

Speculative Fever: Investor Contagion in the Housing Bubble

Patrick Bayer*

Kyle Mangum[†]

James W. Roberts[‡]

February 2016

Abstract

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 virtually non-existent. 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 CODES: D40, D84, R30

Keywords: Speculation, Housing Markets, Asset Pricing, Financial Intermediaries, Asset Bubbles, Contagion

*Department of Economics, Duke University and NBER. Contact: patrick.bayer@duke.edu.

[†]Department of Economics, Georgia State University. Contact: kmangum@gsu.edu

[‡]Department of Economics, Duke University and NBER. Contact: 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. We are grateful for Duke University’s financial support. Any errors are our own.

1 Introduction

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 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

¹Similarly, 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.

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

³See, for example, DeLong, Shleifer, Summers, and Waldmann (1990), Scharfstein and Stein (1990), Shleifer and Summers (1990), Topol (1991), Froot, Scharfstein, and Stein (1992), Lux (1995), Lux (1998), Morris (2000), Allen and Gale (2000), Corcos, Eckmann, Malaspinas, Malevergne, and Sornett (2002), Scheinkman and Xiong (2003), Prasanna and Kapadia (2010).

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

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

that of professional, “non-infected” investors as a way of gauging the relative sophistication of these investors.

Four important features of the 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, Lee, Tracy, and van der Klaauw (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 United States; comparatively, accessing comprehensive individual investment data for other financial markets is typically more challenging. Fourth, the geographical 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 estimate 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 investors⁷), it will generally be a difficult task to determine the set of other investors’ portfolios that any given investor has knowledge about.

Our focus is on investors 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

⁶Using a different methodology, Bayer, Geissler, Mangum, and Roberts (2014) estimates a similar jump in purchases by novice investors in the Los Angeles metropolitan area over this same period.

⁷One 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.

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), but obvious candidates abound, including 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 1989-2012. 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 8 percent within the next year and up to 20 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 9 percent and 19 percent over the same horizons, respectively. The magnitudes of both forms of investor contagion change over the course of the housing boom, especially the flipped property channel, the effect of which is greatest in the peak years, 2004-2006.

⁸See, 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, Davis, Greenstone, and Walker (2015).

⁹In 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.”

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. 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. We estimate that the impact of neighborhood-level contagion increased over the course of the boom because (i) the average exposure of homeowners to nearby investment activity rose sharply as real estate speculation increased and expanded 2004-2006 and (ii) the estimated effect of exposure on homeowner investment behavior increased over the course of the boom, peaking in the 2004-2006 period. Our most conservative analysis implies that neighborhood contagion was responsible for 11.2 percent of the speculative real estate investments and 10.4 percent of investors at the peak of the boom.

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 than 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, Ivkovic, Smith, and Weisbenner (2008), or Banerjee, Chandrasekhar, Duflo, and Jackson (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). 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. 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. 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 social learning and social utility mechanisms.¹⁰ 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.¹¹ Second, there are so many obvious ways that learning could take place.

¹⁰An exception is the recent experimental paper, Bursztyn, Ederer, Ferman, and Yuchtman (2014), on peer effects in financial decisions, which finds that both social learning and social utility channels affect behavior.

¹¹Indeed, the Bursztyn, Ederer, Ferman, and Yuchtman (2014) paper finds that the social learning channel is strongest when agents observe the behavior of a relatively sophisticated group.

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 form professional networks (of repair services, inspectors, attorneys, etc.) from more experienced investors. Third, while we find significant effects of neighbors’ investing on one’s own investment decision, we also find that simply living near investment properties that are flipped also induces investor entry. In this latter case, the infected individual can observe the investor’s return on the flipped property from the publication of transactions prices (which were easily available online during this period) regardless of whether they ever interacted directly. On this point, it is further suggestive that we measure the property effect to be strongest during the period of most rapid price appreciation in 2004-2006, when especially large returns would have been routinely observed.¹²

The rest of the paper is organized as follows. Section 2 describes our source data and the construction of the estimation sample in detail. Section 3 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 4 compares the investment performance of influenced investors compared to all other investors in the housing market. Section 5 concludes. The appendices contain a Monte Carlo study of our estimator, described below, and additional figures and tables referenced throughout the paper.

2 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.

2.1 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. Our sample includes the complete universe of housing transactions in the five largest counties in the Los Angeles metropolitan area (Los Angeles, Orange, Riverside, San Bernardino, and Ventura), from 1988 to 2012. For each transaction, the data contain the names of the buyer and seller, the transaction price, the address, the date of sale, and numerous property characteristics

¹²The investor neighbor effect tempers during this period, perhaps because existing investors were becoming less sanguine about the potential for future returns.

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 is a condominium, was subdivided or split into several smaller properties and re-sold, the price of the house was less than \$1¹³ or flagged as not an arms-length transactions, the house sold more than once in a single day, the price or square footage was in the top or bottom one percent of the sample,¹⁴ or there is a potential inconsistency in the data such as the transaction year being earlier than the year the house was built. The remaining sample consists of nearly five million transactions spanning 94 quarters (the data end in mid 2012), with roughly 50 thousand transactions per quarter. We report summary statistics below, after describing additional sample selections.

Because the Dataquick data set contains information on any 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.¹⁵

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. The price index is computed using a standard repeat sales approach.¹⁶ 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. Transactions prices averaged \$187,000 in the 1990s.

¹³A 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.

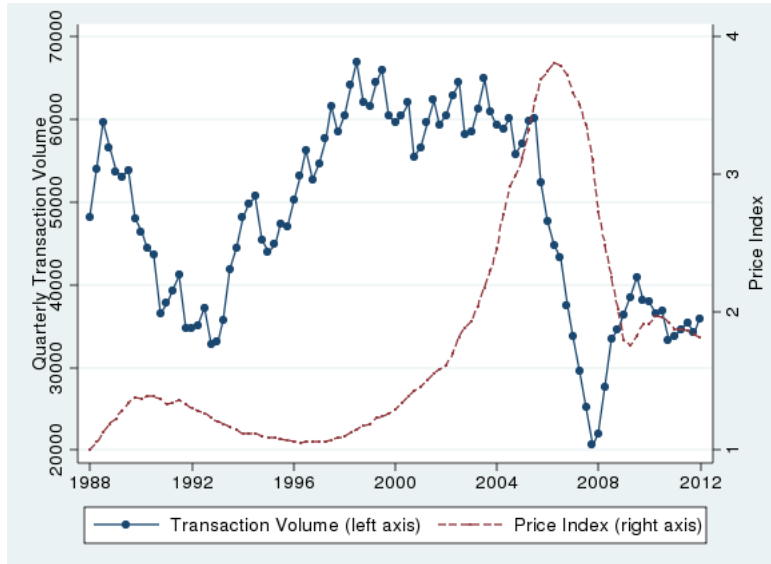
¹⁴These may reflect coding errors.

¹⁵HMDA 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, McMillen, Murphy, and Timmins (forthcoming). 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, McMillen, Murphy, and Timmins (forthcoming).

¹⁶In 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.

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.



NOTES: 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 on log transaction price; 1988Q1 is normalized to 1. See footnote 16.

2.2 Designating Investors and Investment Activity

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, we emphasize our objectives 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 construct a portfolio of properties that each individual owns over the course of our sample—a “profile”—and this requires us to match potentially multiple transactions to particular individuals. Of course, 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).¹⁷ 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.

Next, we need to identify the possible channels through which each non-investor may be influenced to become an investor. The two channels available in these data are whether a neighbor begins investing in the housing market and if a nearby home is bought and sold as an investment property. 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 highlight how the obvious 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 ones.

Our measures of investment activity are derived from buyer/seller names and transaction dates observed in the Dataquick data. We build a profile of each individual’s holdings over time by combining all of the property holding profiles associated with a given name. In the data, buyer and seller names are detailed and typically include middle initials and often the names of a spouse or co-borrower. In our analysis we are interested in individuals entering the investor market, and so we use the names associated with each transaction to exclude purchases by businesses, nonprofits, and various government organizations (“institutions”). We conduct an extensive name cleaning algorithm to flag institutions, separate first, middle, spouse, and surnames, and to standardize punctuation and spacing.¹⁸

To provide a more detailed characterization of the data, Table 1 reports counts of observations by identification of name (post 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

¹⁷While many individuals view their primary residence as “an investment,” we are interested in those taking the next step in investing: buying additional properties as investments.

