households that previously had been aloof from these speculative ventures" beginning to participate in the market.

[2]See, for example, Calvo and Mendoza (1996), Basu (2002), Chari and Kehoe (2003), Burnside, Eichenbaum and Rebelo (2016).

[3]See, for example, DeLong et al. (1990), Scharfstein and Stein (1990), Shleifer and Summers (1990),

These models typically characterize a fundamental information problem in which rational investors use the activity of others to learn about movements in market fundamentals, but also often include a subset of naïve agents (e.g., noise traders) that engage in herd behavior for reasons that may not be entirely motivated by rational decision-making.[4] Despite the long-standing theoretical and practical interest in asset bubbles in general, and investor herd behavior in particular, the existing empirical evidence on investor contagion has been very limited, rarely moving beyond anecdotal accounts or a characterization of the observed correlation in investor behavior.

In this paper we study individual investor behavior in the recent housing boom and bust in the U.S. and, in so doing, provide some of the first causal evidence on the causes and consequences of investor contagion.[5] In our attempt to move beyond anecdotal characterization of this phenomenon, we aim to achieve three primary goals: establishing a causal effect of others' investment activity on the likelihood that an individual becomes an investor in the housing market; quantifying the contribution of investor contagion to the overall amount of speculative investing in the housing market; and comparing the performance of "infected" investors, who are drawn into the market, to that of professional, "non-infected" investors as a way of gauging the relative sophistication of these new investors and the consequences of the contagion we document.

Four important features of the U.S. housing market make it a compelling and particularly well-suited setting for studying investor contagion. First, the housing market experienced a substantial rise and fall over the 2000s, with housing prices increasing 50 percent nationally, and upwards of 200 percent in some metropolitan areas, before tumbling back to roughly pre-boom levels by the end of the decade. Second, there was a great deal of speculation in the housing market during the boom. At the height of the market from 2004-2006, for example, Haughwout et al. (2011) estimate that 40-50 percent of all homes in the states that experienced the largest housing booms were purchased as investment properties.[6] Moreover, as we show in this paper, and very much in line with the above-mentioned historical accounts of new investors entering during booms, much of this investment was made by new investors. Third, housing transactions are a matter of public record, as the deed for each home, along with any liens on the property, must be recorded at the time of purchase. As a result, the universe of home purchases, including transaction price and buyer and seller names, is available for nearly all markets in the U.S.; comparatively, accessing comprehensive individual investment data for other financial markets is typically more challenging. Fourth, the geographical

Topol (1991), Froot, Scharfstein and Stein (1992), Lux (1995), Lux (1998), Allen and Gale (2000), Morris (2000), Corcos et al. (2002), Scheinkman and Xiong (2003), Prasanna and Kapadia (2010).

[4]See Banerjee (1992), Kirman (1993), Orlean (1995), Shiller (1995), Chamley (2004), and Jackson (2010) for a broader characterization of rational herd behavior in economic models.

[5]Below we describe the existing empirical literature on the topic.

[6]Using a different methodology, Bayer et al. (Forthcoming) estimates a similar jump in purchases by novice investors in the Los Angeles metropolitan area over this same period.

nature of the housing market provides us with a natural way to identify channels through which contagion may occur, as potential investors may take cues from nearby real estate activity (our paper's estimates will indicate that this is in fact the case). In contrast, were we to consider stock market portfolio decisions, for example, we would need a compelling way to designate from whence an investor may contract such a speculative fever. While this may be possible for certain subsets of the population (for example one might be able to get direct measures of investors' peers for a select group of investors[7]), it will generally be a difficult, if not impossible, task to determine the set of other investors' portfolios that any given investor has knowledge about.

Our focus is on individuals who purchase houses for investment purposes. We aim to identify cases where individual investors are drawn into the market because of the activity of other investors. To provide causal evidence on this type contagion, we utilize a nearest-neighbor research design that identifies the causal effect of nearby investment activity on a potential investor's behavior by estimating the impact of hyper-local investment activity (on his or her residential block), while controlling for similar measures of activity at a slightly larger neighborhood (on other nearby blocks). This type of research design has been used extensively in the recent empirical literature on neighborhood effects to identify a variety of spatial spillovers including employment referrals, foreclosures, and school choice, to name a few examples.[8] This approach to establishing causality leverages another feature of the housing market useful for our purposes: an individual's ability to purchase a house on one specific block versus the next is largely driven by the availability of homes at the time of purchase.[9] This sharply limits household sorting at the block level (a fact that we confirm in our data below), and thus largely mitigates the concern that a positive effect of very nearby investment activity on an individual's likelihood of becoming an investor represents only a spurious correlation. In the analysis below we provide several key pieces of support for the validity of our research design.

Using this approach, we examine two ways that someone may be influenced by nearby real estate investment activity: either an immediate neighbor has recently begun investing in the housing market, or a property in the immediate neighborhood was recently "flipped." As we further explain below, our aim is not to identify the specific mechanism(s) through which such contagion occurs via these channels per se (although we will have something to say about this when we analyze the choice of lenders), but obvious candidates abound, includ-

[7] One example is Duflo and Saez (2002) who have data from a large university and divide departments into subgroups along demographic lines to assess whether investor decisions, such as whether to enroll in a Tax Deferred Account or which mutual fund vendor to choose, is affected by other employees of the same department.

[8] See, for example, Bayer, Ross and Topa (2008), Linden and Rockoff (2008), Campbell, Giglio and Pathak (2011), Currie, Greenstone and Moretti (2011), Anenberg and Kung (2014), Currie et al. (2015).

[9] In our analysis we provide direct evidence in favor of this identifying assumption and also present a number of placebo tests and alternative specifications designed to test the robustness of the main results to different definitions of what constitutes "hyper-local."

ing word-of-mouth between neighbors related to information, optimism, or technical know-how about flipping homes and, perhaps more directly, a direct and vivid demonstration of the potentially large returns from short-term real estate investing in a booming market.

We apply our research design using data on nearly five million housing transactions in the Los Angeles metropolitan area from 1988-2012. To minimize any concern about the validity of our research design, we take a conservative approach and focus on investment activity within 0.10 miles of a household, roughly a city block. Our results imply sizable contagion effects for both new investors and flipped homes. The presence of each neighbor that begins to invest in housing within 0.10 miles of a household increases that household's probability of also investing in housing by eight percent within the next year and up to 14 percent over three years. The presence of a flipped property that has just been re-sold, the other channel of contagion that we consider, raises the probability of that household investing by nine percent and 16 percent over the same horizons. To strengthen the external validity of our results, we demonstrate consistent findings in both the San Francisco and Boston markets.

Moreover, the sizable effects that we document likely understate the true magnitude of investor contagion for at least three reasons. First, our research design identifies the effect of immediate neighbors and flipped homes relative to those just a short distance away. If those slightly more distant neighbors and homes also have an impact on investment activity, our estimates will understate the full extent of investment contagion. Second, our analysis considers the impact of all local investment activity regardless of whether a homeowner is aware of it. If, for example, homeowners only interact or learn about investment activity from half of their neighbors, the true impact of neighbors on one another would be twice as large. And, finally, our analysis, by design, only captures this neighborhood channel of investor contagion and, therefore, misses any impact that a homeowner's wider circle of family, friends, and acquaintances might have on investment behavior (e.g. Bailey et al. (2018)). Our analysis focuses on this neighborhood channel not necessarily because we believe that it is the most critical channel of contagion, but rather because it gives us some leverage to use a research design that credibly isolates causal effects.

Having established contagion in real estate investment activity, we provide an estimate of how much neighborhood investor contagion contributed to the level of speculative investing over the course of the Los Angeles housing boom. The contagion effect at a point in time varies as exposure to other investors and baseline entry rates fluctuate, but on average, we find that the conservative estimate of the neighborhood effect accounts for 10.3 percent of new investors over the course of the boom. The San Francisco and Boston markets had slightly less total investment activity than Los Angeles during their booms, although we also find similar marginal impacts of contagion in those markets as well.

We close the paper by exploring the performance of investors drawn into the market through this channel. In particular, we muster three pieces of evidence that investors subject to neighborhood influence at their time of entry (hereafter, "infected investors") perform worse than all other investors. First, infected investors earned inferior returns on properties they bought and sold, through three channels we decompose–buying at prices higher and selling at prices lower than other investors, relative to market, and suboptimal market timing. Moreover, we show that infected investors were more likely to hold their properties past the peak of the boom, and hence were more subject to capital losses and a failure to capture any market appreciation realized up to that point. Finally, we show that infected investors were more likely to default as prices plummeted. Because they purchased investment properties with lower initial equity stakes, however, their overall exposure to downside risk was somewhat limited relative to other investors. Overall, the results of our analysis suggest that investors infected by activity in their immediate neighborhood are substantially less sophisticated than the more general population of investors in the market.

Our paper is related and contributes to the empirical literature on peer effects in investment decisions. It has long been noted that in scenarios where agents lack perfect information about the potential costs and benefits of taking an action, like investing in the housing market, they may take informational cues from their peer group, even if this information is not correct (e.g., Bikhchandani, Hirshleifer and Welch (1992) or Ellison and Fudenberg (1993)). Of course, they may also take "social" cues from this group as they attempt to "keep up with the Joneses" by mimicking what others do (e.g. Bernheim (1994)). A number of papers have used these possible mechanisms related to social learning and social utility to motivate the empirical study of peer effects in investing (e.g., Duflo and Saez (2002), Hong, Kubik and Stein (2004), Brown et al. (2008), or Banerjee et al. (2013)). In connecting our paper to this literature one should be careful in the interpretation of the phrase "peer effect." In our paper we do not attempt to systematically measure one's peers; rather we are interested in the effect of nearby investment activity on one's investment behavior. This nearby activity might be generated from one's peers (perhaps a neighbor buys an investment property), but it need not be (as explained above, we also look at the effect of flipped houses in one's neighborhood regardless of where the flipper lives).

There has been recent interest in exploring peer effects in the housing market. McCartney and Shah (2016) use the same research design as we do and find that households are more likely to use a particular lender when more peers in their census block do. Gupta (forthcoming) uses exogenous shocks to interest rates in adjustable rate mortgages (ARMs) to identify spillover effects from one household defaulting on a mortgage to neighboring properties which are not directly affected by the ARM resetting.[10] Patacchini and Venanzoni (2014) look at the influence of a household's friendship network (proxied for by a measure of

[10]Other work investigating foreclosure contagion effects include Munroe and Wilse-Samson (2013) and Towe and Lawley (2013).

the household's children's friendships, which are surveyed in the U.S. National Longitudinal Survey of Adolescent Health) within a neighborhood to explore the impact peers have on housing quality, such as whether survey respondents report that their house is "well-kept." Like us, Bailey et al. (2018) look at peer effects in household purchase decisions, but focus on owner-occupier purchase decisions as opposed to the investment property decisions we examine. Also, they measure social networks/peers using Facebook, which, as we note below, highlights the conservatism in our approach to defining social networks based solely on geography. We contribute to this literature by providing what is to our knowledge the first application to real estate investing, as well as the first implementation of the type of identification strategy utilized in this paper to study investment decisions (although, as mentioned above, McCartney and Shah (2016) use the same design to look at mortgage refinancing decisions, some of which could be driven by investment motives). Additionally, we further extend this literature by exploring the performance of those investors drawn into the market by others' actions, as well as quantifying the effect of this contagion on the overall level of speculative investment that occurred during the recent housing boom and bust (we note that Bailey et al. (2018) find that individuals driven to buy homes because of the experiences of their Facebook friends tend to buy at higher prices). This is especially important for studying the consequences of peer effects in investing during "bubble-like" episodes, which is not the focus of the above-mentioned papers.

Reflecting on our findings, a natural question one might ask is how exactly the influence occurs. The empirical literature on peer effects in investing mentioned above usually cannot distinguish between the social learning and social utility mechanisms mentioned above.[11] While a precise delineation of these channels is not possible in our analysis, there are several reasons to suspect that social learning (of various kinds) plays an important role. First, the theoretical literature on speculative bubbles often points to some sort of learning about market fundamentals from others' behavior.[12] Second, there are so many obvious ways that learning could take place. For example, novice investors could change their beliefs about market fundamentals or the possible payoffs from investing, or they could learn practical "tricks of the trade" or from professional networks (of repair services, inspectors, attorneys, etc.) from more experienced investors. Third, we provide empirical support for one such learning mechanism at play in the data: lender referrals. Specifically, we find that new investors are significantly more likely to use the same lender as a nearby neighbor investor than an investor who lives only a bit further away.[13] The effect is strongest for atypical lenders, where,

[11]An exception is the recent experimental paper, Bursztyn et al. (2014), on peer effects in financial decisions, which finds that both social learning and social utility channels affect behavior.

[12]Indeed, Bursztyn et al. (2014) find that the social learning channel is strongest when agents observe the behavior of a relatively sophisticated group.

[13]Our analysis of lender choice is similar to that of Maturana and Nickerson (forthcoming) who use commonalities in teaching schedules to identify peers among school teachers in Texas and study the likelihood that a teacher refinances their mortgage when there is an increase in mortgage refinancing among their peers.

a priori, one would expect referrals to be most important. As a falsification test, we show that there is no effect of using the same lender that was used to flip a nearby property, where the actual investor may live much farther away, which would make the communication to refer a specific lender less likely.

The rest of the paper is organized as follows. Section I describes our source data and the construction of the estimation sample in detail. Section II describes the research design, gives evidence in support of its identifying assumptions and presents our estimates of investor contagion, closing with a simple and straightforward way to gauge the overall magnitude of contagion's effect on the housing market in general. Section III provides evidence that supports a social learning mechanism via the referral of lenders. Section IV compares the investment performance of influenced investors compared to all other investors in the housing market. Section V concludes. The Online Appendices contain a Monte Carlo study of our estimator, described below, and additional figures and tables referenced throughout the paper.

I. Data

In this section we introduce the data and describe in detail how we construct our sample of investment properties and investors (both potential and actual) in the housing market. See Bayer, Mangum and Roberts (2020) for more information.

A. *Overview of Sample*

The primary data set that we have assembled for our analysis is based on a large database of housing transactions compiled by Dataquick Information Services, a national real estate data company, which acquires data from public sources like local tax assessor offices. We have the complete record of property transactions for the greater Los Angeles, San Francisco, and Boston metropolitan areas from 1988 to 2012. For most of our analysis, we focus our attention on the largest and most volatile of these, Los Angeles.[14] For each transaction, the data contain the names of the buyer and seller, the transaction price, the address, and the date of sale. A county tax assessor file with property information is matched to the transactions through a unique property identifier. This provides numerous property characteristics including, for example, square footage, year built, number of bathrooms and bedrooms, lot size, and, importantly for our purposes, the latitude and longitude of the parcel.

From the original census of transactions, we drop observations if a property was subdivided or split into several smaller properties and re-sold, the price of the

[14]Los Angeles consists of the California counties of Los Angeles, Orange, Riverside, San Bernardino, and Ventura; San Francisco of Alameda, Contra Costa, Marin, San Mateo, and San Francisco counties of California; and Boston of Essex, Middlesex, Norfolk, Plymouth, and Suffolk counties of Massachusetts. Dataquick also provided data on New Hampshire counties bordering Massachusetts considered to be in the Boston MSA, but these are small in population and the data do not begin until the late 1990s, so we exclude them.

house was less than \$1[15] or flagged as not an arms-length transactions, the house sold more than once in a single day, or there is a potential inconsistency in the data such as the transaction year being earlier than the year the house was built. While these transactions are excluded from our analysis of investment activity, we still use the information on a transfer of ownership to mark when an individual's tenure of holding the property ends. We additionally drop properties whose price or square footage was in the top or bottom one percent of the sample as these are either highly unusual properties or simply reflect coding errors. Our research design will be based on a "neighbor" match of properties by distance, so we focus on single family homes because in condominium properties the "neighbor" proximity is less likely to depend on surface distance, although we explicitly check for robustness to this sample selection including related analysis of multifamily complexes. The remaining sample consists of nearly five million transactions in Los Angeles spanning 94 quarters. The San Francisco and Boston samples contain, respectively, 1.1 million and 584 thousand transactions.

Because the Dataquick data set contains information on liens registered against the property (i.e. mortgage information), we are also able to attach information on the income, race and ethnicity of homebuyers by matching Dataquick records to public data from the Home Mortgage Disclosure Act (HMDA) for a majority of transactions. As we show in Section 3 below, this demographic information allows us to examine household sorting at fine levels of geography, thereby providing a direct check on a key assumption underlying our research design.[16] We report summary statistics for the three cities in Online Appendices C and D.

As our paper focuses on the entry of new investors into a "bubble-like" market, we now describe the price dynamics of the Los Angeles housing market during our period of study. Figure 1 shows the basic dynamics of prices and transaction volume for the Los Angeles metropolitan area over the study period. (Similar depictions of San Francisco and Boston appear in the Online Appendix in Figure D1.) The price index is computed using a standard repeat sales approach.[17] Following a rapid increase in prices in the late 1980s, the early 1990s were a cold market period for Los Angeles, with prices declining by roughly 30 percent between 1992 and 1997 and transaction volume averaging only a little more than 30,000 houses per quarter during this period. Transaction prices averaged

[15] A nominal or zero price suggests that the seller did not put the house on the open market and instead transferred ownership to a family member or friend.

[16] HMDA requires mortgage companies to report information about every mortgage application and these data are made public on an annual basis. We merge HMDA data with Dataquick by matching on the basis of lender name, loan amount, Census tract, and year following the procedure described in Bayer et al. (2016). The merge results in a high quality match for approximately 75 percent of the sample. The merge fails in some instances due to the lack of a unique match (e.g., if two Bank of America loans for \$250,000 are registered in the same Census tract) and in others due to the use of alternate lender names in the two samples (e.g., for a subsidiary) that we were not able to verify as being part of the same company. Summary statistics for housing attributes for the matched subsample are very similar to the full sample as shown in Bayer et al. (2016).

[17] In particular, the index is based on the year-quarter fixed effects in a regression of the log transaction price on house fixed effects and year-quarter fixed effects.

\$187,000 in the 1990s. Starting in the late 1990s and continuing until mid 2006, the Los Angeles housing market experienced a major boom, with house prices tripling (prices averaging \$511,000 in 2006) and transaction volume nearly doubling. Just two years later almost all of the appreciation in house prices from the previous decade had evaporated and transaction volume had fallen to record low levels (less than 20,000 houses per quarter).

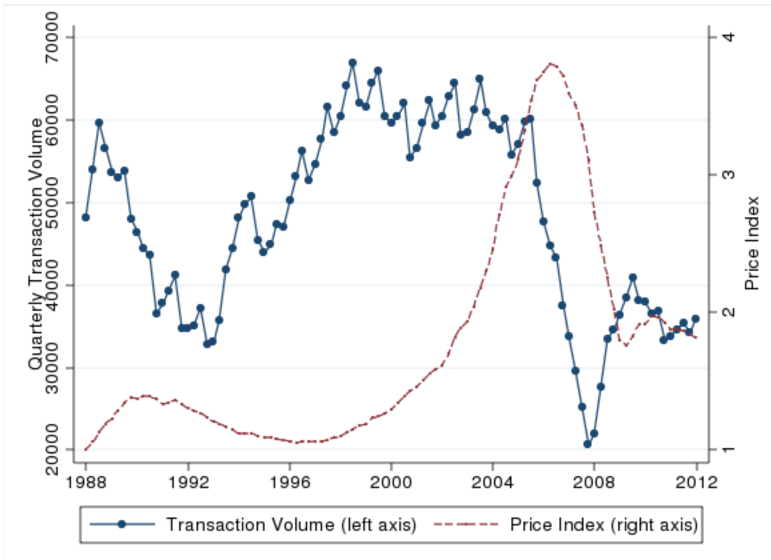

Figure 1. Quarterly Transaction Volume and Price Index in the Los Angeles Area.

Note: Uses transactions data as described in the main text for Los Angeles, Orange, Riverside, San Bernardino, and Ventura counties in California. Reported transaction volumes are smoothed by a three-quarter moving average. The price index is the quarterly dummy point estimate from a repeat sales regression of log transaction price; 1988Q1 is normalized to 1. See footnote 17.

B. Designating Investors and Measuring Investment Activity

Overview

The primary goal of this paper is to study investment activity in the housing market and, in particular, to examine whether homeowners are drawn into speculative investing by observing investment activity in their neighborhood. In this subsection we detail how we construct the key variables for our analysis.

Before proceeding in detail, we discuss our broad objectives for the data and the potentially complicating factors involved in achieving them. First, we want to use transactional data to identify whether a homeowner is an investor in the housing market, and if so, when he or she became one. There is no explicit designation of "investment" in the transactions data. Thus, we will need to infer investments for the individuals in the transactions record. Not every property that an individual owns is for investment purposes–some may be their primary residence (and they may have multiple primary residences during the sample if they move). While many individuals view ownership of their primary residence as "an investment," we are interested in those taking the next step in investing: buying additional properties for financial gain (whether through rental payment or capital appreciation), not regular consumption of housing services.

Our method, in short, is to build a profile of each individual's holdings over time by combining all of the property holding profiles associated with a given name, observing instances in which a property holding significantly overlaps with another, implying it is being held as investment and not for regular consumption of housing. Thus, for each person in the data we need a method for separating out, at each point in time, the primary residence from investment properties. Since the construction of many of these variables will rely on observing transactions and individuals over time, there is an obvious potential for sample selection bias at the beginning of the data (e.g., how could we know whether a person that is observed to purchase a home on the first day of our data is buying that home as an investment or a primary residence?). Thus, we will use the first half of the data as a "burn-in" period for constructing many of our variables. It is for this reason we focus on cities with a long transaction record in the Dataquick files.

Next, we need to identify the possible channels through which each non-investor may be influenced by neighborhood activity to become an investor. The "investment exposure" channels available in these data involve both *people* and *properties*: (i) whether a neighbor begins investing in the housing market ("people"), and (ii) if a nearby home is bought and sold as an investment ("properties"). Thus, we will use our designations of which individuals become investors (and when this happens), and which properties are transacted as investments (and when this happens) to measure a potential investor's exposure to each channel.

We elaborate on these and other data construction issues below. We also note that any possibilities for measurement error introduced by our sample creation will, in general, bias our analysis towards *not* finding any effect of nearby investment activity on a potential investor entering the market, thus implying that our estimates of causal effects are conservatively measured.

Designating Investment Transactions and Investor Entry

Our measures of investment activity are derived from buyer/seller names and transaction dates observed in the Dataquick data.[18] In the data, buyer and seller names are detailed and typically include middle initials and often the names of

[18]Our measure is most similar to Haughwout et al. (2011) , which uses credit records to designate overlapping holdings in mortgage records. Chinco and Mayer (2015) uses property tax addresses to flag "out-of-town" investors, a set of investor types obviously not of interest for our neighborhood effects study. Recent work on occupancy misreporting (Piskorski, Seru and Witkin (2015), Griffin and Maturana (2016), Mian and Sufi (2017)) suggests owner-occupied flags on mortgage applications are not a reliable way to measure investment activity, although we do conduct robustness checks on this.

a spouse or co-borrower. In our analysis we are interested in individual persons entering the investor market, and so we use the names associated with each transaction to exclude purchases by businesses, nonprofits, and various government organizations (collectively, "institutions"), although these might generally be considered investments of another type. We conduct an extensive name cleaning algorithm to flag institutions, to separate first, middle, spouse, and surnames, and to standardize punctuation and spacing.[19]

To provide a more detailed characterization of the data, Table 1 reports counts of observations for Los Angeles by identification of name (after cleaning). We will refer back to this table a number of times in this section. The full data set contains 3.8 million unique names conducting nearly 4.8 million transactions. Of these, 3.6 million names are "personal names" (e.g. George Akerlof) and not institutional names (e.g. First Bank of California). Among personal names, 57 percent include a middle name or initial, and three-quarters have a middle name or initial and/or a co-owner name (such as a spouse). We treat names with co-owners as distinct–"George Akerlof" and "George Akerlof and Janet Yellen" are two separate name profiles–which means that we likely understate the count of investors, since in some cases, the same Akerlof may be purchasing both properties.

Given our use of name-matching to build portfolios, some misclassification of ordinary home-owners as investors (and vice versa) is inevitable. In general, we expect such misclassification to diminish the measured differences between investors and regular homeowners by cross-contaminating the two categories, thereby attenuating our key parameter estimates. Below we illustrate the robustness of our results to various potential sorts of measurement error. The corresponding results indicate misclassification errors are not a serious problem for our main results, and in many cases suggest that such error leads to conservative estimates of the main effects presented throughout this paper.

Table 1 reports counts of individuals and properties flagged as investments. About one million of the transactions could possibly be categorized as investments, but a substantial portion (45 percent) are purchases by institutions, which we exclude. Another five percent are overlapping holdings for very common names (e.g. John Smith, Jose Lopez). In order to avoid classifying two individuals with a common name as one investor, we exclude any single name for which we observe more than 42 transactions.[20] Hence, we focus on personal names that are not excessively common.

After standardizing buyer names, we then leverage the transaction dates to construct a portfolio of properties that each individual holds over the course of our sample–an ownership "profile." We use the complete profile of property holdings over time for each individual to construct several important measures for our analysis, beginning with the designation of the set of individuals that

[19]The code for name cleaning algorithm is available upon request.

[20]Forty-two is a somewhat ad hoc number resulting from our inspection of the name distribution. Results are not sensitive to varying this cutoff.

Table 1—Purchaser Name Counts, Transactions Data 1988-2012.

Category Description	Transactions	Unique Buyers
All	4,756,715	3,520,786
Institutional Name	512,461	150,380
Personal Name	4,244,254	3,370,406
Excessively Common	50,052	694
Individual	4,194,202	3,369,712
Detailed	2,972,634	2,532,148
Middle Present	2,325,049	1,945,557
Spouse Present	1,575,864	1,420,694
Noninvestments	3,616,267	3,350,752
Investments (any)	577,935	262,429
Flips	102,202	50,761
Overlapping Tenures	543,159	253,763
Home at Entry ID'ed	356,491	175,306
No Home at Entry ID'ed	221,444	87,123
First property is flip	83,329	26,035
Multiple initial properties	101,379	36,798
First property held < 2 years	10,336	5,129
Insufficient overlap with investments	26,400	19,161

Note: The table shows counts of unique names and of transactions for several categories of names identified in the transaction register data for Los Angeles. A "common name" is a non-institutional name that is listed as buyer for over 42 distinct transactions.

become "investors" at some point during the sample period. We designate as investor an individual simultaneously holding two or more properties, with the entry date marked by the first occurrence of purchase of a second property (in overlap, i.e., without sale of the first). We allow for important exceptions related to the possibility that individuals may jointly hold two properties for a brief period in the course of moving homes within the metropolitan area. If the individual's original property is sold within six months of purchasing a second property, then the corresponding brief period of multiple holdings does not count towards the definition of an "investor." The overlap time is extended to twelve months if the second property would be the individual's only investment in our sample. Our goal here is to be conservative in characterizing multiple holdings as investments rather than the result of ordinary search frictions in the housing market.

Designating Investors At Risk of Entry

Having defined the set of individuals that become investors during the data period, we next distinguish each individual's "primary residence" from her investment properties. Identifying the primary residence is important because it determines the point in space around which we measure the homeowner's exposure to nearby investment activity. For individuals that never hold overlapping properties this is straightforward, as we simply classify their only property holding as their primary residence. For individuals that are classified as investors, we are primarily concerned with properly designating their primary residence during the period before they acquire multiple holdings. We treat the time from their first appearance in the sample until they simultaneously hold multiple properties (that is, the time of their investing entry, by our definition) identically to non-investors, designating their single property holding as their primary residence, with some important exceptions when we suspect that the initial property actually may have been itself an investment property.[21] To minimize instances of mis-designating what may actually be an investment property as a primary residence, we define an investor's primary residence as unassigned in her pre-entry period if any of the following conditions hold: (i) the individual was observed to purchase multiple properties within the first six months of entering the sample, (ii) the individual purchased multiple properties from the same seller in the same year, including one that would have been designated as the primary residence, (iii) the initial property that would have been designated as the primary residence was held for less than two years, or (iv) the time period of the individual's holding of multiple (i.e., investment) properties did not overlap with a non-investment property. In using these restrictions, our goal is to avoid classifying a primary residence in cases where the observed behavior looks suspiciously like that of an investor rather than regular homebuyer. In the last case, the concern is that we cannot reliably measure their exposure to investment activity prior to their time of entry. In general, measurement error should attenuate the entry effects we intend to estimate by contaminating the non-investor/investor designation on the outcome side of the regression.

The lower panel of Table 1 reports counts of individual buyers and transactions in investment properties. We can confidently identify the primary residence prior at the time of entry for two-thirds of buyers with personal (and not excessively common) names. We refer to an active primary residence (i.e. at a point in time between purchase and resale) as a "tenure."

We define all of the remaining overlapping property holdings of investors (i.e., those not designated as primary residences) as "investment properties." Among investment properties, we further designate whether a property was "flipped," i.e., sold again after a brief holding period, which we set to a period of less than two years.[22] This distinction is motivated by the possibility that this form of investment behavior may have been particularly visible to immediate neighbors, especially if the property was held empty during the investor's holding period. Importantly, our measure of flipped homes only counts these short-tenured sales of properties that have been classified as investment properties. This distinguishes flipped investment properties from short tenures by neighbors who may have had to re-sell quickly due to changes in life circumstances.

[21] We may observe investors by our overlapping property definition without ever observing a primary residence-like purchase. Two examples are someone who had purchased their primary residence in the Los Angeles area prior to the beginning of our sample in 1988, or an investor from outside the metro area. The latter case, "out-of-town" investors, may be important actors in the housing market (see Chinco and Mayer (2015), Favilukis and Van Nieuwerburgh (2017)), but are by definition not the focus of our neighborhood effects study.

[22] In related work, Bayer et al. (Forthcoming) found the two-year tenure definition to be a reasonable approximation of an investor's intent to "flip."