¹⁸The name cleaning algorithm is available from the authors upon request.

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, thereby attenuating our key parameter estimates. Below we illustrate the robustness of our results to various potential sorts of measurement error. The corresponding results support the natural intuition that misclassification 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 40 transactions. Hence, we focus on personal names that are not excessively common.

Table 1: Transaction, Name, and Investment Property Counts.

Category	1988-2012	
	Transactions	Unique Buyers
All	4,756,715	3,781,998
Institutional Name	512,461	168,037
Common Name	50,052	695
Personal Names	4,194,202	3,613,266
Middle	2,325,049	2,063,358
Spouse	1,575,864	1,471,192
Either Middle or Spouse	2,972,634	2,671,472
Neither Middle nor Spouse	1,221,568	941,794
Non-investment properties	3,616,267	3,350,829
Investment properties	577,935	262,437
Investment Property, Primary Res. ID'ed	356,491	175,306
Investment Property, Primary Res. Not ID'ed	221,444	87,131
Disqualifications:		
Multiple initial properties (i,ii)	133,831	48,646
First property held < 2 years (iii)	61,513	19,391
Insufficient overlap with investments (iv)	26,100	19,094

NOTES: The table shows counts of unique names and of transactions for several categories of names identified in the transaction register data. A “common name” is a non-institutional name that is listed as buyer for over 40 distinct transactions. The Roman numerals under “Disqualifications” correspond to the reasons listed in the main text.

We use the complete profile of property holdings over time for each individual to construct

several important measures for our analysis. We begin by designating the set of individuals that become “investors” at some point during the sample period. In general, any individual that simultaneously holds two or more properties is designated an investor with two 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. In particular, if the individual’s original property is sold within six months of purchasing a second property, or the original primary residence is sold within twelve months, and the second property would be the individual’s only investment in our sample, then the corresponding brief period of multiple holdings does not count towards the definition of an “investor.” 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.

Having defined the set of individuals that become investors during the sample period, we next characterize each individual’s “primary residence” and “investment properties.” Identifying the primary residence is important because it determines the point around which we measure how exposed a homeowner is to nearby investment activity. For individuals that never become investors 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 (i.e., become investors) identically to non-investors, designating their single property holding as their primary residence, with some important exceptions when we suspect that the initial property purchased in the sample period itself may have been an investment property. This would be the case, for example, for a property purchased by an investor from out of state or by someone who had purchased their primary residence in the Los Angeles area prior to the beginning of our sample in 1988. 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 the 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 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.

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

We define all of the remaining overlapping property holdings of investors (other than those designated as primary residences) as “investment properties.” Among investment properties, we also designate whether a property was “flipped,” i.e., sold again after a holding period of less than two years. 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.

Figure 2 reports the time series for three proxies for overall investment activity in the Los Angeles market between 1993-2012 derived from our transaction data set.¹⁹ 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.²⁰ Our overall measure of investment activity also tracks quite well that of Haughwout, Lee, Tracy, and van der Klaauw (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.²¹

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

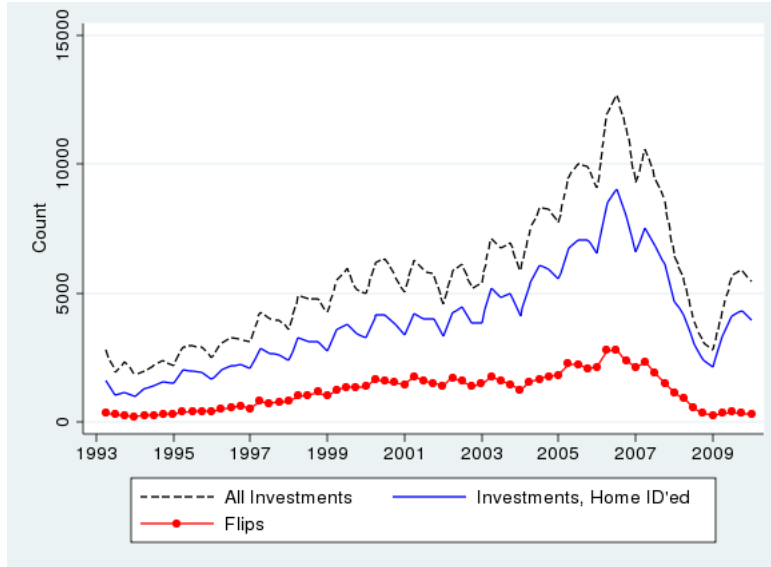
¹⁹As noted above, identification of investors is especially noisy in the first few years of transactions data.

²⁰While 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.

²¹Haughwout, Lee, Tracy, and van der Klaauw (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.

likely to understate the true measure, especially near the beginning of the sample period.²² After growing mechanically in the early years of the transaction data, the counts of active tenures stabilize around 2000, as shown in Figure 8 in Appendix B, further supporting our focus on the period after 2000 when the construction of these measures of investment had stabilized.

Figure 2: Investor Purchasing Activity Over Time.



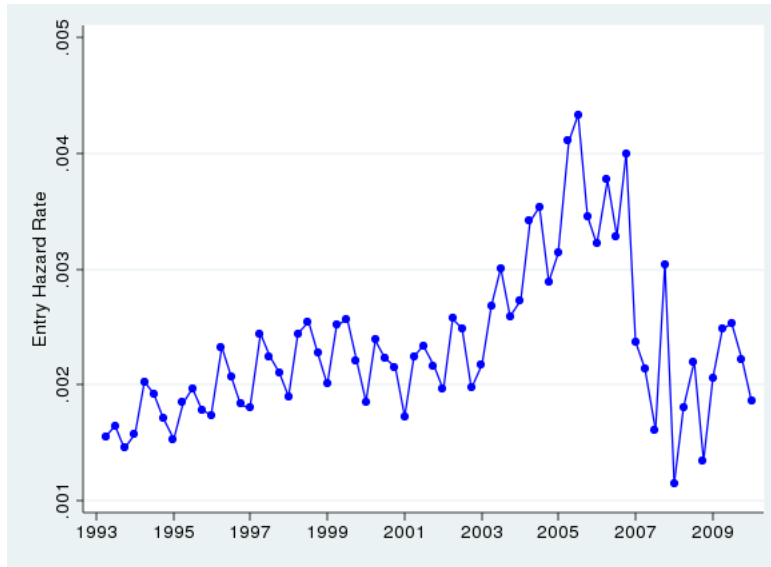
NOTES: The figure plots three quarterly data series that measure investment activity. Definitions are given in the main text. For investment properties, the timing is the quarter of purchase, and for flips it is the quarter of sale.

Also speaking to this issue, 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.

In Table 2 we report summary statistics for transactions by personal names (i.e. not

²²To 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.

Figure 3: Investor Entry Over Time.



NOTES: The figure plots a count of quarterly investor entry - i.e. time of purchase of an individual's first investment property.

institutions or excessively common names) over the entire sample period and in our focus period of 2000-2007. In the main sample period, transaction prices are obviously higher, but the composition of homes (by property attributes and adjusted values) is comparable. Homes in Los Angeles tend to be newer and more expensive than those in many other U.S. cities. The vast majority of buyers take out a mortgage, with an average loan to value ratio of 88 percent in 2000-2007. Homebuyers in the Los Angeles metropolitan area are racially and ethnically diverse with an average income of \$107,000 over the main sample period.²³

Table 3 reports summary statistics for the set of transactions we identify as investments in the period 2000-2007. 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 tend to be slightly older, smaller, and of slightly lower value. The subset of investment properties for which we have identified the location of the investor's primary residence is presented in the middle columns; the statistics are quite similar. The rightmost columns report statistics for flips. Compared to other investment properties, these are slightly older, smaller, and lower value, and are more likely to be purchased in cash.

²³Data from HMDA started becoming available in the early 1990s, so there are some years in which no HMDA match could take place. Match rates improved from the early 1990s to the 2000s.

Table 2: Transaction and Property-level Summary Statistics.