Figure 2 reports the time series for three measures of investment activity in the Los Angeles market between 1993-2012 derived from our transaction data set. We report counts of investment properties and then break out two subsets that are important for our analysis: those with an identified primary residence for the investor, and those that were flipped (i.e., re-sold within two years). The dynamics of the three series are very similar, with a clear peak around 2006. Afterwards, counts began to fall, although investments actually continued as a relatively high share of transactions.[23] Our overall measure of investment activity also tracks quite well that of Haughwout et al. (2011), which is based on a measure of multiple property holdings reflected on individual credit reports as multiple mortgage payments, providing further support that our name matching procedure is reasonable.[24]

In this figure one can see a "burn-in" period in how we construct our measures. As mentioned above, an obvious issue with our classification of primary residences and investment properties is that we must observe overlapping properties to flag investment properties, and the first purchase that we observe in the data for a given individual is assigned as the primary residence, subject to our disqualifications. In this way, our measure of investment activity is likely to understate the true level, especially near the beginning of the sample period.[25] After growing mechanically in the early years of the transaction data, the counts of active tenures stabilize around 2000, as shown in Online Appendix Figure C1, further supporting our focus on the period after 2000 when the construction of these measures of investment had stabilized.

Online Appendix Table C2 reports summary statistics for the set of transactions we identify as investments in the period 2000-2007 compared to non-investments. We report statistics for all investment properties in the period, and also the subsets of properties for which we identified the primary residences of the investors and those which are short-tenured "flips." The sample of investment property transactions is, in general, representative of the overall sample, though investment properties, especially flips, tend to be slightly older, smaller, and of slightly lower value. The subset of investment properties for which we have identified the

[23]While not the primary focus of this paper, it is interesting to note that investors remained quite active in the post-peak period, buying and holding properties rather than re-selling them quickly. In this way, investors may have helped to stabilize the market during the housing bust. See also Bayer et al. (Forthcoming).

[24]Haughwout et al. (2011) focuses on the early/mid 2000s, characterizing national trends and those in "bubble states," including California. In using a sample of credit reports, it does not face the "burn-in" issue we describe in using name matches, but shows dynamics similar to those we find.

[25]To get a sense of whether our measures of investment activity are reasonable in this latter portion of the sample period, we conducted a simple analysis of investment properties as a share of all quick sales (sales within two years of purchase) observed in the data. The overall count of quick-sales, of course, does not depend on our classification of investment properties. As expected, the ratio of these measures is initially quite low and rising over time, presumably as a greater percentage of investment properties are properly classified. Importantly, however, this ratio levels off by 1995 and, in fact, this ratio is exactly 34 percent for the periods 1995-2000 and 2001-2006, respectively - suggesting that the classification of investment properties is likely quite consistent from 1995 forward. Still, to be conservative, we begin most of our analysis in 2000.

location of the investor's primary residence is presented in the middle columns; the statistics are quite similar, indicating this is not an unusually selected sample of properties.

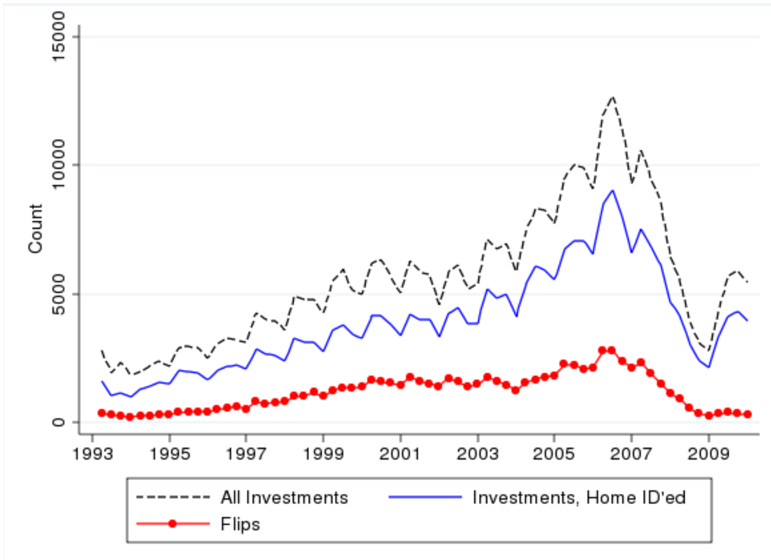

Figure 2. Investor Purchasing Activity Over Time.

Note: The figure plots three quarterly data series that measure investment activity for Los Angeles. Definitions are given in the main text. For investment properties, the timing is the quarter of purchase, and for flips it is the quarter of sale.

Figure 3 displays the time series of the hazard rate of investor entry behavior. For each individual homeowner with an active tenure, entry is defined as the date of the purchase of a first investment property. The hazard rate measures the fraction of active tenures who become new real estate investors in the given quarter. As in Figure 2, there is a strong upward trend in entry by new real estate investors over the course of the housing boom from 2000-2006, with a sharp falloff thereafter. Thus, the spike in volume in the boom shown in Figure 2 is due in part to the increased rate of entry of new investors.

Having characterized the tenures in primary residences of all non-investors and investors when possible, we define any homeowner observed in a primary residence as being "at-risk" of becoming a real estate investor, borrowing some language from epidemiology. Homeowners remain at-risk until they either (i) purchase an investment property (entering as defined above) or (ii) sell their primary residence and leave the sample.[26] Buying a new primary residence keeps the individual in

[26]As with the initial name-matching algorithm, we expect this procedure for defining homeowners "at-risk" of becoming first time real estate investors to introduce a small amount of misclassification

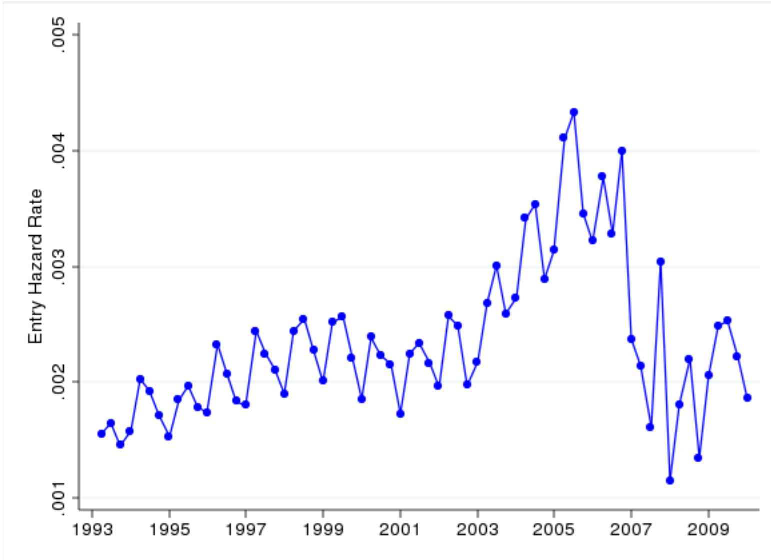

Figure 3. Investor Entry Over Time.

Note: The figure plots the hazard rate of quarterly investor entry - i.e. time of purchase of an individual's first investment property - for Los Angeles.

the sample as a new at-risk tenure.

Table 2 reports the at-risk tenures and entry rates for Los Angeles in our estimation period of 2000-2007 after excluding investor names for whom we could not confidently identify a primary residence. There are approximately 2.1 million unique at-risk tenures with an average of 50 months observed between 2000-2007 for a total of 104 million at-risk tenure months. We classify 63 thousand of these as investors. Since every active tenure in the metro area is considered at-risk and we observe them at a monthly rate, investor entry rates are relatively low by definition–three percent of individuals, entering at a monthly hazard rate of 0.06 percent.

Table 2 also reports the distribution of investors by their number of investment purchases. Most investors in our sample purchased only one or two investment properties, though the distribution has a long upper tail, with a handful transacting many–in some instances, dozens–of properties.[27] While this upper tail of

error into our main analysis, as some investment properties purchased by individuals residing outside the study area or in an existing home purchased prior to 1988 may be characterized as primary residences. We expect the number of such misclassified investment properties to be a small percentage of the overall stock of "at-risk" primary residences and for this error to attenuate the findings - in this case because the homeowner does not actually reside in the location that we have designated as their primary residence. We also examine robustness to more stringent definitions of owner occupancy.

[27]Recall that any name with more than 42 purchases is excluded to avoid excessively common names. Institutions are also excluded.

presumably more professional investors is interesting, it is important to note that our analysis is focused on the initial decision to enter as a real estate investor and, thus, is more centered on understanding the behavior of the novice investors comprising the majority of this distribution who entered during the boom (Figure 3).

Table 2—Estimation Sample Summary Statistics: Entry and Investing Frequency.

At-Risk Tenures and Entrants		Purchase Activity	
At-risk tenures (N)	2,114,687	Mean Investments per Investor	1.53
At-risk tenure-months (NT)	104,665,796	Pct. Purchasing 1	71.57
Entrants	62,947	Pct. Purchasing 2	17.23
Entry Rate (%)	2.98	Pct. Purchasing 3	5.70
Entry Hazard Rate (% x 10,000)	6.01	Pct. Purchasing 4+	5.49

Note: The table reports summary statistics of the primary estimation sample for Los Angeles; *entry hazard rate* is the outcome of interest. Additional statistics show the purchasing frequency of investors conditional on entry. Other definitions given in the main text.

Exposure to Neighborhood Investment Activity

Having characterized transactions as investments and flips, and individuals as investors or non-investors, we next conduct an analysis of the spatial patterns of exposure to investment activity and entry. With interest in studying the entry behavior of new investors, we perform a spatial match of all at-risk investors (those who eventually enter and those who do not) to their exposure to the two forms of investing activity, people and properties. We leverage the location information and transaction dates available in the data to build a panel dataset of at-risk individuals observed over the course of their tenure in their primary residence.

Specifically, we use the location of the primary residence of at-risk homeowners to construct two measures of exposure to nearby investment activity, measured by the Great Circle distance calculated using the properties' latitude and longitude information from the tax assessor file. The first is the measure of people: whether the individual's neighbors are engaged in real estate investment. We construct counts of the number of neighbors who have entered as investors over several lags of time, constructing the exposure measure for various distance bands around the individual's primary residence. Our second measure of neighborhood investment activity is based on investment properties. We construct counts of the number of investment properties that were sold following a holding period of less than two years ("flips"), again at various lags and distance bands.

Thus, using each at-risk individual's location as a reference point in space, we measure the exposure to investors (i.e., people, wherever they may buy investment properties) and to nearby investments (i.e. properties, wherever their owners may reside). These two measures of exposure to investment activity will form the basis of our research design, functioning as both treatment and control variables depending on the exposure radius, as we explain in Section II. The expected impact of misclassification error in designating these investments would be to attenuate estimated effects by classical measurement error. Note that each of these measures vary over both space (individuals in some neighborhoods observe more investment activity than others) and time (some periods have more investment activity than others).

To each at-risk tenure month, we have matched the two forms of investment activity occurring within one-tenth mile distance rings and in annual lags up to four years. In the lower panel of Table 3, we report the summary statistics for exposure within a one year lag for the 0.1, 0.3, and 0.5 mile rings–the treatments and controls in our primary estimation specification. (These statistics are taken over the pooled sample, so they average over spatial and temporal variation.) The level of exposure rises mechanically with the expanding radius, and the outer rings are inclusive of the inner ring exposure, so that any exposure within 0.1 mile, for example, is also within 0.3 and 0.5 miles. At the innermost ring, much of the variation occurs at the extensive margin–zero exposure versus non-zero. Most at-risk tenures have at least some exposure to investment in their broader neighborhood of the 0.5 miles radius.

Table 3—Estimation Sample Summary Statistics: Exposure to Investing Activity.

Distance (miles)	Mean	Std. Dev.	Pct. w/. 0	Pct. w/. 1	Pct. w/. 2+
Investor Neighbors within:					
0.1	0.28	0.62	78.76	16.82	4.42
0.2	0.81	1.14	51.75	28.89	19.36
0.3	1.55	1.69	31.06	28.60	40.33
0.4	2.49	2.32	18.15	22.18	59.66
0.5	3.60	3.02	10.88	15.32	73.80
Flips within:					
0.1	0.18	0.52	85.46	11.86	2.67
0.2	0.55	0.96	64.91	23.27	11.82
0.3	1.05	1.46	46.51	27.55	25.94
0.4	1.70	2.06	32.49	26.34	41.16
0.5	2.47	2.71	22.66	22.53	54.81

Note: The table reports summary statistics for investment exposures, the explanatory variable of interest, in the primary estimation sample for Los Angeles. Definitions given in the main text. Spatial rings of exposure are inclusive of the narrower rings (e.g. 0.1 mile is also within 0.3 mile).

II. Estimating Contagion in Real Estate Investing

A. *Research Design*

The primary goal of our analysis is to identify the causal impact of neighborhood investment activity on the entry of homeowners into speculative real estate investing. Specifically, we seek to examine whether at-risk homeowners are affected by the recent entry of their neighbors into real estate investing and/or by observing homes being bought and quickly re-sold in their neighborhood.

The main challenges with identifying contagion along these lines is that (i) homeowners are not randomly assigned to neighborhoods and (ii) unobserved factors operating at the neighborhood level may affect the investment activity of all residents. Both of these issues might give rise to correlation in investment activity at the neighborhood level that is not causal.

To deal with these issues, we follow a research design that has been used extensively in the recent literature on neighborhood effects and local spillovers. The idea is to examine the influence of hyper-local investment activity (e.g., on one's own block) while controlling for comparable activity on other nearby blocks. In practice, we will measure the effect of activity within a radius of one-tenth of a mile, while conditioning on activity in wider (e.g. 0.30 or 0.50 mile) bands.[28] Additional nonparametric controls for time period and broader geographic definitions can be included as well.

Formally, there are two key identifying assumptions that underlie this approach. The first is that household sorting or other unobserved predictors of investment activity do not vary in a significant way across this geographic scale, due, for example, to the fact that search frictions may limit a household's ability to select the exact block on which they live. That is, while a homebuyer may be able to identify a neighborhood in which they would like to live, their ability to pick an exact block is limited by the homes listed for sale when they are searching. The second identifying assumption–or more precisely, a necessary condition for detecting nonzero effects–is that neighborhood interactions are more likely to take place at hyperlocal geographies. The effect that we estimate, for example, will contrast the response of homeowners to activity within 0.10 miles with that just a bit farther away. If neighborhood interactions were not stronger at hyperlocal geographies, the estimated effect would be zero. Of course economic theory does not define the scale at which such interactions truly take place. Our selection of radii are intended to be a comparison of a city block's distance to a set of blocks. To the extent that interactions also operate (perhaps at a lower intensity) at a broader geographic scale, our analysis will tend to understate the full size and scope of these neighborhood interactions.

It is worth noting at the outset that these identifying assumptions are unlikely to hold perfectly in the real world. We expect that there will be some block level sorting, however limited, and neighborhood spillovers to decay with distance in a more continuous way, even if they are much stronger among immediate neighbors. Each of our exposure measures is continuous, so intuitively, this research design of inner ring/outer disc draws on the sharp contrast between the two spatial

[28]One way to see the intuition of the design is to recall most of the inner ring variation is at the extensive margin (zero or nonzero). The regressions will control for neighborhood level investment activity, and randomly some at-risk tenures get exposure very close by. The Monte Carlo exercise in Online Appendix A shows the statistical properties of random local exposure in comparison to our estimated effects.

processes at play. That is, we expect the correlation between the unobserved *attributes* of neighbors that might affect their propensity to become real estate investors to be only slightly stronger on the same versus nearby blocks, and we are interested in whether *investment activity* reveals a pattern of much stronger correlation at very close geographic proximity.

Such is the general intuition of nearest-neighbor research designs for measuring neighborhood effects. In our current context, there may additionally be concerns that dynamically-varying *information* about real estate investing is decaying rapidly and that even fine spatial scales could still be subject to common information shocks not available to the broader neighborhood, so we will also test and control for measures of such localized information in the real estate market.

The following figures illustrate the basic idea of this research design using data from our sample. First, to examine how the degree of household sorting increases with geographic proximity, we compare a homeowner's attribute x_i with the mean of those attributes within successive annuli (i.e., open rings, or "doughnuts") of width d drawn around the homeowner's residence, $\bar{x}_{i,rd,(r+1)d}$, for $(r = 1, 2, \dots)$. For example, if X is income, $\bar{x}_{i,(2)0.1,(2+1)0.1}$ is the mean income in an annulus whose inner radius is 0.2 miles and outer radius is 0.3 miles. We then take the average absolute value of the differences between x_i and $\bar{x}_{i,rd,(r+1)d}$ over the sample for 20 $(r = 1, \dots, 20)$ bins of 0.1 mile rings $(d = 0.1$ miles). Figure 4 reports the proportional differences between a homeowner's and neighbors' attributes as a function of the distance between the homeowner and the neighbors. The figure reveals that for race/ethnicity and household income at the time of purchase, this difference increases quite gradually with distance. That is, homeowners are only slightly less similar, in terms of race/ethnicity and income, to their neighbors 0.1-0.2 miles away relative to neighbors within 0.1 miles, and again slightly less to those 0.2-0.3 miles away, and so on.

Beyond typical sorting on attributes, there may be spatial similarities in price dynamics that provide information about the real estate market to local residents, so we conduct the same spatial gradient exercise for price growth.[29] We run a standard repeat-sales regression by county to sweep out the common trend in the market, and then take residuals from this regression for each property transaction.[30] From these residuals, we can form a local index of price dynamics that may be relatively more informative for nearby properties. In Figure 4, we report the spatial similarity in these price residuals, taken as the difference between the focal properties residual and those in 0.1 mile annuli, just as we did with homeowner attributes. We find there is very little difference in the price growth residual across the spatial scales.

In contrast to the smooth spatial patterns in Figure 4, Figure 5 illustrates the relatively sharp increase in very local neighborhood investment activity for

[29]We are grateful to two anonymous referees for suggesting this analysis.

[30]The residual is the observed price minus the predicted price, $r_{it} = p_{it} - \hat{p}_{it}$, where $\hat{p}_{it} = \hat{\alpha}_i + \hat{\delta}_i$, with $\hat{\alpha}_i$ the property fixed effect and $\hat{\delta}_i$ the quarter fixed effect from the county-level repeat sale regression.

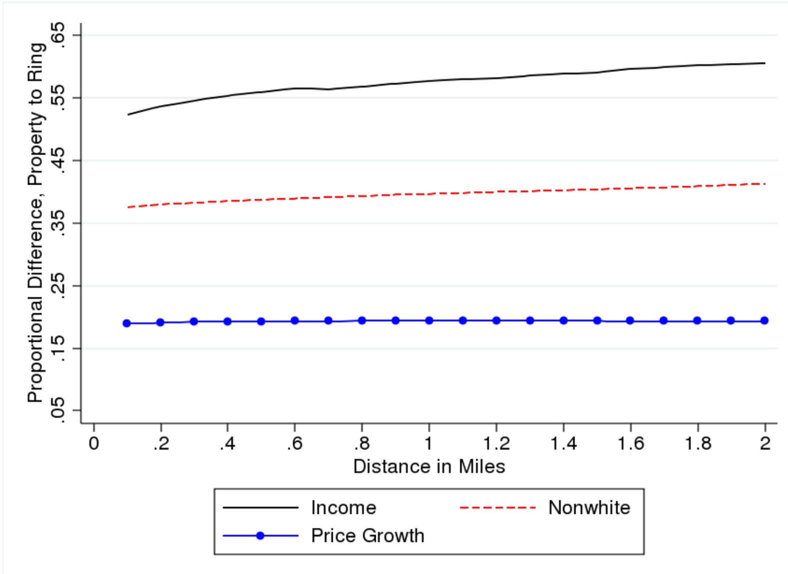

Figure 4. Comparability in Homeowner Attributes Over Space.

Note: The figure reports the average proportional difference between the attribute of a single property and the average of that attribute for property transactions within rings (annuli) of d tenths of a mile.

at-risk homeowners who entered as real estate investors in a given period. The figure plots the difference in neighborhood investment activity within the past year for at-risk homeowners who entered as real estate investors versus those who did not become investors, again calculating these measures for the twenty 0.1 mile-wide rings from 0.1-2.0 miles away. The graphs correspond to our two main measures of neighborhood investment activity, entry into real estate investment by immediate neighbors and a "flipped" property in the neighborhood. Figure 5 reveals a difference at 2 miles, and a pattern of a slightly increasing differences as the geographic scale closes from 2.0 to 0.5 miles that closely resembles the pattern in Figure 4.

Strikingly, as the geographic scale shrinks even further, the differences in exposure to both measures of recent neighborhood investment activity rise sharply. In this way, Figure 5 implies that those at-risk homeowners who became first-time real estate investors in a given period were much more likely to have been exposed to neighborhood investment activity at very close proximity to their homes than those at-risk homeowners who did not become investors. Importantly, by controlling directly for investment activity within 0.30 or 0.50 miles, all of the results that follow isolate as the causal effect only the sharp increase in entry associated with neighborhood activity at a very fine geographic scale (0.10 miles). We conducted an additional Monte Carlo-style placebo test of our inner/outer ring research design that further supports this approach as a way to identify causal effects of

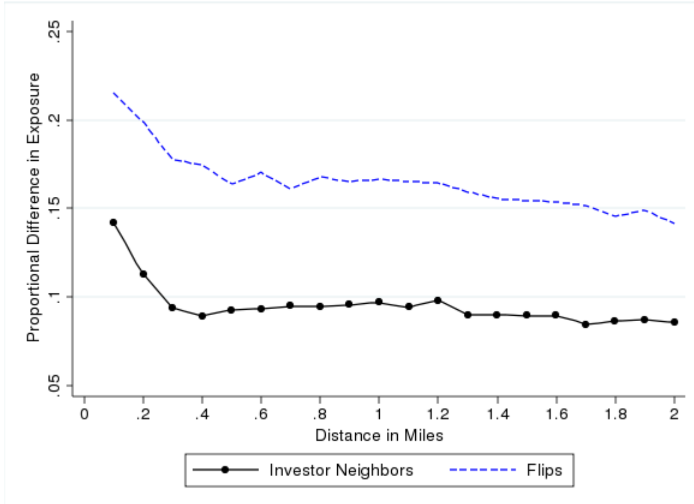

Figure 5. Differences Between Investor Entrants and Non-Entrants in Exposure to Investing Activity.

Note: The figure reports the average proportional difference in exposure to investment activity within the last year between identified investors and non-investors. The non-investor comparison group is drawn from a random sample (with replacement) of non-investor tenures active in the investor's entry month.

neighborhood activity on an individual's likelihood of becoming an investor. A description and the results of this test are provided in Online Appendix A. The effects we find in the analysis below are well outside the distribution effect sizes if investment activity were randomly assigned to be within a nearby ring, suggesting that some special influence is indeed occurring at very close proximities.

While the contrasting spatial patterns of Figures 4 and 5 are visually apparent, it is also clear that the differences in attributes between very near and slighter more distant neighbors are not literally zero. (Obviously there must be some differences, on average, at small spatial scales in order to generate differences at larger spatial scales.) This leaves open the possibility that even these smooth differences are consequential in predicting the spatial pattern of investing. An assumption of this research design is that individuals with similar entry propensity do not cluster together within very fine geographies. We believe this to be reasonable, since the thinness of the housing market often precludes such small scale selection, but we readily acknowledge that the assumption may not hold exactly. At issue for our research design is whether the hyperlocal sorting that does occur is *predictive of investing in a meaningful way.* We examine the issue statistically in the following exercise.

We want to test for excess correlation between an individual's likelihood of investing and that of nearest neighbors relative to those neighbors living slightly farther away. We first measure the likelihood that an individual becomes an investor using the observable attributes in the data (income and race) in a simple regression of entry on these attributes, $e_i = \gamma' X_i + \varepsilon$, where X is the vector of attributes; in practice, we use categorical variables for racial group, income tercile, and an indicator for whether the income and race information is missing because of failure to match to HMDA data. (We will account for the dynamically varying local real estate market attributes directly in our main outcome regressions below.) This model yields an expected likelihood that an individual i would be an investor, $\hat{e}_i = \hat{\gamma}' X_i$. The purpose of this regression is to build an index of entry likelihood based on observable attributes X, not to make any causal claims about these attributes.[31]

We then form a sample of at-risk homeowners matched to neighboring at-risk homeowners up to a distance of 0.5 mile and measure the spatial correlation in the entry index in a regression.[32] As in our primary analyses below, we capture correlation within the 0.5 matched ring, and then measure excess correlation as the ring narrows to 0.3 and 0.1 miles. The regression is:[33]

$$\hat{e}_i = \beta_0 + \beta_{0.1} \hat{e}_j I(Dist(|i-j|) \leq 0.1) + \beta_{0.3} \hat{e}_j I(Dist(|i-j|) \leq 0.3) + \beta_{0.5} \hat{e}_j I(Dist(|i-j|) \leq 0.5) + \varepsilon_i \tag{1}$$

where i and j are two distinct at-risk tenures, $Dist$ means the distance in miles between the properties, and $I(\cdot)$ is an indicator function. This test adds a few insights compared to a simple observational analysis of Figure 5. First, it combines the attributes into an index rather than relying on univariate correlation alone. Second, it accounts for the potential predictive power of each attribute, which, as noted above, was not captured by looking at the gap in attribute means alone, and allows use of nonlinear models of prediction such as logit. As a result, the test allows for a quantification of the potential bias from excess sorting.

Table 4 reports the estimated coefficients. There is clearly sorting on entry likelihood at the widest ring of 0.5 miles, as indicated by the coefficient estimate of 0.17. However, there is no additional sorting on entry propensity at the inner rings as the estimated coefficients inside are smaller than the outer ring by multiple orders of magnitude (the estimates are slightly negative but the magnitudes are very small). The conclusion is that sorting at close proximity is not more predictive of entry once conditioning on the broader neighborhood.

This analysis provides direct support for the nearest-neighbor research design in our context. To augment the nearest neighbor design, we also take the additional step of conditioning directly on observable individual and neighborhood attributes in the analysis below - finding, as expected, that these additional controls have little impact on the estimated effects.

[31] Indeed, if these attributes are correlated with predictive unobservables, the test of the research design becomes even stronger.

[32] Because of the massive adjacency matrix created by a complete matching of all properties to all properties, we randomly sample a maximum of 30 within the half-mile buffer. Further, the at-risk tenures are limited to those that begin within five years of one another.

[33] Our notation here is different from our main entry regressions because this test uses a paired sample–at-risk tenures paired with other at-risk tenures–not a summary of local exposures around a focal tenure.

TABLE 4—REGRESSION OF INDIVIDUAL EXPECTED ENTRY PROBABILITY ON LOCAL NEIGHBORHOOD EXPECTED ENTRY.

Summary of Entry Propensity Index		
Mean	5.2446	5.2438
Std. Dev.	1.3228	1.3254
N At Risk Tenures	1,297,383	1,297,383
Matched Homeowner Sample Regression		
Entry Model	LP	Logistic
Index X w/i. 0.10 mi	-0.2758	-0.2808
	(0.024)	(0.024)
Index X w/i. 0.30 mi	-0.0928	-0.0911
	(0.015)	(0.015)
Index X w/i. 0.50 mi	17.1712	17.2241
	(0.054)	(0.054)
Cons.	4.3129	4.3074
	(0.003)	(0.003)
N Matches	36,466,363	36,466,363

Note: The table reports the coefficients and standard errors (in parentheses) from a regression of individual entry probabilities in matched owner-occupant pairs in distances up to 0.5 miles. The entry propensity model in column 1 is linear probability (LP) and in column 2 is a logistic regression (Logistic). The matched-pair regression is linear in the reported distance increments.

B. Baseline Results

We now present our main results based on estimating regression specifications that relate entry as a real estate investor to measures of recent neighborhood investment activity, exploiting the inner ring/outer disc research design. In the panel dataset we construct, the level of observation is the monthly at-risk tenure, which is defined as any active tenure in which the homeowner has not yet engaged in real estate investing. The primary specification is a linear probability hazard regression that relates investor entry to recent neighborhood investing activity. An investor remains in the data until they enter as investors, exit their tenure, or the study period ends. We focus on the period 2000-2007, which includes the period of house price appreciation and overall increase in investing activity for the entire greater Los Angeles area. The basic estimating equation for our research design is

$$e_{it} = \beta_0 + \sum_{r}^{R} \left[\beta_r^N N_{t-s:t,0:rd} + \beta_r^F F_{t-s:t,0:rd}\right] + \varepsilon_{it} \tag{2}$$

where outcome e_{it} is an indicator for whether the at-risk individual i enters as an investor in month t. Our main explanatory variables of interest are the exposure to neighbors entering as investors, N, and properties being flipped, F. As described above, each of these is measured as the cumulative count in a recent time period,

$t-s:t$, and within some distance ring, $0:rd$. We start with a one year lag, so that $s = 11$ (months), but will expand to study the profile over preceding lags.[34] As the above section showed, the distance rings are of primary concern. Our research design is to measure the effect of exposures at inner ring distance, $r = 1$, controlling for exposures at the wider disc. Note that each ring starts at 0, so that the outer discs are inclusive of the inner ring exposures. Thus, the coefficients of (β_1^N, β_1^F) measure the additional effect of the exposure occurring within the nearest ring over and above the effect of exposures occurring in 0 to Rd. As before, we use distance increments of $d = 0.1$ miles. Our ring steps are $r \in \{1, 3, 5\}$. In this hazard specification, there is both spatial and temporal variation in the level of housing activity. Intuitively, we can identify an effect by comparing two at-risk tenures, one with neighbor investors and one without, or by comparing the propensity of an at-risk individual to enter when there has been recent investing activity to a period when there was not.

Table 5 reports our main results from estimating Equation (2). We include each type of treatment, nearby neighbor investors and nearby flipped properties, jointly in the same regression.[35] Coefficient estimates are followed by hazard ratios, the change in propensity to enter attributable to one exposure of the explanatory variables relative to a baseline unexposed at-risk tenure. The results presented in column 1 show that there is a positive and significant effect of activity within 0.10 miles on the propensity to enter as a real estate investor in a given month. The results are easiest to interpret when expressed relative to the unexposed hazard. The coefficients imply an increased hazard of 15 percent when a neighbor becomes an investor, and a 22 percent increase from having a flipped property in one's immediate neighborhood.