	All Personal Names, 1988-2012			All Personal Names, 2000-2007		
	<i>N</i>	Mean	Std. Dev.	<i>N</i>	Mean	Std. Dev.
Transactions						
Year	4,194,202	1999.49	6.80	1,575,048	2003.17	2.11
Price (\$)	4,194,202	274,819.60	199,120.90	1,575,048	358,556.60	219,566.70
Value (\$ 2000)	4,194,202	216,817.00	140,076.90	1,575,048	207,641.10	119,715.10
Loan Present?	4,194,202	0.89	0.31	1,575,048	0.94	0.23
LTV	3,741,681	0.87	2.53	1,485,399	0.88	0.24
Nonwhite	1,761,752	0.49	0.50	1,005,642	0.53	0.50
Income	1,854,524	101.40	132.83	1,069,242	107.50	131.44
Properties						
Year built	2,056,770	1969.37	21.39	739,298	1971.09	22.46
Sq. ft	2,056,770	1,662.66	646.96	739,298	1,675.39	671.07
No. beds	2,056,770	3.07	0.94	739,298	3.07	0.95
No. baths	2,056,770	2.17	0.77	739,298	2.17	0.78

NOTES: The table shows transaction and property-level summary statistics for data that cover five counties in the Los Angeles area (Los Angeles, Orange, Riverside, San Bernardino, and Ventura). The sample is cleaned as described in main text. Loan to value ratio (LTV) is measured relative to the price paid at the time of initial purchase. Value is the transaction price deflated by a metro-wide price index to year 2000 dollars.

Table 4 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.²⁴ While this upper tail of 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 more novice investors comprising the majority of this distribution.

2.3 Exposure to Neighborhood Investment Activity

Having characterized the primary residence 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. Homeowners remain at-risk until they either (i) purchase an investment property (as defined above) or (ii) sell their primary residence and leave the sample.²⁵ Buying

²⁴Recall that any name with more than 40 purchases is excluded to avoid excessively common names. Institutions are also excluded.

²⁵As 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 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.

Table 3: Transaction and Property-level Summary Statistics for Investment Properties, 2000-2007.

	Investments			Investments, Investor Home ID'ed			Flips		
	<i>N</i>	Mean	Std. Dev.	<i>N</i>	Mean	Std. Dev.	<i>N</i>	Mean	Std. Dev.
Transactions									
Year	256,773	2003.68	2.09	171,498	2003.76	2.09	33,319	2002.88	1.91
Price (\$)	256,773	350,800.90	214,677.70	171,498	357,117.30	212,665.00	33,319	292,332.80	201,745.80
Value (\$ 2000)	256,773	181,785.90	103,323.30	171,498	183,590.00	101,599.40	33,319	170,889.90	106,590.50
Loan Present?	256,773	0.93	0.25	171,498	0.94	0.23	33,319	0.84	0.37
LTV	238,890	0.90	0.19	161,777	0.90	0.18	27,876	0.89	0.19
Properties									
Year built	187,615	1968.32	22.67	153,493	1968.22	22.45	26,805	1963.16	23.59
Sq. ft	187,615	1,547.11	622.00	153,493	1,530.70	610.06	26,805	1,477.65	588.54
No. beds	187,615	2.98	0.96	153,493	2.98	0.95	26,805	2.91	0.92
No. baths	187,615	2.05	0.77	153,493	2.04	0.76	26,805	1.97	0.77

NOTES: The table shows transaction and property-level summary statistics our designated investment properties purchased in the period 2000-2007. The sample is cleaned as described in main text. Loan to value ratio (LTV) is measured relative to the price paid at the time of initial purchase. Value is the transaction price deflated by a metro-wide price index to year 2000 dollars.

Table 4: Summary of Investors' Purchasing Behavior.

	All 1988-2012			Entry 2000-2007		
	All	Primary Res.	ID'ed	All	Primary Res.	ID'ed
<i>N</i>	262,437		175,306	105,723		62,906
Mean	2.20		2.03	1.70		1.53
SD	3.02		2.77	1.43		1.21
Pct. with 1 Purchase	59.46		65.58	63.53		71.62
Pct. with 2 Purchases	21.02		16.92	21.89		17.24
Pct. with 3 Purchases	7.31		6.52	7.17		5.70
Pct. with 4+ Purchases	12.21		10.98	7.41		5.44

NOTES: *N* refers to counts of investors by name; institutions and excessively common names are excluded. Entry 2000-2007 refers to investors who purchased their first investment property during this time period. Statistics exclude investors who entered prior to 2000 but purchased during 2000-2007, and include any properties purchased after 2007 by investors within the 2000-2007 entry cohorts.

a new primary residence keeps the person in the sample as a new at-risk tenure.

We use the location of the primary residence of at-risk homeowners to construct two measures of exposure to nearby investment activity.²⁶ The first is a measure of whether the individual's neighbors are engaged in real estate investment. In particular, we construct counts of the number of neighbors who have entered—i.e., begun purchasing investment properties—in a given period, constructing this measure for various distance bands around the individual's primary residence. Our second measure of neighborhood investment activity is based on investment properties that are flipped within the individual's neighborhood. In particular, we construct counts of the number of investment properties that were sold in a given period following a holding period of less than two years, again at various distance bands.

These measures of exposure to investment activity will form the basis of our research design, functioning as “treatment” and “control” variables depending on the exposure radius. Note that each of these measures varies over both space (some neighborhoods have more investment activity than others) and time (some periods have more investment activity than others).

We close this section by providing summary statistics of the two measures of investment activity matched to at-risk tenures with plausibly identified primary residences for our primary estimation sample in Table 5. After discarding property tenures unlikely to be primary residences as described above, or with investors observed to enter before 2000 and after 2007, we are left with 2,114,646 unique at-risk tenures active sometime during the period of 2000-2007. The typical at-risk is observed for nearly 50 months of the available 96 (with censoring occurring at entry for investors), making nearly 105 million at-risk tenure-month observations. There are 62,906 observed entries, for a monthly hazard rate of six hundredths of one

²⁶We use the Great Circle distance calculated using the properties' latitude and longitude information from the tax assessor file.

percent.

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 5, 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.²⁷ The level of exposure rises mechanically with the expanding radius, and the outer rings are inclusive of the inner ring exposure. 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 (defined here as 0.5 mile radius). This offers intuition about the research design we employ, which asks, of the activity occurring within the wider rings, what fraction is occurring within a hyper-local (e.g. within 0.1 mile) radius? In essence, we will compare the investing behavior for at-risk tenures with very local exposure to those without, conditioning on the amount of exposure in the broader neighborhood.

Table 15 in Appendix B provides the joint distribution of two forms of investment activity at the 0.1 mile ring. While there is some overlap, the table shows these are distinct objects.²⁸

Table 5: Estimation Sample (2000-2007) Summary Statistics.

Frequencies							
At-risk Tenures	Avg Months Observed	At-Risk Tenure-Months	Entries	Entry Hazard Rate (x10,000)			
2,114,646	49.50	104,664,593	62,906	6.0102			
Exposure to Investment Activity Within Past Year							
	Mean	SD	Min	Max	Pct. w/. 0	Pct. w/. 1	Pct. w/. 2+
Investor Neighbors							
w/i. 0.1 mi	0.28	0.62	0	13	78.76	16.82	4.42
w/i. 0.3 mi	1.55	1.69	0	22	31.06	28.60	40.34
w/i. 0.5 mi	3.60	3.02	0	40	10.88	15.32	73.80
Flipped Properties							
w/i. 0.1 mi	0.18	0.52	0	13	85.46	11.86	2.67
w/i. 0.3 mi	1.05	1.46	0	29	46.51	27.55	25.94
w/i. 0.5 mi	2.47	2.71	0	40	22.66	22.53	54.81

NOTES: The table reports summary statistics for outcomes and righthand side variables in the primary estimation sample. 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).

²⁷These statistics are taken over the pooled sample, so they average over spatial and temporal variation.

²⁸This also shows that investors do not necessarily purchase investments on their own blocks, which both instills confidence that primary residence and investment properties are in indeed separately identified, and introduces interesting avenues for future research regarding the spatial pattern of an investor’s purchasing.

3 Estimating Contagion in Speculative Real Estate Investing

3.1 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 for a profit 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 basic 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 (176 yards/160 meters, about the size of a typical city block), while conditioning on activity in wider (e.g. 0.30 or 0.50 mile) bands, as well as dummies for standard neighborhood definitions (ZIP code).

Formally, there are two key identifying assumptions that underlie this approach. The first is that household sorting (or other unobserved predictors of investment activity) does 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 is that neighborhood interactions take place at hyper-local 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 further away. If neighborhood interactions were not stronger at hyper-local geographies, the estimated effect would be zero. Of course economic theory does not define the scale at which such interactions truly take place, and so to the extent that they 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.

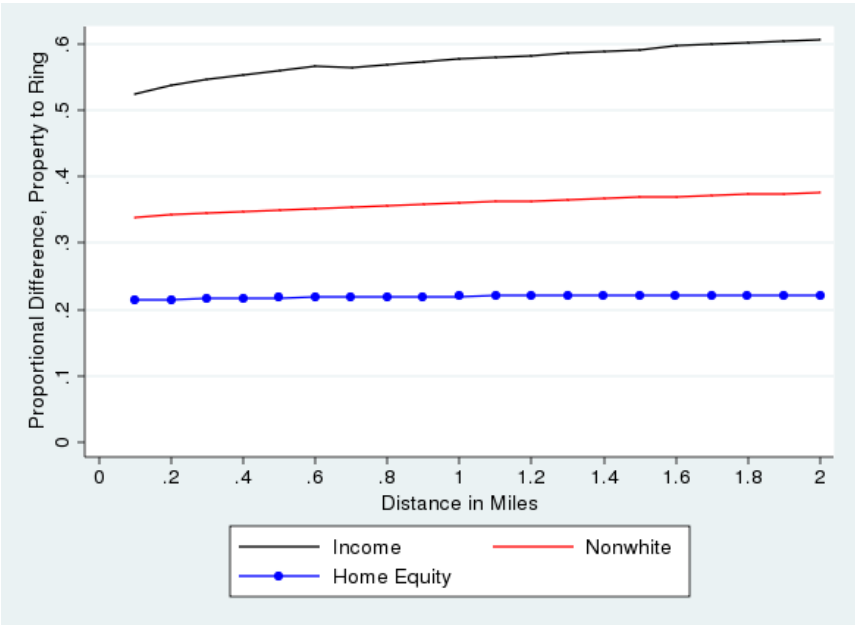
It is worth noting at the outset that these identifying assumptions are unlikely to hold perfectly in the real world. We expect that there will be some block level sorting, however limited, and neighborhood spillovers to decay with distance in a more continuous way, even if they are much stronger among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

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 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, $\bar{x}_{i,(n)d,(n+1)d}$, for $(n = 1, 2, \dots)$.²⁹ We then take the average absolute value of the differences between x_i and $\bar{x}_{i,(n)d,(n+1)d}$ over the sample for 20 ($n = 1, \dots, 20$) bins of 0.1 mile rings ($d = 0.1$ mile). 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. The figure also documents that there is virtually no distinction between immediate neighbors’ initial home equity at time of purchase from that of neighbors further away.

In contrast, Figure 5 illustrates the relatively sharp increase in very local neighborhood investment activity for at-risk homeowners who entered as real estate investors in a given period. In particular, Figure 5 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, (i) entry into real estate investment by immediate neighbors and (ii) a quick sale of an investment property (“flip”) 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.

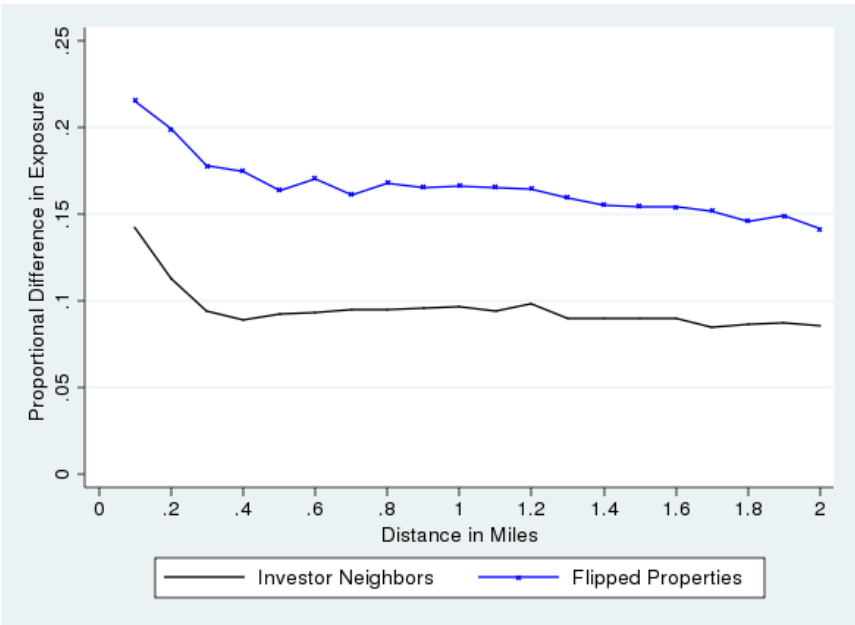
²⁹For example, if X is income, $\bar{x}_{i,(2)0.1,(2+1)0.2}$ is the mean income in an annulus whose inner radius is 0.2 miles and outer radius is 0.3 miles.

Figure 4: Comparability in Homeowner Attributes Over Space.



NOTES: 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.

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



NOTES: 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.

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).

While these figures provide some initial perspective on the assumptions underlying our analysis, in the analysis that follows we also present a number of robustness specifications that provide additional evidence that the key assumptions underlying our research design appear to be plausible.

Finally, we conduct 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 neighborhood activity on an individual's likelihood of becoming an investor. A description and the results of this test are provided in Appendix A.

3.2 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/outer ring research design. 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. We focus on the period 2000-2007, which includes the period of house price appreciation and overall increase investing activity, for the entire greater Los Angeles area.³⁰

In this hazard specification, there is both spatial and temporal variation in the level of housing activity. Thus, 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. As Table 5 showed, entry is an uncommon event, especially when measured at the monthly level. Over the entire period of 2000-2007, the average monthly entry rate was 0.06 percent.

We begin by using a one year lag, so that the righthand side variables are counts of invest-

³⁰There may be interesting heterogeneities in the degree to which types of households or neighborhoods are susceptible to nearby influence, but we leave these for future work.

ment activity occurring within the last year.³¹ The distance bands are defined inclusively, so that any activity in the inner band is also measured in the outer band. Thus, coefficients can be interpreted as the additional impact of the inner band beyond that of being included in the outer band.

Table 6 reports results from our main specifications. It includes results from each type of treatment, nearby neighbor investors and nearby flipped properties, jointly in the same regression.³² Coefficient estimates are followed by hazard ratios, the change in propensity to enter attributable to the explanatory variable(s) 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 mile 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 baseline unexposed hazard of becoming an investor in a given month. 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/outer band research design, adding controls for activity within 0.30 miles and 0.50 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. As discussed above, we take this to be a conservative estimate of the causal impact of one additional investor neighbor or flipped property on investor entry.

Columns 3 and 4, respectively, add year-quarter and ZIP code fixed effects, and Column 5 includes both.³³ 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 mile coefficients fall with the inclusion of both the ZIP code and quarterly dummies specification in column 5, while the 0.10 mile points remain stable. This suggests that ZIP code fixed effects and the 0.50 mile bands function similarly as controls. We adopt the specification reported in column 3, with 0.30 and 0.50 band controls and quarterly dummies, as our preferred specification.

To further illuminate the spatial dimension of the effects documented in Table 6, Table 7

³¹The 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.

³²Results are qualitatively similar, but quantitatively larger, when using separate specifications to measure the effects of each type of exposure.

³³We use the baseline hazard rate from column 2 so that hazard ratios are interpretable across specifications with and without fixed effects.

Table 6: Linear Probability Hazard Regression of Investor Entry on Exposure to Investing Activity.