Column 2 utilizes our inner ring/outer disc research design, adding controls for activity within 0.3 miles and 0.5 miles. Controlling for these broader neighborhood measures reduces the estimated impact of each measure of exposure by about half. Measured as a percentage increase in the baseline hazard, each investor neighbor in one's immediate neighborhood increases the propensity to enter in a given month by 8 percent, while each property flip increases it 9.3 percent. This is the effect of the exposure within 0.1 miles in excess of its being within 0.3 and 0.5 miles. As discussed above, we take the innermost ring to be a conservative estimate of the causal impact of one additional investor neighbor or flipped property on investor entry.[36]

[34]The count ignores investments that occurred before the at-risk tenure was active (e.g 11 months ago when the tenure was active for only 10 months) since these would not actually be observed by the at-risk homeowner. We include dummies for early in the at-risk tenure to account for the associated censoring. We also exclude the first six months of the at-risk tenure from the estimation, since an overlap of less than six months is by our definition not an investment.

[35]Results are qualitatively similar, but quantitatively larger, when using separate specifications to measure the effects of each type of exposure.

[36]For comparison, the effect of a investor neighbor exposure within 0.3 miles, if we were to include it as part of causal effects, adds an additional 1.2 percent in excess of its being within 0.5 miles. The wider marginal effect is smaller than the same-block effect, but it still would add appreciably to the total effect in Table 11 because exposure rates are higher in the wider rings.

We can also include flexible controls for unobservables by time periods or by discrete neighborhood definitions such as ZIP codes. In Equation (3), the terms D_t and D_z represent, respectively, dummy variables for quarter of observation and ZIP code.[37] The outer disc terms already control for neighborhood trends over time and space, centered around the at-risk individual in space and around the period of observation in time, but D_t would account for common shocks to the market as a whole and D_z for common unobserved features of discrete neighborhoods. The most exhaustive nonparametric control is to interact these and enter ZIP-by-quarter dummies.

$$e_{it} = \sum_{R}^{r} \left[\beta_r^N N_{i,t-s:t,0:rd} + \beta_r^F F_{i,t-s:t,0:rd}\right] + D_t + D_z + \varepsilon_{it} \tag{3}$$

The remaining specifications in Table 5 further test the research design with versions of Equation (3). Columns 3 and 4, respectively, add year-quarter and ZIP code fixed effects, and Column 5 includes both.[38] The coefficients and hazard ratios are very stable, indicating that the outer ring research design has effectively controlled for both spatial and temporal trends in exposure. It is interesting to note that the 0.50 miles coefficients fall with the inclusion of both the ZIP code and quarterly dummies specification in column 5, while the 0.10 miles points remain stable. This suggests that the 0.50 miles disc functions sufficiently as a control for neighborhood trends. As a more stringent control, column 6 uses the interacted ZIP-by-quarter effects and again the inner ring coefficients are unchanged. We adopt the specification reported in column 2, with 0.30 and 0.50 disc controls, as our preferred specification, since it produces very similar results as the others at much lower computational cost.

All of these results clustered standard errors at the at-risk tenure level to allow for correlation in the errors between observations for the same individual, but there is potential for wider spatial and temporal correlation. More stringent forms of clustering that account for additional spatial and temporal correlation in the error structure, including double clustering as suggested by Petersen (2009), are provided in the Online Appendix in Table B1. None of the conclusions is affected by these alternative clusterings.

C. *Controls for Individual and Neighborhood Attributes*

A potential concern with this research design is that the outer discs do not sufficiently control for all the determinants of entry propensity, and if the omitted determinants also vary at fine spatial scales or with nearby exposures, the

[37]The results are similar when using census tracts instead of ZIP codes. See Table B2 in Online Appendix B.

[38]We use the baseline hazard rate from column 2 so that hazard ratios are interpretable across specifications with and without fixed effects.

Table 5—Linear Probability Hazard Regression of Investor Entry on Exposure to Investing Activity: Baseline Model.

	1	2	3	4	5	6
			Regression Coefficients			
Investor Neighbor						
w/i. 0.1 mi	0.8190	0.3228	0.3269	0.3462	0.3479	0.3260
	(0.0462)	(0.0524)	(0.0524)	(0.0526)	(0.0526)	(0.0526)
w/i. 0.3 mi		0.0495	0.0527	0.0659	0.0712	0.0659
		(0.0268)	(0.0268)	(0.0269)	(0.0269)	(0.0269)
w/i. 0.5 mi		0.1685	0.1257	0.1479	0.0792	-0.0666
		(0.0148)	(0.0149)	(0.0154)	(0.0155)	(0.0161)
Flip						
w/i. 0.1 mi	1.1888	0.3769	0.3763	0.4110	0.4061	0.4069
	(0.0574)	(0.0652)	(0.0651)	(0.0655)	(0.0655)	(0.0657)
w/i. 0.3 mi		0.0590	0.0580	0.0845	0.0841	0.0877
		(0.0330)	(0.0330)	(0.0332)	(0.0332)	(0.0333)
w/i. 0.5 mi		0.3159	0.2707	0.2033	0.1213	0.1505
		(0.0178)	(0.0178)	(0.0188)	(0.0189)	(0.0199)
Constant	5.3251	4.0720	4.3344	4.3640	4.8112	5.2812
	(0.0271)	(0.0389)	(0.0399)	(0.0461)	(0.2348)	(0.0515)
			Marginal Effect to Hazard Rate			
Investor Neighbor						
w/i. 0.1 mi	0.1538	0.0793	0.0803	0.0850	0.0854	0.0803
	(0.0089)	(0.0129)	(0.0129)	(0.0129)	(0.0129)	(0.0129)
Flip						
w/i. 0.1 mi	0.2233	0.0926	0.0924	0.1009	0.0997	0.0924
	(0.0111)	(0.0160)	(0.0160)	(0.0161)	(0.0161)	(0.0160)
Qtr Effects			yes		yes	
ZIP Effects				yes	yes	
ZIP X Qtr Effects						yes

Note: The outcome is whether the at-risk tenure enters as an investor in a given month. Standard errors in parentheses are clustered at the at-risk tenure level. Coefficients have been multiplied by 10,000 for readability.

inner rings could pick up spurious effects. Our preliminary analysis in Section II.A largely assuages this concern, but we can test the sensitivity of our results to inclusion of individual and neighborhood attributes. The latter are especially interesting, because beyond helping to address concerns related to omitted observables, local market indicators could comprise other sources of information relevant to the individual's entry propensity.

Equation (4) represents an expansion of the baseline (2) with the inclusion of additional controls. The time-invariant entry propensity index from Section II.A, $\hat{e}_i$, measured from attributes of the at-risk individuals (income and race) is included directly. Next, a time-varying individual attribute, one's contemporaneous level of equity in their primary residence, appears as w_{it}. We measure equity from the projected current home value, using county-level appreciation at time t relative to the time of purchase, minus the expected mortgage balance computed from the loan's vintage and initial loan-to-value ratio taken from the transaction record.[39]

Other measures of activity in the local real estate market around at-risk individual i appear in the vector $A_{i,t-s:t,0:rd}$. We incorporate measures of the local area price appreciation relative to a county-wide index, as described in II.A, and similarly, the level of transactions activity, measured as a proportion of the total stock.[40] These may provide local, salient signals to at-risk individuals about the profitability of investment entry which could be potential confounds of our main effect, and are independently interesting regardless. We denote these in distance rings $0 : rd$ of at risk tenure i in lags of $t - s : t$, just as we did with the investor exposures.

$$e_{it} = \beta_0 + \sum_R^r \left[\beta_r^N N_{i,t-s:t,0:rd} + \beta_r^F F_{i,t-s:t,0:rd}\right] + \gamma_1 \hat{e}_i + \gamma_2 w_{it} + \sum_R^r \gamma_r^A A_{i,t-s:t,0:rd} + \varepsilon_{it} \quad (4)$$

Table 6 adds more controls for the individual at-risk tenure and neighborhood. The first column, 0, is our baseline from Table 5 (column 2 of that table). Column 1 adds the individual-level controls only. The entry index is itself highly predictive of investment activity–the lower panel shows a standard deviation increase in the index increases entry likelihood by 25 percent–but as expected, this does not change the estimate of investment exposure effects. The estimated effect of the equity measure is negative, since in the cross section individuals with very high home equity are less likely to become real estate investors. In any case, its inclusion does not affect the main results.

Column 2 uses controls for other neighborhood activity in the real estate market. These may be important as a source of information to the individual about

[39]The results are robust to ignoring the loan information and simply using current value less purchase price.

[40]This proportion is $\frac{\sum_{0:rd} T_{i,t-s:t,0:rd}}{\sum_{0:rd} P_{i,0:rd}}$, the sum of transactions with a distance rd of property i in the last $t - s$ months, divided by the number of properties within the distance rd.

Table 6—Linear Probability Hazard Regression of Investor Entry on Exposure to Investing Activity: Including Individual and Neighborhood Controls.

		0	1	2	3
Regression Coefficients					
Investor Exposures					
Investor Neighbor	w/i. 0.1 mi	0.3228	0.3119	0.3040	0.2977
		(0.0524)	(0.0528)	(0.0525)	(0.0527)
	w/i. 0.3 mi	0.0495	0.0441	0.0472	0.0454
		(0.0268)	(0.0268)	(0.0268)	(0.0268)
	w/i. 0.5 mi	0.1685	0.1728	0.1542	0.1632
		(0.0148)	(0.0168)	(0.0149)	(0.0176)
Flip	w/i. 0.1 mi	0.3769	0.3650	0.3328	0.3274
		(0.0652)	(0.0655)	(0.0653)	(0.0655)
	w/i. 0.3 mi	0.0590	0.0529	0.0440	0.0445
		(0.0330)	(0.0331)	(0.0332)	(0.0332)
	w/i. 0.5 mi	0.3159	0.2982	0.2745	0.2737
		(0.0178)	(0.0179)	(0.0180)	(0.0197)
Additional Conrols					
Personal Attributes	Entry Likelihood Index		81.0124		78.5396
			(2.0573)		(2.1790)
	Current Home Equity		-0.6970		-0.6451
			(0.3624)		(0.3449)
Local Price Change	w/i. 0.1 mi			0.9429	0.8452
				(0.1588)	(0.1616)
	w/i. 0.3 mi			1.5887	1.3023
				(0.3556)	(0.3643)
	w/i. 0.5 mi			3.9490	3.1459
				(0.3981)	(0.4791)
Local Transaction Rate	w/i. 0.1 mi			2.9806	2.6461
				(0.4881)	(0.4956)
	w/i. 0.3 mi			0.6701	0.3008
				(1.3727)	(1.3759)
	w/i. 0.5 mi			2.3607	-1.5768
				(1.5034)	(2.0989)
Marginal Effect to Hazard Rate					
Investor Neighbor	w/i. 0.1 mi	0.0793	0.0766	0.0746	0.0731
		(0.0129)	(0.0130)	(0.0129)	(0.0129)
Flip	w/i. 0.1 mi	0.0926	0.0896	0.0817	0.0804
		(0.0160)	(0.0161)	(0.0160)	(0.0161)
Personal Attributes	+1 S.D., Entry Llhd. Index		0.2523		0.2446
			(0.0067)		(0.0070)
	+1 S.D., Current Home Eq.		-0.1948		-0.1803
			(0.1013)		(0.0964)
Local Price Change	+1 S.D., w/i. 0.1 mi			0.0402	0.0361
				(0.0067)	(0.0069)
Local Transaction Rate	+1 S.D., w/i. 0.1 mi			0.0469	0.0416
				(0.0076)	(0.0078)

Note: The outcome is whether the at-risk tenure enters as an investor in a given month. Standard errors in parentheses are clustered at the at-risk tenure level. Coefficients have been multiplied by 10,000 for readability. Income is the individual's stated income on the loan application for his/her primary residence, in year 2000 dollars. Current home equity is imputed using the price-index projected value minus an estimate of the current loan balance using initial equity and an amortization schedule. Local price change and transactions rate are measured for the preceding year in the distance bands noted in the table.

trends in real estate, or in his/her own home wealth that were not picked up by our wealth proxy. The spatial measures of local price appreciation (relative to the county index) and transaction activity in the last year are both positively associated with the propensity to enter, suggesting that above-average price appreciation and an active transactions market are signals to potential investors. We see this as even more evidence that Los Angeles residents are obtaining information from their immediate neighborhood. A standard deviation increase in local prices leads to a four percent increase in entry propensity, and in transactions activity, an almost five percent increase. Interestingly, however, the inclusion of these variables hardly affects the estimated effects of being exposed to our measures of neighborhood investment activity. The outer disc coefficients are slightly attenuated, but the inner ring coefficients are stable.[41] In this way, exposure to investment activity operates as a special channel of information in excess of other local signals. The bottom panel shows that an additional investor or flip exposure alters entry propensity in roughly proportional measure to a two standard deviation increase in local prices or transaction rates. Finally, column 3 includes both individual and neighborhood level controls, and the results carry through similarly.

To conclude our primary analysis, Online Appendix B demonstrates the robustness of our results to a number of potential concerns including measurement error in our name matching procedure and alternative approaches to identifying a primary residence.

D. Multifamily Properties

Our main analysis sample is limited to single-family properties because our research design relies on the quasi-random assignment of exposure in close spatial proximity. From a practical perspective, the natural thinness of the single family housing market means that there is very little additional sorting for households that live within 0.1 versus 0.2 miles of one another. While we have not included condominium properties in our main analysis, they are included in the deeds data. This suggests the possibility of using a similar research design to test for contagion among residents living in the same building or, even more finely, on the same floor within a building.

Applying an analogous research design to measure within-building or within-floor spillovers changes the nature of the identifying assumption, however. While our main analysis is based on the assumption that there is essentially no sorting in the single family housing market across fine geographies, a within-building or within-floor design would require that there is no sorting across buildings within

[41] These local real estate market indicators are more correlated with investment exposures at the outer discs than at the inner rings. For example, the correlation coefficient between price growth and neighbor entry and flips are less than 0.07 at 0.1 miles. Correlations between the 0.5 miles measures are roughly twice as large. These suggest an element of randomness to close proximity exposure, which is consistent with the intuition of our research design.

the same city block or across floors within the same building.

To examine whether there is systematic sorting across condo buildings or floors, we expand Figure 4 to show the spatial gradient of differences in income for residents of single family properties and multifamily properties with more than 20 units. Figure 6 reveals that there is systematic sorting on income for condo residents (i) as geographic proximity decreases from 0.5 to 0.1 miles, (ii) (ii) across condo residents compared to their nearest neighbor (less than 0.1 miles), and (iii) across floors within condo buildings.[42] The particularly steep decline in income differences from 0.1 to 0 miles reflects the increased similarity of residents of a given building to one another versus other condo residents within 0.1 miles. The extra little bump down in income differences at 0 shows that those who reside on the same floor are slightly more similar to one another than others in the same building, perhaps because residences on certain floors are more desirable. In contrast, the flatness of the line for single family residents as distance declines from 0.5-0.1 miles highlights just how little comparable sorting there is among single family residents at these geographies.