	1	2	3	4	5
Coefficient Estimates					
Investor Neighbors:					
w/i. 0.10 mi, 1yr lag	0.820 (0.046)***	0.324 (0.052)***	0.328 (0.052)***	0.347 (0.053)***	0.349 (0.053)***
w/i. 0.30 mi, 1yr lag		0.049 (0.027)*	0.052 (0.027)*	0.066 (0.027)**	0.071 (0.027)***
w/i. 0.50 mi, 1yr lag		0.169 (0.015)***	0.126 (0.015)***	0.148 (0.015)***	0.080 (0.016)***
Flipped Properties:					
w/i. 0.10 mi, 1yr lag	1.189 (0.057)***	0.378 (0.065)***	0.377 (0.065)***	0.412 (0.066)***	0.407 (0.065)***
w/i. 0.30 mi, 1yr lag		0.059 (0.033)*	0.058 (0.033)*	0.084 (0.033)**	0.084 (0.033)**
w/i. 0.50 mi, 1yr lag		0.316 (0.018)***	0.271 (0.018)***	0.203 (0.019)***	0.121 (0.019)***
Hazard Rates					
Investor Neighbors:					
w/i. 0.10 mi	0.154 (0.009)***	0.080 (0.013)***	0.080 (0.013)***	0.085 (0.013)***	0.086 (0.013)***
Flipped Properties:					
w/i. 0.10 mi	0.223 (0.011)***	0.093 (0.016)***	0.093 (0.016)***	0.101 (0.016)***	0.100 (0.016)***
Qtrly effects			Y		Y
ZIP effects				Y	Y

NOTES: 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. *** p<0.01, ** p<0.05, * p<0.1, +p<0.2.

reports the results of a set of simple specifications that consider the impact of nearby activity for different distances in space. Table 7 shows how the effect of neighborhood activity in the past year decays with distance. Consistent with the results presented above, the impact of local activity falls to zero by 0.3-0.4 miles, conditional on the outer ring of 0.5.

Because the extent of household sorting falls off so gradually (Figure 4), it is very difficult to rationalize the pattern of results presented in Tables 6 and 7 with a sorting-based explanation. Likewise, the possibility that some unobserved factor was responsible for generating the correlation between both measures of recent neighborhood investment activity and the at-risk homeowner's entry to real estate investing would have to be incredibly special - having an impact confined to only the households and/or properties on a single city block.

Table 7: Spatial Decay of Contagion Effect.

	1	2	3	4
Distance	w/i 0.10 mi	w/i 0.20 mi	w/i 0.30 mi	w/i 0.40 mi
Investor Neighbors:				
Coefficient (w/i. 0.1 mi)	0.269 (0.0594)***	0.093 (0.0442)**	0.053 (0.0381)	-0.075 (0.0344)**
Hazard (w/i. 0.1 mi)	0.066 (0.0146)***	0.023 (0.0108)**	0.013 (0.0093)	-0.018 (0.0084)**
Cumulative Effect	0.066 (0.0146)***	0.089 (0.0135)***	0.102 (0.0131)***	0.084 (0.0128)***
Flipped Properties:				
Coefficient (w/i. 0.1 mi)	0.341 (0.0737)***	0.055 (0.0551)	0.021 (0.0471)	0.022 (0.0422)
Hazard (w/i. 0.1 mi)	0.084 (0.0181)***	0.013 (0.0135)	0.005 (0.0115)	0.006 (0.0103)
Cumulative Effect	0.084 (0.0181)***	0.097 (0.0168)***	0.102 (0.0164)***	0.108 (0.0159)***
Qtrly effects	Y	Y	Y	Y

NOTES: 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. *** p<0.01, ** p<0.05, * p<0.1, +p<0.2.

To illustrate the time-dimension of the effect of neighborhood investing activity, Table 8 shows how the impact of activity within 0.10 miles decays with additional lags of time since the activity took place. The full impact of exposure to an investor neighbor or flipped property is realized over three years, but falls to zero thereafter. Nearby investment activity is positively temporally correlated, so when jointly included, the year-by-year coefficients are smaller than the single year of Table 6, but the cumulative effect is more than twice that

of Table 6. The cumulative effect of a nearby flip or a neighbor investor is about a 19-20 percent increase in the likelihood of entering as an investor in each case. Interestingly, the impact of having a new neighbor investor accumulates more slowly over three years, while the impact of a nearby flip is largely realized within two years.

That the estimated effects fall to zero at either a distance of 0.40 miles or time greater than three years strongly supports the notion that the geographic location of investment activity is not correlated with the underlying propensity of homeowners to become real estate investors. If the correlations were spurious and there was no causal effect of very nearby and recent activity—that is, if investment activity occurred in neighborhoods that contained homeowners especially likely to become real estate investors—we would expect to see a fair amount of lingering positive correlation at wider distance bands and longer lags presented in columns 4 of these tables.

Table 8: Temporal Decay of Contagion Effect.

Years lagged:	1	2	3	4
Investor Neighbors:				
Coefficient (w/i 0.1 mi)	0.204 (0.0497)***	0.160 (0.0533)***	0.202 (0.0596)***	0.037 (0.0673)
Hazard (w/i. 0.1 mi)	0.073 (0.0201)***	0.057 (0.0199)***	0.072 (0.0211)***	0.013 (0.0221)
Cumulative Effect	0.073 (0.0201)***	0.130 (0.0270)***	0.201 (0.0323)***	0.215 (0.0364)***
Flipped Properties:				
Coefficient (w/i 0.1 mi)	0.233 (0.0592)***	0.191 (0.0636)***	0.124 (0.0726)*	-0.012 (0.0845)
Hazard (w/i. 0.1 mi)	0.083 (0.0250)***	0.068 (0.0252)***	0.044 (0.0265)*	-0.004 (0.0297)
Cumulative Effect	0.083 (0.0250)***	0.151 (0.0338)***	0.194 (0.0400)***	0.190 (0.0464)***
Qtrly effects	Y	Y	Y	Y

NOTES: 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. *** p<0.01, ** p<0.05, * p<0.1, +p<0.2.

Table 9 examines how the effect size changes over time by estimating the hazard regressions separately for three, three-year intervals. The hazard ratios are also computed separately by interval, since the baseline entry rate changes over time. The effect of nearby flipped properties (lower panel) within 0.10 is much stronger near the peak of the boom in 2004-2006, which was also when exposure was highest (Figure 2). In contrast, the impact of

neighbor investors (upper panel) is largest at the beginning of the time period in 1998-2000 and declines as the boom nears the peak. While the impact of neighbor investors remains positive throughout the entire boom, the effect is roughly half the size in 2004-2006, compared with 1998-2000.

Taken together, these results are consistent with the notion that potential novice investors may have received very different signals from nearby flips versus investor neighbors as the market reached its peak in 2004-2006. Essentially any property that was sold after a relatively short holding period between 2004-2006 would have resulted in a massive return for the investor and, thus, might have been especially attractive as a signal of potential gains for new investors. In contrast, (at least a subset of) experienced investors may have become less sanguine about the potential for large returns as the market reached its peak and became more tempered in encouraging their novice neighbors to get into real estate investing. These results are consistent with the notion that a learning mechanism is in play.

Table 9: Linear Probability Hazard Regression Results, Triennially, 1998-2006.

	1 (1998-2000)	2 (2001-2003)	3 (2004-2006)
At-Risk Tenures	34,553,535	39,578,719	40,972,704
Entries	18,858	21,615	29,772
Entry Rate ($\times 10,000$)	5.46	5.46	7.27
Investor Neighbors:			
Coefficient (w/i. 0.1 mi)	0.466 (0.0932)***	0.357 (0.0832)***	0.264 (0.0835)***
Hazard (w/i. 0.1 mi)	0.107 (0.0215)***	0.080 (0.0186)***	0.058 (0.0183)***
Flipped Properties:			
Coefficient (w/i. 0.1 mi)	0.250 (0.1219)**	0.161 (0.1003)	0.568 (0.1024)***
Hazard (w/i. 0.1 mi)	0.058 (0.0280)**	0.036 (0.0224)	0.125 (0.0226)***
Qtrly effects	Y	Y	Y

NOTES: 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, + $p < 0.2$.

3.3 Tests of Research Design and Robustness of Main Results

Table 10 presents a number of additional robustness checks related to sample selection. (Column 0 gives the baseline results from Table 6 above.) The first two columns check the

sensitivity of the baseline results to our ability to identify at-risk tenures in the data. Section 2 described how we inferred whether the purchaser of the home was at-risk (i.e. used the home as their primary residence).³⁴ The data contain two other sources of information, though also imperfect, on whether the property was owner-occupied. First, the HMDA data match includes a flag for whether the loan application was for an owner-occupied home. Second, the assessor data match includes information on the owner’s home mailing address; matching this to the property address gives another indicator for whether the home is considered owner-occupied.³⁵ The results in columns 1 and 2 imply that limiting our analysis to these tenures that meet these more stringent definition of an at-risk homeowner does not materially affect our results; the effect sizes are very similar to the baseline estimates reported in the first row of the table.

The next three columns, 3-5, consider the name matching algorithm used to infer investments. Column 3 drops any investor whose properties span an area wider than 50 miles, under the suspicion that these may be two different individuals with the same (relatively uncommon) name. The loss of observations is small, indicating this is of low incidence, and the effect sizes are quite similar. Column 4 uses only names with middle names/initials and/or spouses listed, since these are less likely to be duplicated. The effect size is slightly smaller for flips, but larger for investor neighbors, especially as measured by the hazard ratio. Column 5 drops any names that are combinations of common names, defined as both first and last being in the top 20 percent of names observed in the data. Recall that we have already excluded any name with more than 40 properties attached, which removed common names like John Smith and Jose Lopez. This is an additional flag for a name like Michael Thompson, where Michael and Thompson are common, but Michael Thompson was not so common as to already be removed.³⁶ This drops an additional 2 million monthly observations, with virtually no effect on the results.

Finally, we use the neighborhood similarity measures constructed to make Figure 4 to classify at-risk tenures as more or less similar to their neighbors. We standardize the attributes and add the (absolute value of) the deviations between the 0.1 and 0.1 to 0.3 mile rings to create a “similarity index.”³⁷ We then repeat the analysis for the 75 percent of

³⁴Note that we also must observe the investor’s primary residence to include him in the spatial match of the investor neighbor righthand side variable. These sample selection checks refer to inclusion of the at-risk tenures (lefthand side variable).

³⁵Note that because the assessor data is overwritten each year by Dataquick, it reflects information from 2011. As a result, this measure of owner-occupancy can only be used for tenures that persist into the 2011 assessment year, limiting the number of observations for which this flag is useful primarily to those late in the sample.

³⁶The commonality of first (and last) names was calculated ignoring the presence of middle names/initials or spouse names.

³⁷The results are quite similar when using 0.1-0.5 mile differences, and when weighting the index by the

tenures that are most similar to their neighbors. The effect on flips is negligible, and the effect on the investor neighbors is slightly reduced,³⁸ suggesting that even in areas where there is the least sorting, we find very similar results to those from our baseline analysis.

Lastly, we investigate the robustness of our analysis to alternative binary choice models and functions of exposure to investment activity. Our main results use a linear probability model with counts as the explanatory variables. This is for computational reasons, ease of interpretation, and to avoid an incidental parameter problem when including large numbers of fixed effects. We check the robustness to functional form in Table 16 in Appendix B. The table reports results from a linear probability model with an indicator variable for the inner ring, as well as those from alternative nonlinear binary response models (probit and logit). The results are all quite similar to those from the baseline specification.

While the point estimates move up and down slightly across the specifications reported in Tables 10 and 16, taken as a whole, they strongly support the robustness of our baseline results to a number of potential measurement and specification issues.

3.4 The Impact of Investor Contagion

The regressions above measure whether there is a statistically meaningful effect of exposure to an investor neighbor or flipped property on entry probability. We next use these point estimates to calculate the magnitude of the effect of the contagion mechanism on total investor entry and purchase volume. The aggregate effect depends on (1) the effect on propensity measured in the regressions, and (2) the degree of exposure to each of the treatments. From Table 5 above, we see that in the years 2000-2007, typical exposure to investor neighbors was 0.28 within 0.1 miles, and 0.18 for exposure to flipped properties, although exposure rates vary over time (Figures 2 and 3). Therefore, we calculate the additional entry propensity for each month and report averages over several ranges of years. It is important to note that the following is purely an accounting exercise at observed volumes and prices, and we do not attempt to counterfactually predict real estate prices or buy/sell behavior at alternative prices or number of investors.

We can write the change in entry propensity as

$$\Delta r_{dt} = (\rho_{d,T}^I X_t^I + \rho_{d,T}^F X_t^F) \tag{1}$$

attributes' power in predicting investor entry.

³⁸The reduction appears to be the result of changing the set of neighborhoods included in the sample rather than a violation of the identifying assumptions. When limiting the sample to ZIP codes with more inner/outer similarity, but including the more dissimilarly situated at-risk tenures themselves, the estimates are similar to column 6.

Table 10: Robustness of Baseline Results.

	0 Baseline	1 Owner Occupancy	2 HMDA or Assessor File	3 HMDA File	4 Name Match	5 Narrow Purchase Area	6 Detailed Names Drop Common Names	7 Similarity Index
Investor Neighbors Coefficient (w/i. 0.1 mi)	0.328 (0.052)***	0.312 (0.063)***	0.359 (0.078)***	0.336 (0.052)***	0.357 (0.051)***	0.324 (0.053)***	0.217 (0.054)***	
Hazard (w/i. 0.1 mi)	0.080 (0.013)***	0.081 (0.016)***	0.085 (0.018)***	0.085 (0.013)***	0.109 (0.016)***	0.081 (0.013)***	0.056 (0.016)***	
Flipped Properties Coefficient (w/i. 0.1 mi)	0.377 (0.065)***	0.425 (0.079)***	0.421 (0.096)***	0.362 (0.064)***	0.287 (0.063)***	0.382 (0.066)***	0.335 (0.064)***	
Hazard (w/i. 0.1 mi)	0.093 (0.016)***	0.110 (0.021)***	0.099 (0.023)***	0.092 (0.016)***	0.087 (0.019)***	0.096 (0.017)***	0.086 (0.02)***	
Nonentries	104,601,687	73,814,177	53,775,098	104,535,809	85,636,559	102,360,145	75,037,810	
Entries	62,906	43,485	34,890	60,831	37,547	60,859	45,533	
Raw Entry rate ($\times 10,000$)	6.014	5.891	6.488	5.819	4.384	5.946	6.068	
N	104,664,593	73,857,662	53,809,988	104,596,640	85,674,106	102,421,004	75,083,343	

NOTES: The table presents various robustness checks for our main results. 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, + $p < 0.2$.

where X_t^I and X_t^F indicate exposure to an investor neighbor or a flipped property, respectively, at time t . The increase in hazard rate as reported in the regression tables above is represented by $\rho_{d,T}^I$ (investor neighbors) and $\rho_{d,T}^F$ (flipped properties); these are indexed by the distance radius admitted, d , and the effect length measure, T . We present the results for the full three-year impact horizon, using only the innermost ring (0.1 mile) effect (Table 8).

The results are reported in Table 11. The first row reports the simple monthly average increase in the likelihood of investor entry. Using the innermost ring of 0.1 mile, the typical Los Angeles area at-risk tenure is 8.1 and 10.4 percent more likely for the periods 2001-03 and 2004-2006, respectively, to become an investor because of the presence of the neighborhood contagion effect.³⁹

The second row of Table 11 shows the cumulative impact of the contagion mechanism on the volume of investor purchases, including any additional investment beyond the purchase made at entry. About one-third of investors entering in the years 2001-2006 purchase more than one property, and of these investors, the mean number of investment properties purchased is 2.9.⁴⁰ The calculation of contagion’s impact on purchases is determined by the effect on cumulative entry and how likely these active investors are to purchase additional properties after their first. We write the impact on volume at time t as

$$\Delta v_t = \Delta r_t + \sum_s^{t-1} (v_s^{>1} \times \Delta r_s) \quad (2)$$

where $v_s^{>1}$ represents the additional purchases by active investors. We find that neighborhood contagion increased the number of investment purchases by 10 percent over the period 2001-2006.

Aggregate effect size depend critically on exposure rates, yet the exposure rates reported above account only for investors we have been able to spatially match, i.e., those with their primary residence identified. However, we know these are not the only investors operating in the Los Angeles market. If we were to scale our estimates to include investors for whom we have not confidently identified a primary residence, the estimated average increase to entry would rise to 13.7 percent and the increase in purchases would rise to 14.7 percent in the 2004-2006 period.

Another way to measure the magnitude of the contagion effect is through an epidemiological concept known as the “basic reproduction rate,” the number of new cases that an

³⁹We have also calculated the cumulative increase in the size of the investor pool over the reported time periods. This is in principle different from the likelihood-of-entry calculation, because it is weighted by the number of at-risk tenures present each month. However, since the count of at-risk tenures is very stable between 2001-2006 (as shown in Figure 8), the magnitudes are quite similar to the first row of Table 11.