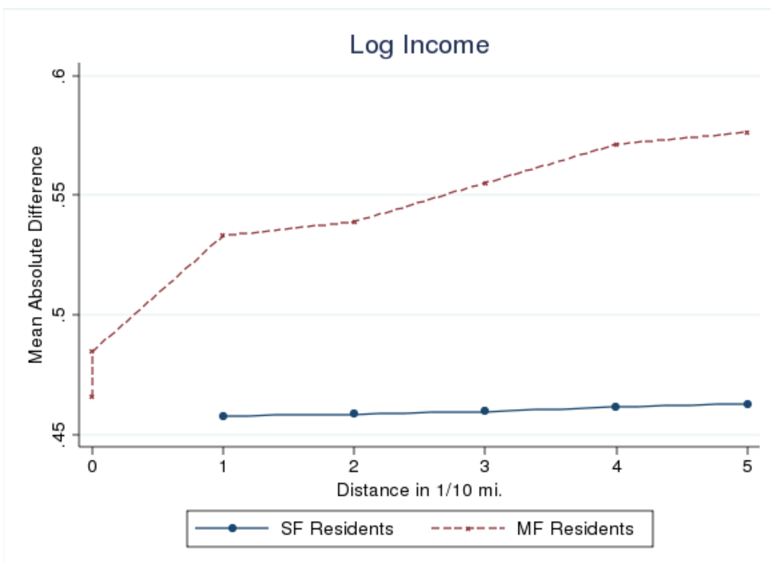

Figure 6. Comparability in Homeowner Incomes By Units in Structure.

Note: SF: Single Family, MF: Multifamily. The figure reports the average proportional difference between the income of a resident and the average income of residents within rings (annuli) of d tenths of a mile. The multifamily building series has an additional point at distance zero representing the difference between at-risk tenures and neighbors in the same floor group.

Given the substantial sorting at the building level, we focus only on whether

[42]Details for how "neighbors" are defined within a multifamily complex are given in Online Appendix E.

individuals exposed to neighbor investors or flipped properties on the same floor within a building are more likely to become real estate investors, while controlling for exposure to investment behavior on other floors within the same building. Even in this case, we present the results with caution given the clear evidence of some sorting across floors shown in Figure 6.

Table 7 shows the results of our within-building analysis. In comparison to our baseline surface distance model, the building level exposure ("Same Building") functions as the neighborhood, and the unit group exposure ("Same Floor Group"), functions as the hyperlocal near-neighbor exposure. In the hazard regression models, we measure the marginal effect of an entering investor neighbor or a flip occurring in the near neighbor group relative to its occurrence within the same building (address). We include fixed effects for the four multifamily structure types described in Online Appendix E, as well as for the various multifamily community numbering systems. The results imply that having a property flipped or a neighbor investor on the same floor within a multifamily building substantially increases the propensity that one becomes an investor, holding constant investment activity within the same building. Note that while this "near/nearer" design qualitatively bears resemblance to the surface distance model, we are cautious in making quantitative comparisons because the inner and outer groups are not directly comparable (in size or in concept) between the two distance methods.

Overall, while we are less confident that our research design is appropriate in this case due to concerns about sorting, the evidence is consistent with a neighborhood effects explanation for investor contagion. For the remainder of the paper, we return to our main sample and research design using single family properties only.

E. Marginal Effects by Neighborhood Type

Not all neighborhoods afford the same opportunity to interact with one's neighbors. In Table 8, we estimate the baseline model separately for neighborhoods of low, medium, and high density, as measured by the population density by the 2000 decennial census tract. We split the 3,400 census tracts into terciles, which in our sample of single family properties results in more at-risk tenures in lower density tracts. We have checked but do not report that each type of neighborhood shows a lack of hyperlocal sorting on entry propensity as shown for the full sample in Table 4.

The estimated contagion effect is not constant across neighborhoods but exhibits an inverted-U shape, with the largest effects in medium density places. These medium density neighborhoods, which consist primarily of tightly packed single family homes in Los Angeles and its suburbs, seem to promote the greatest neighborhood effects. In less dense places, it appears interactions among neighbors are relatively lower, but in higher density places, the marginal effect of one neighbor is lower. That the impact of a given neighbor or flipped property is smaller in high density areas of the city seems like a natural consequence of the

Table 7—Entry Hazard Regressions Using Within-Building Exposure Models in Large Multi-family Properties.

	1	2	3	4
Regression Coefficients				
Investor Neighbor				
Same floor group	1.2681	1.2396	1.2791	1.2643
	(0.5456)	(0.5457)	(0.5477)	(0.5491)
Same Building	0.2850	0.3156	0.3287	0.4283
	(0.1422)	(0.1427)	(0.1482)	(0.1520)
Flip				
Same floor group	1.0717	1.0424	1.0837	1.1017
	(0.7558)	(0.7560)	(0.7572)	(0.7609)
Same Building	1.2399	1.2501	1.2444	1.2726
	(0.2031)	(0.2033)	(0.2047)	(0.2065)
Constant	8.8202	8.5862	8.7872	8.7073
	(0.1607)	(0.2269)	(0.1644)	(0.1676)
Marginal Effect to Hazard Rate				
Investor Neighbor				
Same floor group	0.1438	0.1405	0.1450	0.1433
	(0.0621)	(0.0621)	(0.0623)	(0.0625)
Flip				
Same floor group	0.1215	0.1182	0.1229	0.1249
	(0.0859)	(0.0859)	(0.0860)	(0.0864)
Fixed Effects:				
Structure Type		yes		
Unit No. Type			yes	
Structure X Unit No. Type				yes

Note: The table presents hazard regression results for a variant of the nearest neighbor model for multi-family buildings in which the “outer neighborhood” is exposure within-building and the “inner neighborhood” is exposure within the same floor/floor group. See the text for details and variable descriptions. Standard errors in parentheses are clustered at the property tenure level. Coefficients have been multiplied by 10,000 for readability.

fact that individuals are less likely to interact with any particular neighbor on dense city blocks.

Table 8—Contagion Effects by Neighborhood Density.

Tract Density Tercile	1 Low	2 Medium	3 High
Summary Statistics			
Tenures (N)	1,020,196	729,717	364,431
Obs (NT)	50,535,208	36,594,549	17,519,271
Entries	27,149	21,881	13,910
Entry Rate (x10,000)	5.37	5.98	7.94
Regression Coefficients			
Investor Neighbor			
w/i. 0.1 mi	0.3416	0.4172	0.2002
	(0.0790)	(0.0875)	(0.1115)
w/i. 0.3 mi	0.1327	0.0701	-0.0633
	(0.0412)	(0.0432)	(0.0587)
w/i. 0.5 mi	0.0441	0.1966	0.1616
	(0.0233)	(0.0239)	(0.0330)
Flip			
w/i. 0.1 mi	0.2725	0.6210	0.2910
	(0.0981)	(0.1143)	(0.1298)
w/i. 0.3 mi	0.1083	0.0768	0.0144
	(0.0514)	(0.0537)	(0.0698)
w/i. 0.5 mi	0.3060	0.3226	0.2331
	(0.0282)	(0.0292)	(0.0367)
Constant	4.0539	3.6525	5.8877
	(0.0505)	(0.0740)	(0.1215)
Marginal Effect to Hazard Rate			
Investor Neighbor			
w/i. 0.1 mi	0.0843	0.1142	0.0340
	(0.0195)	(0.0241)	(0.0189)
Flip			
w/i. 0.1 mi	0.0672	0.1700	0.0494
	(0.0242)	(0.0315)	(0.0220)

Note: The outcome is whether the at-risk tenure enters as an investor in a given month. Standard errors in parentheses are clustered at the at-risk tenure level. Coefficients have been multiplied by 10,000 for readability.

F. *Temporal Pattern and Cumulative Effects*

The analysis thus far has focused on the impact of a one-year lag in exposure in order to test for the presence of a neighborhood effect. A broader analysis of the effects of contagion should test the length and strength of the total impact of exposure. In Table 9, we include measures of exposure for the inner rings and outer discs for lags up to four years.[43]

[43] In the notation of Equation (2), we are now using a summation over lags s corresponding to month ranges of $\{0 : 11, 12 : 23, 24 : 35, 36 : 47\}$.

For brevity, the table reports only the innermost ring effect size, though we have included all lags of outer discs as controls in accordance with our research design. We report the single lag hazard rate as well as the cumulative effect; that is, how an exposure changes an individual's propensity to enter as the exposure ages from one to four years. The results show a statistically detectable effect of each type of investing activity up to three years, and each appears to die off by the fourth year. Thus, an individual exposed to an investor neighbor is 14 percent more likely to enter within three years of initial exposure. For flips, the three year cumulative effect is 16 percent. We use these cumulative effect sizes to measure the total impact of contagion below.

Table 9—Temporal Pattern of Contagion Effect.

Years Lagged:	1	2	3	4
Investor Neighbor				
w/i. 0.1 mi	0.1964	0.1571	0.1969	0.0054
	(0.0541)	(0.0533)	(0.0565)	(0.0593)
Marginal Effect to Hazard Rate				
partial, w/i. 0.1 mi	0.0502	0.0392	0.0494	0.0015
	(0.0132)	(0.0130)	(0.0138)	(0.0145)
cumulative, w/i. 0.1 mi		0.0893	0.1388	0.1402
		(0.0177)	(0.0211)	(0.0238)
Flip				
w/i. 0.1 mi	0.2100	0.2142	0.1537	0.0370
	(0.0671)	(0.0677)	(0.0707)	(0.0797)
Marginal Effect to Hazard Rate				
partial, w/i. 0.1 mi	0.0604	0.0573	0.0403	0.0094
	(0.0164)	(0.0165)	(0.0173)	(0.0195)
cumulative, w/i. 0.1 mi		0.1177	0.1580	0.1674
		(0.0221)	(0.0262)	(0.0303)

Note: The outcome is whether the at-risk tenure enters as an investor in a given month. Standard errors in parentheses are clustered at the at-risk tenure level. Coefficients have been multiplied by 10,000 for readability.

G. *Contagion In Other Cities*

Our analysis has centered on the Los Angeles market, mainly because it experienced a major housing cycle event and there is a long transaction record which allows us to identify investing behavior. To explore the external validity of our results, we now expand our analysis to two other metro areas for which Dataquick maintains a long transaction record since 1988, the San Francisco Bay Area and the greater Boston, Massachusetts area. We clean the data, flag investors, and build a spatial match of exposures for single family properties in San Francisco and Boston just as we did for Los Angeles.

Figures depicting their housing markets and summary statistics of the estimation samples are reported in Online Appendix D. Figure D1 shows the time series of transactions, price indices, and investment activity. Similar to Los Angeles, San Francisco experienced a large boom and bust in prices and transactions over the period 2000-2007 (with strong price appreciation beginning slightly earlier than Los Angeles). Investment activity showed a similar time pattern, though the average level was lower than Los Angeles. About 2.3 percent of at-risk individuals become investors (compared to three percent in Los Angeles), and investments and flips account for about 9 and 1.7 percent of transactions, respectively, compared to 16 and 3.1 percent in Los Angeles. Boston also experienced a period of strong price appreciation, but the growth was more steady and gradual than the California markets.[44] Transaction levels are relatively more steady (and more seasonal), though they do decline beginning in 2005. Investment activity is notably less cyclical in Boston. About two percent of individuals are marked as investors, and investments and flips make up, respectively, 8.1 and 1.7 percent of transactions.

The relevant question is then, given the differences in these markets compared to Los Angeles, are there still neighborhood effects in investing? Table 10 reports hazard regressions of the form of Equations (2) and (4) for San Francisco and Boston data. For each, we report the baseline model with outer disc exposures as controls, as well as robustness checks that include dummies for time and broader neighborhood (columns 2 and 5) or individual and local neighborhood controls (columns 3 and 6) as we did in Table 6 for Los Angeles. The upper panel reports regression coefficients, and the lower panel the marginal effect sizes.

The table shows clear evidence that the neighborhood effects in both cities are of a remarkably similar pattern and magnitude to those in Los Angeles. In San Francisco, an investor neighbor in the innermost ring changes entry propensity by about eight percent, and a flip changes entry propensity by about 10 percent. In Boston, the effect sizes are about eight and 11 percent, respectively. Contagion effects may be highly local, but they are not confined to one market.

H. The Impact of Investor Contagion

We close the section on the main results by calculating an estimate of the market-wide effect of investor contagion. The regressions above measured whether there is a statistically meaningful effect of exposure to an investor neighbor or flipped property on entry probability. The marginal effect conditions on an exposure, while the aggregate effect depends on both the change to entry propensity measured in the regressions and the degree of exposure to each of the treatments. We next calculate the total impact the presence of neighborhood effect had on the markets we study. It is important to note that the following is purely an accounting exercise at observed exposures, and we do not attempt to counterfactually predict entry or exposure at alternative prices or volumes of transactions.

[44]Since the run-up in prices was longer in Boston, we extend the main data period there to include 1998 and 1999.

Table 10—Contagion in Other Cities.

	1	2	3	4	5	6
	San Francisco Bay Area			Boston Area		
			Regression Coefficients			
Investor Neighbor						
w/i. 0.1 mi	0.2598	0.2036	0.2162	0.3242	0.2956	0.2806
	(0.1029)	(0.1036)	(0.1032)	(0.1302)	(0.1312)	(0.1319)
w/i. 0.3 mi	0.1827	0.1438	0.1239	0.0910	0.0867	0.0888
	(0.0542)	(0.0543)	(0.0548)	(0.0769)	(0.0774)	(0.0777)
w/i. 0.5 mi	0.3304	0.1580	0.2352	0.0005	-0.0609	0.0002
	(0.0286)	(0.0304)	(0.0290)	(0.0426)	(0.0470)	(0.0433)
Flip						
w/i. 0.1 mi	0.3220	0.2586	0.2372	0.4679	0.4752	0.4159
	(0.1636)	(0.1639)	(0.1646)	(0.1937)	(0.1946)	(0.1957)
w/i. 0.3 mi	0.2369	0.2006	0.1535	0.0668	0.0640	0.0641
	(0.0820)	(0.0821)	(0.0829)	(0.1166)	(0.1165)	(0.1181)
w/i. 0.5 mi	0.2988	0.1981	0.2158	-0.0157	-0.0671	-0.0280
	(0.0444)	(0.0463)	(0.0451)	(0.0693)	(0.0729)	(0.0707)
Constant	2.5855	-0.4134	1.2674	3.6226	1.7298	3.1494
	(0.0610)	(0.3442)	(0.5062)	(0.0591)	(1.0637)	(0.4774)
			Marginal Effect to Hazard Rate			
Investor Neighbor						
w/i. 0.1 mi	0.1005	0.0787	0.0836	0.0895	0.0816	0.0775
	(0.0398)	(0.0401)	(0.0400)	(0.0360)	(0.0362)	(0.0364)
Flip						
w/i. 0.1 mi	0.1246	0.1000	0.0917	0.1292	0.1312	0.1148
	(0.0633)	(0.0634)	(0.0637)	(0.0535)	(0.0538)	(0.0540)
Qtr Effects		yes			yes	
ZIP Effects		yes			yes	
Spatial Controls			yes			yes
Individual Controls			yes			yes

Note: The outcome is whether the at-risk tenure enters as an investor in a given month. Standard errors in parentheses are clustered at the at-risk tenure level. Coefficients have been multiplied by 10,000 for readability.

To measure the effect of contagion, we need to determine (borrowing more epidemiological language) the contact rate and the transmission rate from one investor to other at-risk individuals. The transmission rates were measured in the regressions. Every investor enters, by definition, meaning they produce β^N more investors for each at-risk contact they have in their neighborhood. But not everyone conducts a flip (some may hold their investment properties for longer periods of time), so only a subset will make contacts through that channel. Thus, the basic reproduction rate (Rothman (2012)) is the weighted average

$$R = \beta^N \cdot C^N + \beta^F \cdot Pr(F) \cdot C^F \tag{5}$$

where the β effect sizes come from the regressions, C^N, C^F refer to the contact rates of each type of exposure (neighbor and flips, respectively), and $Pr(F)$ is the probability that the investor flips a property. Because the contacts we consider are spatial, C is measured as property density in the average investor's (or flip's) neighborhood.[45] The typical property in Los Angeles is within a 0.1 mile radius of about 70 other properties. $Pr(F)$ can be measured directly from the data as the fraction of investment properties that are flipped. The cumulative effect, then, is a product of the reproduction rate and how many recently entered investors there are of each vintage in a given period. The change in the number of investors, Δe_t, is

$$\Delta e_t = \sum_s R_s N_{t,s} \tag{6}$$

where $N_{t,s}$ is the number of investors observed in t to enter in lags of s. We sum over the three lags as measured by Table 9. For each month t, we count how many investors enter in the last year, one to two years, and two to three years, which yields $N_{t,s}$.[46] Finally, we report this as a rate of increase over a world in which no neighborhood effect was present; if e_t is observed entry rates, we measure the rate of increase as $\frac{\Delta e_t}{e_t - \Delta e_t}$.

Table 11 reports the result of this calculation for the three metro areas. On average over the study period, contagion produced 10.3 percent more investors operating in the Los Angeles market than there would have been absent a neighborhood effect mechanism. The prevalence of investor entry and flip occurrence increased as prices rose, which produced more investors in raw terms, although baseline entry rates also went up, so that the rate of increase from contagion was actually slightly smaller during 2004-2006 by this metric. Effect sizes for San

[45] We measure densities using the full assessor record, which includes properties that did not transact during our data period. We limit to single family properties to remain consistent with our research design.

[46] We measure density separately for each investor cohort, although it turns out to be of little consequence.

Francisco and Boston are of a similar magnitude, from nine to 14 percent. Thus, even though these markets saw less investing activity on average, the proportion attributable to contagion was comparable to Los Angeles.

Finally, note that we are measuring the effect of contagion from the innermost ring (0.1 miles) in excess of the 0.3 and 0.5 discs. If one were to accept the effects at wide distances as causal, the projected impact size would rise according to the coefficient size and also the increase in contact rate admissible in wider areas.

Table 11—Increase in Investors Due to Contagion.

	2000-2007	2001-2003	2004-2006
Additional Investors Created (%)			
Los Angeles	10.31	9.41	8.93
San Francisco	14.34	13.58	11.84
Boston	9.63	9.10	9.02

Note: The table reports effect size calculations based on estimates of the three-year marginal effect of exposure within the inner spatial ring (0.1 mi); see Table 9.

III. Exploring Mechanisms: Information Sharing

The foregoing analysis has tested for the existence of a neighborhood effect and estimated its size. Of course, detecting the presence of the effect does not by itself explain the mechanism that produces it, and whether it categorizes as social learning or social utility. However, the literature on social effects in bubbles often points to social learning, and we find those learning mechanisms to be plausible a priori in this particular context. With this in mind, we test for an information sharing mechanism available in the data: the use of particular lenders to finance investment purchases. Here and following, we return to our focal market, Los Angeles.

The transactions record includes the name of any lender(s) with liens on the property. We use this information to test whether investors are more likely to use the same lender as their investor neighbors. As an initial matter, we check how likely two neighbors or any type are to use the same lender.[47] In Table 12, we use the matched sample from the entry sorting index test of Section II.A to measure the propensity of any two owner-occupant-flagged properties to use the same lender when financing the transaction. In new residential developments, vertical relationships between construction, sales, and lending are much more likely to occur, so we split the analysis into three sets organized by property age. Column 1 limits to properties no more than two years old, column 2 to 3-5 years old, and the last column older than 5 years. New properties are far more likely to be financed by the same lender, and the gradient is steeply increasing as we

[47]We thank a referee for suggesting this exercise.

narrow from 0.5 to 0.1 miles. There is a substantial drop off in propensity for even slightly older properties, and the gradient flattens considerably. Column 3, which limits to properties at least five years since construction, shows a flat 2.75 percent propensity demonstrating that there is no evidence of excess likelihood of using the same lender at small spatial scales compared to the slightly larger neighborhood.[48]

TABLE 12—MEAN PROBABILITY OF OWNER-OCCUPANTS USING THE SAME LENDER.

	1	2	3
Category	New Construction	Recent Construction	Older Construction
Property Age	$\leq$ 2 years	3-5 years	> 5 years
Distance			
0.0-0.1 mi.	0.0940	0.0267	0.0274
0.1-0.2 mi.	0.0663	0.0252	0.0274
0.2-0.3 mi.	0.0514	0.0246	0.0274
0.3-0.4 mi.	0.0442	0.0247	0.0274
0.4-0.5 mi.	0.0424	0.0244	0.0275

Note: The table reports the mean probability of owner-occupants using the same lender to finance their purchase transactions in one-tenth mile rings. The categories refer to property age at time of transactions; that is, the time elapsed between construction date and transaction date. Results by time elapsed between transactions are available on request.

The test of whether two investors use the same lender for investment transactions involves a change of sample and two more degrees of separation. For this test, we (i) limit the spatial match to two paired investors instead of all owner occupants, and (ii) compare the lender not for their home property, but the one used for their respective investment properties, which are likely much farther apart. Considering a reference investor as the center of a circle, we match her with, for example, two other investors within a 0.5 mile radius, one of those very close (say, less than 0.1 miles) and another farther (say, 0.4 miles). We measure the likelihood that the center investor uses the same lender for her investment (not primary residence) as her neighbor who lives down the block used for his investment, as compared to the same-lender propensity for the focal investor and the neighbor 0.4 miles away. We exclude from the radius investors who entered after the center investor, and exclude from the observations any transactions that occur after the central investor's transaction of interest.[49] We limit the matches to a maximum distance of one-half mile between investors, and exclude investments without a loan present. Finally, in light of Table 12, we drop properties less than five years old.

Table 13 reports on the results of this matching procedure. Of nearly two million

[48]Note that 'age' refers to the time lag between construction date and transaction date, not necessarily the time between the matched transactions. We have found (in unreported results, available on request), that conditioning on years elapsed between transactions changes the average same-lender propensity, but not the spatial gradient.

[49]These matches would appear on the other side, i.e. with treatment and outcome reversed, elsewhere in the matched dataset. Because of the timing conditions, the pairings are not duplicated.

matches of proximity 0.5 miles or less between the investors, 1.4 percent use the same lender for financing their investment transaction as did their previously entering neighbor. This is slightly less than the owner occupant sample in Table 12, but we are comparing a set of investment properties typically farther than 0.5 miles apart. We then estimate a regression of the same-lender indicator on an indicator that the investors are very close, within 0.1 mile, and report in the lower panel the marginal change to propensity. Column 1 reports that this propensity increases nearly 15 percentage points for neighbors very close by. Columns 2 and 3 split the lenders into uncommon and common lenders, the distinction being less than or more than a two percent market share of all mortgage originations in the year of investment purchase.[50] This separates large lenders (such as Wells Fargo or Countrywide) from smaller and more local firms with less brand recognition and possibly different underwriting procedures. The average same-lender incidence is mechanically lower for uncommon lenders and higher for common lenders,[51] but there is a notable difference in marginal effect of very close neighbors. When an investor uses an atypical lender to finance an investment purchase, their very close neighbors are 25 percent more likely to use the same one for their investment property. For common lenders, the marginal increase from very close neighbors is just seven percent.

Table 13—Probability of Using the Same Lender as Spatially Proximate Investors and Flips.

	1	2	3	4	5	6
Matched To:	Investor Neighbor			Flip		
Lender Type	All	Uncommon	Common	All	Uncommon	Common
			Summary Statistics			
Obs.	1,613,106	728,500	176,659	997,841	464,001	103,437
Mean	0.014	0.006	0.092	0.015	0.007	0.101
St. Dev.	0.118	0.077	0.289	0.121	0.080	0.301
			Marginal Effect of Nearest Matches			
Within 0.1 mi	0.1526	0.2531	0.0685	0.0148	0.0298	-0.0116
	(0.0332)	(0.0705)	(0.0349)	(0.0345)	(0.0790)	(0.0375)

Note: The data consist of investment properties matched to the investor's exposure investment activity within 0.5 miles. The spatial match in columns 1 to 3 is that of investors to their investor neighbors; columns 4 to 6 much investors to flips nearby the investors. "Uncommon" lenders are those that appear on less than two percent of transactions in the year of investment purchase, and "common" appear on more than two percent. Newly constructed homes (up to 5 years in age) are excluded. Standard errors in parentheses are clustered at the investor level.

One possible concern with this result is that certain lenders may operate in particular neighborhoods and not in others, which would give rise to spatial patterns in the use of lenders. There are good reasons to believe this is not the case here.

[50]The cutoff was chosen after inspection of the market share distribution. The result is robust to alternative choices of cutoff.

[51]Of course, if one investor uses a common lender and the other an uncommon, the propensity for same lender is zero by definition.

First, the match has already conditioned on being within a half-mile radius, so that we measure only the increased propensity to use the lender of an extremely close neighbor relative to one slightly farther away. Moreover, the results from Table 12 indicate that there is not much hyper-local spatial correlation in the use of the same lender in general.

Second, we can use nearby flipped properties as a placebo test. The remaining columns of Table 13 conduct the same exercise, but spatially matching investors to flips occurring in their neighborhood instead of investor neighbors (i.e., properties instead of people). In this match, the center investor is near the flip and may not know its purchaser or any details on the transaction, and hence we think of it as a placebo test of the peer effect. The outcome is again whether the center investor uses the same lender as an investment property, except now it is whether the flip (not its buyer) is in the same neighborhood and on the same block. The raw probability of having the same lender is comparable to when we match by investor proximity: 1.5 percent on average, 0.6 percent for uncommon lenders, and 10 percent for common lenders. But the increase for very close matches is not present for nearby flipped properties, with the marginal changes to propensity estimated to be zeros and measured with similar precision as columns 1 to 3.

Taken together, the evidence of Table 13 strongly suggests a referral-like process. When the investors are near each other, new entrants are more likely to use the same lender as their neighbors to finance their investments. When new entrants are near a flip, on the other hand, they are much less likely to have a relationship with the buyer, and we see no effect on choice of lender.

Certainly this is not the only channel of information sharing; indeed, the average incidence is low enough to see that it is not the whole story of contagion. Rather, we see this an illuminating example of a particular form of information sharing. There are possibly many dimensions on which information sharing could be present, from other referrals such as realtors, contractors, or closing attorneys, to various "tricks of the trade" in locating properties and executing transactions. That we see evidence of this behavior in one observable component suggests the existence of spatial "peer effects" through information sharing.

IV. Neighborhood Influence and Investor Success

While the language of "contagion" may seem pejorative, it is not obvious a priori whether investors susceptible to influence in their immediate neighborhood should perform better or worse than other investors. On the one hand, extremely local information may be especially salient and useful, while on the other, the influenced investors may be naïvely mimicking behavior around them without any special skill or insight. This is, in principle, an empirical question. The goal of this section is to examine the subsequent performance of investors that are drawn into the real estate market because of the influence of neighbor investors or nearby flipped properties.

The analysis presented in this section reports the results of regressions that relate various aspects of an investor's performance to measures of neighborhood investment activity at the time of investor entry. The observation level is the investment property and the explanatory variable is whether the investor (the purchaser of the property) was influenced at the time of entry. Of course, we do not actually observe whether any one particular investor was "influenced," only whether he or she was exposed. Some potential investors who are exposed would have entered anyway as a result of the baseline rate of investor entry during this period, and some exposed at-risk tenures will not enter. Thus, to properly interpret effects for those actually drawn in by neighborhood activity, we must scale the magnitude of the measures of neighborhood investment activity by the inverse of the estimated cumulative marginal effect of investment activity on entry from Section 3.[52] Note that this scaling is a matter of interpreting the effect size, and does not affect the detection of statistical significance. We estimate regressions of the form

$$y_p = \alpha_1[\hat{\beta}^N I(N_{entry})] + \alpha_2[\hat{\beta}^F I(F_{entry})] + D_{cohort} + \varepsilon_p \tag{7}$$

where y_p is the outcome for property p, α_1, α_2 are the coefficients of interest, $I(\cdot)$ are indicators for investor neighbor and flip exposure, and $\hat{\beta}$ are the marginal changes to entry rate hazards from Table 5. The regression compares outcomes cross-sectionally between investors with and without exposure. Of course, exposure varies over time, and different entering cohorts face different price and transaction regimes that could affect their behavior, and the cohorts may have been more or less sophisticated for other reasons. We do not want to attribute cohort effects to contagion influence. Therefore in all specifications we include dummies for the year of the investor's entry, D_{cohort}, as controls.

Table 14 reports the results, beginning with capital returns. In column 1, the outcome is the annualized gross rate of return on properties observed to be sold within our data period. To be clear, we do not observe all transaction costs at purchase or sale, or what is occurring while the property is held (e.g. whether it is rented), so this is not an exact accounting of the investor's returns. It is, however, a readily available metric of success and a natural starting place. The average investment property purchased and observed to be sold earns about 12 percent per year, with a very wide standard deviation. However, the influenced entrant earns substantially less: 24 percent less for influence from a neighbor and 10 percent for a flip, even controlling for entry cohort.[53] Columns 2 and 3 go deeper to test whether influenced investors fared worse on purchase or sales prices relative to their expected values (using a county-level price index). In

[52]This adjustment is analogous to distinguishing a "treatment on the treated" vis-a-vis an "intent-to-treat" effect.

[53]This does not mean that influenced investors earn negative returns, since they may hold their properties longer than a year (and they typically do). Also, recall that these effects were scaled by the change to entry propensity.

other work, (Bayer et al. (Forthcoming)), we find these excess returns at time of transaction, especially purchase discounts, make up a substantial share of investors' gross returns. The average investor in this period is buying and selling very close to market rates, and the positive coefficients in column 2 indicate that the influenced investor is actually buying slightly above market rate, while the negative coefficients in column 3 indicate the influenced investor is selling slightly below market. Thus, influenced investors seem on average less adept at locating "good deals" and negotiating transactions. The remainder of the performance gap would then be related to market timing. Column 4 examines market timing directly by comparing the projected sales price minus the projected purchase price, using county-level repeat sales index models for projection.[54] Influenced investors indeed earn inferior market-based returns, on average, even conditioning on entry cohort.