⁴⁰Recall that our designation of investors limited the number of investment properties to a maximum of 40.

infected case will create.⁴¹ The reproduction rate depends on (1) the transmission rate, (2) the contact rate, and (3) the incubation period. The transmission rate is our point estimate from Table 8 with the associated incubation period of 36 months. There are two forms of transmission—investor neighbors and flipped properties—and we need to scale the latter transmission rate by the probability that an investment property is flipped (sold within 24 months). The contact rate is simply the number of households within a distance ring d ; that is, a measure of property densities. The density varies spatially and will therefore be different for each specific case, but for simplicity we use the simple average to find the typical magnitude. Together, the average number of new investors one entry creates is measured as

$$R = \sum_{t=1}^T [\beta_{d,T}^{Inv} + \beta_{d,T}^{Flip} \times v_t \times Pr(sell_t^{t-s \leq 24})] \times C_d \quad (3)$$

where $\beta_{d,T}^{Inv}, \beta_{d,T}^{Flip}$ are our point estimates at distance d and horizon (incubation period) T . v_t is the volume of purchases made by the active investor pool, and $Pr(sell_t^{t-s \leq 24})$ is the probability that an investment property purchased at time s is sold at time t within 24 months of s , as 24 months or less triggers the transmission at rate $\beta_{d,T}^{Flip}$. The sale propensity after holding for t months varies over time with the price cycle, but for simplicity we will use the average sale probabilities over the time periods reported in the table. (For investors entering 2001-2006, the probability that an investment property is sold within two years is about 0.25.) Finally, C_d is the contact rate, the number of at-risk tenures within distance d . The typical property in the greater Los Angeles market is exposed to 54.7 other properties within 0.1 mile.

The third row of Table 11 reports the results from this calculation, revealing that an entering investor produces 0.15 new investors through contagion, combining the impact on his neighbors and the impact of his purchases on the surrounding at-risk tenures. Using the inverse of this statistic, this means that about every six investors will generate a new investor purely through neighborhood contagion.

4 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 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

⁴¹See for instance, chapter 6 of Rothman (2012).

Table 11: Magnitude of Effect of Contagion on Investor Entry and Purchases.

Measure	Definition	2001-2003	2004-2006	2001-2006
Average Increase in Entry Probability (%)	$\sum_{I,F} \Delta \text{Entry Propensity}_{I,F} \times \text{Rate of Exposure to } (I, F)$	8.12	10.40	9.26
Additional Investment Purchases (%)	$\Delta \text{Stock of Investors} \times \text{Purchase Rates}$	8.76	11.20	9.96
Multiplier Effect of One Investor	$(\Delta \text{Entry Propensity} + \Delta \text{Entry Propensity} \times \text{Flip Probability}) \times \text{Contact Rate}$	0.16	0.15	0.16

NOTES: 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 8.

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. To isolate the performance of those drawn into real estate investing by their immediate neighbors, we follow a similar research design to that used in Section 3 above, inferring causal effects by estimating the impact of hyper-local neighborhood investment (within 0.10 miles) while controlling for the same measures of investment at slightly larger distance bands (0.50 miles).

Accurately measuring the impact of being drawn into real estate investing by neighborhood investment activity requires one additional adjustment relative to the specifications reported in Section 3. In essence, when analyzing the impact of neighborhood influence on investor performance, we must account for the fact that the exposure to nearby investment activity only raises the likelihood of entry fractionally: many 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 estimate 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 a comparable specification in Section 3. This adjustment is analogous to accounting for the impact of the instrument in the first stage of a two stage least squares regression.

Finally, all of the analysis reported in this section of the paper conditions on the time of investor entry to allow for different cohorts to have varying levels of sophistication.⁴² In

⁴²As noted in the discussion of Table 6, conditioning on the outer ring exposure already accounts for most

this way, all estimates are implicitly based on the returns of investors drawn into investing through the neighborhood contagion mechanisms relative to all other investors that entered the market in the same period.

4.1 Return on Re-Sold Properties

We begin our analysis of investor performance by examining various components of the investor's return for properties that were re-sold during the sample period in Table 12. Subsequent tables will also examine the likelihood that investors continued to hold properties through the peak of the housing boom and whether they eventually defaulted.

The first column of Table 12 reports results for the overall returns for homes purchased by real estate investors between 2000-2006. The results imply that, relative to other real estate entrants, investors that were drawn into the market as a direct result of neighborhood investment activity do markedly worse. In particular, those drawn in by exposure to investor neighbors realize returns that are 16 percentage points lower than the typical investor, while those drawn in by observing a nearby flipped property experience total returns that are 9 percentage points lower.⁴³

To further understand the basis for these lower returns, columns 2 and 3 of Table 12 report results for two specific elements of the total return: the amount that investors pay when they buy the property relative to expected market price and the amount they receive when they sell the property (again relative to market). Note that this conditions on market timing and measures performance stemming from the transactions themselves. These results indicate that investors drawn into the market by both types of neighborhood investment activity perform worse than other investors when both buying and selling properties. Homeowners drawn into investing by observing nearby flipped by properties pay an especially high price at the time of purchase - 6.6 percent more than other investors, while the comparable number for those drawn in by neighbor investors is 4.1 percent. Investors drawn in through both neighborhood channels sell their properties for 2-4 percent less than other investors.

By comparing the results from column 1 to those shown in columns 2 and 3 it is also possible to infer the relative market return of investors drawn in by neighborhood activity compared to other investors, which is a function of both the timing and location of their purchases. The calculation of this remaining portion of the total return that is not attributable to buying relatively high and selling relatively low implies that investors drawn in by either form of neighborhood activity earn about an 8 percent lower market rate of return on their

of the time variation anyway.

⁴³On an annualized basis, the effects sizes are returns lower by 7 and 4 percentage points, respectively, for investor neighbor and flip influence.

holdings relative to other new investors.

It is interesting to note that these effects seem confined to the boom and bust period of the 2000s. We do not find materially different outcomes for influenced and uninfluenced investors in the 1990s, as Table 17 in Appendix B shows.

Table 12: Returns on Properties Purchased in 2000-2006 With Sale Observed.

	1	2	3
	Total Returns $\frac{p_{sale} - p_{buy}}{p_{buy}}$	Buyer Discount $\ln(p_{buy}) - \ln(\hat{p}_{buy})$	Seller Premium $\ln(p_{sell}) - \ln(\hat{p}_{sell})$
<i>N</i>	76,821	76,821	76,821
Mean	0.1492	0.0121	-0.0048
SD	0.5939	0.1511	0.1860
Investor Neighbor <i>w/i</i> 0.1mi	-0.1620 (0.0592)***	0.0412 (0.0164)**	-0.0425 (0.0201)**
Flip <i>w/i</i> 0.1mi	-0.0868 (0.0497)*	0.0663 (0.0137)***	-0.0241 (0.0169)+
Controls:			
Time of Entry	Y	Y	Y
Outer Ring (<i>w/i</i> 0.5 mi) Exposure	Y	Y	Y

NOTES: The data include investment purchases made 2000-2006 with observed sale any time before the end of the data in mid-2012. The sample is trimmed of outliers with returns larger than 200 percent in absolute value. \hat{p} is expected market price predicted using county-specific price indices. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, + $p < 0.2$.

4.2 Holding Properties Past the Peak

The results presented in Table 12 reveal that investors drawn into the market through the neighborhood contagion mechanisms perform significantly worse than other investors across each dimension of buying and selling for properties that were re-sold during the sample period. But conditioning on the set of properties that were re-sold potentially masks another important part of performance: the propensity to re-sell properties in time to avoid the severe drop in prices in the subsequent housing bust.

To examine differences in selling behavior just before the market peak, Table 13 reports a series of specifications that examine the extent to which investors continued to hold the properties that were purchased in the pre-peak years of 2004-2006 before the market collapsed.