The realized returns models condition on observing a sale. Column 5 expands the scope of analysis of market timing to test whether influenced investors were more likely to hold on to their properties "too long," defined here to be past the point of break-even in the price cycle. For this, we limit to properties purchased during the period of rapid price acceleration, 2003-2006Q2,[55] and do not condition on observing a sale.[56] While a majority of investors held their properties past the price index level at time of purchase, investors influenced by investor neighbors and flips were, respectively, 15 and eight percentage points more likely to do so.

Finally, columns 6, 7, and 8 consider the equity position and foreclosure incidence for the influenced investors. Equity position at time of purchase is specifically in the transaction record, but foreclosure has to be inferred by observing the property tenure ending with ownership transferring to an institutional entity (like a bank).[57] Column 6 shows that while the average investor makes about a 17 percent down payment, influenced investors take out mortgages with LTVs 11 to 13 percentage points lower. Column 7 uses a discrete measure of low equity, an indicator for whether the down payment consisted of less than five percent of purchase price. While low equity is not uncommon for investment properties in Los Angeles (slightly more than a third of investments; see Table C2), influenced investors bought with low equity at a rate of 18 to 21 percentage points more often. Finally, and perhaps as a result of the findings in columns 4 to 7, we find that influenced investors were about twice as likely to end up in foreclosure.

In summary, it appears the influenced investor on average fared worse on a variety of metrics, despite the possible information sharing that may have been taking place. If this were only a product of the influenced investor being a neo-

[54]There could of course be spatial heterogeneity in market dynamics. The indices are county specific, but that is the extent of spatial detail measured in column 4. Finer local area price differentials would be accounted for in the residual models in columns 2 and 3.

[55]All five counties in greater Los Angeles peaked in 2006Q2, although rates of ascent prior, and rates of decline afterwards, varied somewhat.

[56]We exclude properties flagged as foreclosed-upon; more on that in column 8.

[57]Note that this is the end result of the foreclosure process, and we are not measuring loan performance/delinquency in the interim.

Table 14—Effect of Entry Contagion on Outcomes for Investment Properties.

	1	2	3	4	5	6	7	8
	Rate of Return (ann'lized)	Buy Residual	Sell Residual	Market Return	Held Past Breakeven	Pct. Down	Low Equity	Foreclosed
Obs.	49,808	49,808	49,808	49,808	45,758	108,671	108,671	108,671
Mean	0.121	0.011	-0.003	0.078	0.607	0.175	0.347	0.116
St. Dev.	0.326	0.154	0.189	0.229	0.488	0.253	0.476	0.320
				Regression Coefficients				
Inv. Neighbor w/i. 0.1mi	-0.248	0.061	-0.058	-0.112	0.155	-0.129	0.181	0.074
	(0.0377)	(0.0193)	(0.0237)	(0.0245)	(0.0616)	(0.0211)	(0.0397)	(0.0265)
Flip w/i. 0.1mi	-0.104	0.069	-0.020	-0.047	0.080	-0.117	0.216	0.127
	(0.0307)	(0.0157)	(0.0193)	(0.0199)	(0.0508)	(0.0173)	(0.0326)	(0.0217)
Purchases Made	2000-2007	2000-2007	2000-2007	2000-2007	2003-2006Q2	2000-2007	2000-2007	2000-2007
Entry Year	2000-2007	2000-2007	2000-2007	2000-2007	2000-2007	2000-2007	2000-2007	2000-2007

Note: The table reports summary statistics and estimates of marginal effects of investment exposure at time of entry on several outcomes for investment properties. Column 1 calculates the annualized rate of return on properties observed to be sold by the investor, trimming outliers greater than 200 percent. Column 2 calculates the implied residual at the time of purchase relative to expected market rate as measured by a county-level quarterly price index. Column 3 uses the residual at time of sale. Column 4 is an indicator for whether the property is held longer than the point in which the price index dropped below the time of purchase. Column 5 calculates the percent down payment (equity at time of purchase). Column 6 is an indicator for whether the initial equity was less than five percent. Column 7 measures a foreclosure by the investment property tenure ending in an action by an institutional entity. All specifications include as controls indicator variables for the investor's year of entry. Standard errors are in parentheses.

phyte, then the controls for entry cohort should sweep the effects away, and they do not. Instead, it seems the influenced investor is on average of lower skill. One speculation is that susceptibility to influence is correlated with unsophistication–e.g. taking tips without doing any fundamental research–although we know of no precise way to test this directly. However, if there is a silver lining for the influenced investor, it is that some of the money lost is not their own, but, through default, is shared with their mortgage holders.

V. Conclusion

It is a widely-held notion that financial crises and bubble-like episodes often feature contagion in investment behavior. Colorful anecdotes about novice investors being drawn into a market by others date back to some of the earliest accounts of historical asset price bubbles. However, despite the long-standing theoretical interest in asset price dynamics, and asset pricing models' incorporation of these types of contagion effects, the existing evidence on actual investor contagion is just that: anecdotal.

This paper provides some of the first evidence of contagion in investment behavior during an asset bubble using a research design that credibly isolates causation from correlated investment activity across individuals. We focus on the recent boom and bust in the U.S. housing market and consider whether individual households become investors in the housing market due to the activity of other investors. Our approach to credibly identifying such contagion in investment behavior relies on a nearest-neighbor research design that identifies the causal effect of nearby investment activity on a potential investor's behavior by estimating the effect of hyper-local investment activity while controlling for similar measures of activity at a slightly larger neighborhood. We show this procedure is robust to the inclusion of a wide range of other control variables. Moreover, we advance the literature on "peer effects" in investing by documenting the poor performance of the "influenced" investors - this is critically important for understanding the effects of investor contagion during bubble-like episodes. Using our research design, we show that a contagion effect was present, the magnitude was large even under a conservative measurement, and that the segment of investors drawn into the market via the actions of others performed worse than other investors.

A natural question one might ask is how exactly the influence occurs. The empirical literature on peer effects in investing usually cannot distinguish between social learning and social utility mechanisms. However, we suspect that social learning is relatively more important in our context. First, there is a large theory literature on speculative bubbles that often features such learning; second, there are numerous ways that this type of learning could take place (e.g., novice investors could change their beliefs about market fundamentals by observing others' actions, or learn practical tips for investing); and third, we find evidence of one specific channel through which such communication can be valuable: information shared about which lender to use. Nevertheless, we believe that precisely separating out the origins of the speculative contagion we identify in this paper is an exciting area for future work.

REFERENCES

Allen, Franklin, and Douglas Gale. 2000. "Financial Contagion." *Journal of Political Economy*, 108(1): 1–33.

Anenberg, Elliot, and Edward Kung. 2014. "Estimates of the Size and Source of Price Declines Due to Nearby Foreclosures." *American Economic Review*, 104(8): 2527–2551.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018. "The Economic Effects of Social Networks: Evidence from the Housing Market." *Journal of Political Economy*, 126(6): 2224–2276.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2013. "The Diffusion of Microfinance." *Science Magazine*, 341(6144).

Banerjee, Abhijit V. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics*, 107(3): 797–817.

Basu, Ritu. 2002. "Financial Contagion and Investor Learning: An Empirical Investigation." *IMF Working Paper 2-218.*

Bayer, Patrick, Christopher Geissler, Kyle Mangum, and James W. Roberts. Forthcoming. "Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market." *Review of Financial Studies.*

Bayer, Patrick, Kyle Mangum, and James W. Roberts. 2020. "Replication data for: Speculative Fever: Investor Contagion in the Housing Bubble." *American Economic Association, Inter-university Consortium for Political and Social Research, OPENICPSR-120446.*

Bayer, Patrick, Rob McMillen, Alvin Murphy, and Chris Timmins. 2016. "A Dynamic Model of Demand for Houses and Neighborhoods." *Econometrica*, 84(3): 893–942.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy*, 116(6): 1150–1196.

Bernheim, B. Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy*, 102(5): 841–877.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy*, 100(5): 992–1026.

Brown, Jeffery R., Zoran Ivkovic, Paul A. Smith, and Scott Weisbenner. 2008. "Neighbors Matter: Causal Community Effects and Stock Market Participation." *Journal of Finance*, 63(3): 1509–1531.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. 2016. "Understanding Booms and Busts in Housing Markets." *Journal of Political Economy*, 124(4): 1088–1147.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman. 2014. "Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions." *Econometrica*, 82(4): 1273–1301.

Calvo, Guillermo, and Enrique Mendoza. 1996. "Mexico's Balance-of-Payments Crisis: a Chronicle of a Death Foretold." *Journal of International Economics*, 41(3): 235–264.

Campbell, John, Stefano Giglio, and Parag Pathak. 2011. "Forced Sales and House Prices." *American Economic Review*, 101(5): 2108–2131.

Chamley, Christophe. 2004. *Rational Herds: Economic Models of Social Learning.* Cambridge University Press.

Chari, V.V., and Patrick Kehoe. 2003. "Hot Money." *Journal of Political Economy*, 111(6): 1262–1292.

Chinco, Alex, and Christopher Mayer. 2015. "Misinformed Speculators and Mispricing in the Housing Market." *The Review of Financial Studies*, 29(2): 486–522.

Corcos, Alain, Jean-Pierre Eckmann, Andreas Malaspinas, Yannick Malevergne, and Didier Sornett. 2002. "Imitation and Contrarian Behaviour: Hyperbolic Bubbles, Crashes and Chaos." *Quantitative Finance*, 2(4): 264–281.

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker. 2015. "Environmental Health Risks and Housing Values: Evidence from 1600 Toxic Plan Openings and Closings." *American Economic Review*, 105(2): 678–609.

Currie, Janet, Michael Greenstone, and Enrico Moretti. 2011. "Superfund Cleanups and Infant Health." *American Economic Review*, 101(3): 435–441.

DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." *Journal of Finance*, 45(2): 379–395.

Duflo, Esther, and Emmanuel Saez. 2002. "Participation and Investment Decisions in a Retirement Plan: the Influence of Colleagues' Choices." *Journal of Public Economics*, 85(1): 121–148.

Ellison, Glenn, and Drew Fudenberg. 1993. "Rules of Thumb for Social Learning." *Journal of Political Economy*, 101(4): 612–643.

Favilukis, Jack Y, and Stijn Van Nieuwerburgh. 2017. "Out-of-town Home Buyers and City Welfare." *CEPR Discussion Paper No. DP12283.* working paper, SSRN 2922230.

Froot, Kenneth A., David S. Scharfstein, and Jeremy C. Stein. 1992. "Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation." *Journal of Finance*, 47(4): 1461–1484.

Griffin, John M, and Gonzalo Maturana. 2016. "Who Facilitated Misreporting in Securitized Loans?" *The Review of Financial Studies*, 29(2): 384–419.

Gupta, Arpit. forthcoming. "Foreclosure Contagion and the Neighborhood-Spillover Effects of Mortgage Defaults." *Journal of Finance.*

Haughwout, Andrew, Donghoon Lee, Joseph Tracy, and Wilbert van der Klaauw. 2011. "Real Estate Investors, the Leverage Cycle, and the Housing Market Crisis." *FRB of New York Staff Report 514.*

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein. 2004. "Social Interaction and Stock-Market Participation." *Journal of Finance*, 59(1): 137–163.

Jackson, Matthew O. 2010. *Social and Economic Networks.* Princeton University Press.

Kindelberger, Charles P. 1978. *Manias, Panics and Crashes: A History of Financial Crises.* . 2005 ed., John Wiley & Sons, Inc.

Kirman, Alan. 1993. "Ants, Rationality, and Recruitment." *Quarterly Journal of Economics*, 108(1): 137–156.

Linden, Leigh, and Jonah Rockoff. 2008. "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws." *American Economic Review*, 98(3): 1103–1127.

Lux, Thomas. 1995. "Herd Behaviour, Bubbles and Crashes." *The Economic Journal*, 105(431): 881–896.

Lux, Thomas. 1998. "The Socio-economic Dynamics of Speculative Markets: Interacting Agents, Chaos, and the Fat Tails of Return Distributions." *Journal of Economic Behavior and Organizations*, 33(2): 143–165.

Mackay, Charles. 1841. *Extraordinary Popular Delusions and the Madness of Crowds.* . 1932 ed., Richard Bentley.

Maturana, Gonzalo, and Jordan Nickerson. forthcoming. "Teachers Teaching Teachers: The Role of Workplace Peer Effects in Financial Decisions." *The Review of Financial Studies.*

McCartney, W. Ben, and Avni Shah. 2016. "The Economic Importance of Neighbors: Evidence from Hyperlocal Social Influence Effects in Mortgage Markets." `https://ssrn.com/abstract=2882317`.

Mian, Atif, and Amir Sufi. 2017. "Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005." *The Review of Financial Studies*, 30(6): 1832–1864.

Morris, Stephen. 2000. "Contagion." *Review of Economic Studies*, 67(1): 57–78.

Munroe, David J., and Laurence Wilse-Samson. 2013. "Foreclosure Contagion: Measurement and Mechanisms." `http://www.columbia.edu/~lhw2110/dm_lw_foreclosure_contagion.pdf`.

Orlean, Andre. 1995. "Bayesian Interactions and Collective Dynamics of Opinion: Herd Behavior and Mimetic Contagion." *Journal of Economic Behavior and Organizations*, 28(2): 257–274.

Patacchini, Eleonora, and Giuseppe Venanzoni. 2014. "Peer Effects in the Demand for Housing Quality." *Journal of Urban Economics*, 83: 6–17.

Petersen, Mitchell. 2009. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *Review of Financial Studies*, 22(1): 435–480.

Piskorski, Tomasz, Amit Seru, and James Witkin. 2015. "Asset Quality Misrepresentation by Financial Intermediaries: Evidence From the RMBS Market." *The Journal of Finance*, 70(6): 2635–2678.

Prasanna, Gai, and Sujit Kapadia. 2010. "Contagion in Financial Networks." *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 466: 2401–2423.

Rothman, Seth J. 2012. *Epidemiology: An Introduction.* Oxford University Press.

Scharfstein, David S., and Jeremy C. Stein. 1990. "Herd Behavior and Investment." *American Economic Review*, 90(3): 465–479.

Scheinkman, José A., and Wei Xiong. 2003. "Overconfidence and Speculative Bubbles." *Journal of Political Economy*, 111(6): 1183–1220.

Shiller, Robert J. 1995. "Conversation, Information, and Herd Behavior." *American Economic Review Papers and Proceedings*, 85(2): 181–185.

Shleifer, Andrei, and Lawrence H. Summers. 1990. "The Noise Trader Approach to Finance." *Journal of Economic Perspectives*, 4(2): 19–33.

Topol, Richard. 1991. "Bubbles and Volatility of Stock Prices: Effect of Mimetic Contagion." *The Economic Journal*, 101(407): 786–800.

Towe, Charles, and Chad Lawley. 2013. "The Contagion Effect of Neighboring Foreclosures." *American Economic Journal: Economic Policy*, 5(2): 313–35.