The specifications reported in Table 13 follow the same structure as those reported in Table 12 but consider several different dependent variables. In particular, columns 1-3 examine whether investors held the property through 2007, 2008, and 2009, respectively, while column 4 considers whether the investor held the property past the point where the housing price index for the corresponding county fell below the “break-even” point at which the investor

had purchased the property.⁴⁴ The table reports the marginal effect for each specification, implying that those investors drawn into the market in 2004-2006 by their neighbors were 14.4 percentage points more likely to continue to hold their investments through 2009, while those drawn in by observing nearby flipped properties were almost 13 percentage points more likely to hold their properties through 2008 and 2009. The magnitude of these effects suggest that the investors pulled into the market through the neighborhood contagion mechanism were far less skilled in anticipating the market peak than other investors that entered at the same time.

Table 13: Likelihood of Holding a Property Past the Peak in Prices.

	1	2	3	4
	2007+	2008+	Held Past 2009+	Breakeven Qtr
Frequency	0.754	0.674	0.594	0.640
Investor Neighbor w/i. 0.1 mi Marg. Effect	0.099 (0.0495)**	0.110 (0.0537)**	0.144 (0.0562)**	0.159 (0.0621)**
Flipped Property w/i. 0.1mi Marg. Effect	0.103 (0.0411)**	0.139 (0.0446)***	0.138 (0.0465)***	0.124 (0.0515)**
Controls:				
Time of Entry	Y	Y	Y	Y
Outer Ring (w/i. 0.5mi) Exposure	Y	Y	Y	Y
N	63,345	63,345	63,345	50,121

NOTES: The sample is limited to investment properties purchased 2004 to 2006. For column 4 (“Breakeven Qtr”), the sample is limited to investment purchases before the peak in prices, i.e. 2004-2006:Q1. *** p<0.01, ** p<0.05, * p<0.1, +p<0.2.

4.3 Equity Stake and Default Behavior

The results presented in Table 13 clearly show that the amateurs drawn into the market through neighborhood contagion were less able than other investors to foresee the market peak in time to sell their holdings. Yet if these investors had a limited equity stake in their properties (i.e., had put very little down on their mortgage at the time of purchase), the resulting losses from the severe decline in housing prices in the bust may have fallen on the lenders that had initially extended them credit (or, more accurately, the eventual holders of

⁴⁴All five Los Angeles area counties peaked in Q2:2006, though the rates of ascent before, and descent after, varied slightly between them.

these mortgages).⁴⁵

Table 14 presents the results of three regressions that relate the relative size of the initial equity stake and the likelihood of foreclosure for investors drawn into the market by neighborhood investment activity versus other investors. The analyses include controls by purchase year by county to allow for varying access to credit over time and space.⁴⁶ Column 1 reports the results of a regression of downpayment percentage on our usual measures of neighborhood investment activity, implying that those influenced by neighborhood investment activity had down payments that were five percentage points lower than typical investors. (From the mean transaction prices for investment properties in Table 3, five percent is \$17,500 less equity, on average.) Similarly, column 2 reports the marginal effects of a regression of whether investors purchased homes with very low levels of equity (less than 5 percent down). These results imply that investors associated with the neighborhood contagion mechanisms were eight to ten percentage points more likely to be very-low-equity investors and, thus, had relatively little of their own money at stake.

The final column of Table 14 examines the likelihood of foreclosure directly, reporting a comparable set of marginal effects to those reported throughout Section 4. We measure foreclosure in the data by identifying instances in which a property was re-sold with an institutional seller name instead of the initial buyer's name (indicating, e.g., a bank took control of the property). The results show that investors drawn in by their investor neighbors or drawn in by observing nearby flipped properties are indeed 4 percentage points more likely to default on their investment properties (though the investor effect is towards the boundaries of standard statistical significance with the extensive controls we include). These represent a rather large effect relative to our measure of the overall default rate of 8.7 percent.

The results of Table 14 provide a bit of a silver-lining for investors drawn in through neighborhood contagion relative to other real estate investors, suggesting that they risked less of their own money and, in fact, defaulted at a higher rate than other investors, thereby sharing some of the downside of their investments with their mortgage holders.

⁴⁵It is important to note that regardless of their initial equity stake, investors would have had a strong incentive to sell before the market peak if they had been able to foresee it. Selling before the collapse in prices would have preserved not only their initial down payment but any of the substantial market appreciation that had occurred just prior to the peak and allowed them to avoid any of the costs of default, such as a worse credit scores.

⁴⁶We retain the outer ring exposure controls, which again soak up broad neighborhood variation.

Table 14: Equity and Foreclosure Outcomes.

	1 Pct Equity at Purchase OLS	2 I[Low Equity] Probit	3 Foreclose Probit
Mean	0.164	0.395	0.087
Investor Neighbor w/i. 0.1 mi Marg. Effect	-0.053 (0.0182)***	0.098 (0.0366)***	0.039 (0.0266)+
Flipped Property w/i. 0.1mi Marg. Effect	-0.054 (0.0153)***	0.086 (0.0308)***	0.043 (0.0222)*
Controls:			
Time of Entry	Y	Y	Y
Purchase year by county	Y	Y	Y
Outer Ring (w/i. 0.5mi) Exposure	Y	Y	Y
<i>N</i>	157,889	157,889	157,889

NOTES: The data include investment purchases made 2000-2006. *** p<0.01, ** p<0.05, * p<0.1, +p<0.2.

5 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 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. 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, and in this paper, neither can we. 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 that simply living near to homes that are flipped also influences investment behavior, especially during the periods of most rapid price growth. 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, F., AND D. GALE (2000): "Financial Contagion," *Journal of Political Economy*, 108(1), 1–33.
- ANENBERG, E., AND E. KUNG (2014): "Estimates of the Size and Source of Price Declines Due to Nearby Foreclosures," *American Economic Review*, 104(8), 2527–2551.
- BANERJEE, A., A. G. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2013): "The Diffusion of Microfinance," *Science Magazine*, 341(6144).
- BANERJEE, A. V. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107(3), 797–817.
- BASU, R. (2002): "Financial Contagion and Investor Learning: An Empirical Investigation," *IMF Working Paper*.
- BAYER, P., C. GEISSLER, K. MANGUM, AND J. W. ROBERTS (2014): "Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market," *Working Paper*.
- BAYER, P., R. McMILLEN, A. MURPHY, AND C. TIMMINS (forthcoming): "A Dynamic Model of Demand for Houses and Neighborhoods," *Econometrica*.

- BAYER, P., S. L. ROSS, AND G. 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. D. (1994): “A Theory of Conformity,” *Journal of Political Economy*, 102(5), 841–877.
- BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): “A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades,” *Journal of Political Economy*, 100(5), 992–1026.
- BROWN, J. R., Z. IVKOVIC, P. A. SMITH, AND S. WEISBENNER (2008): “Neighbors Matter: Causal Community Effects and Stock Market Participation,” *Journal of Finance*, 63(3), 1509–1531.
- BURNSIDE, C., M. EICHENBAUM, AND S. REBELO (forthcoming): “Understanding Booms and Busts in Housing Markets,” *Journal of Political Economy*.
- BURSZTYN, L., F. EDERER, B. FERMAN, AND N. YUCHTMAN (2014): “Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions,” *Econometrica*, 82(4), 1273–1301.
- CALVO, G., AND E. MENDOZA (1996): “Mexico’s Balance-of-Payments Crisis: a Chronicle of a Death Foretold,” *Journal of International Economics*, 41(3), 235–264.
- CAMPBELL, J., S. GIGLIO, AND P. PATHAK (2011): “Forced Sales and House Prices,” *American Economic Review*, 101(5), 2108–2131.
- CHAMLEY, C. (2004): *Rational Herds: Economic Models of Social Learning*. Cambridge University Press.
- CHARI, V., AND P. KEHOE (2003): “Hot Money,” *Journal of Political Economy*, 111(6), 1262–1292.
- CORCOS, A., J.-P. ECKMANN, A. MALASPINAS, Y. MALEVERGNE, AND D. SORNETT (2002): “Imitation and Contrarian Behaviour: Hyperbolic Bubbles, Crashes and Chaos,” *Quantitative Finance*, 2(4), 264–281.
- CURRIE, J., L. DAVIS, M. GREENSTONE, AND R. WALKER (2015): “Environmental Health Risks and Housing Values: Evidence from 1600 Toxic Plan Openings and Closings,” *American Economic Review*, 105(2), 678–609.

- CURRIE, J., M. GREENSTONE, AND E. MORETTI (2011): “Superfund Cleanups and Infant Health,” *American Economic Review*, 101(3), 435–441.
- DELONG, J. B., A. SHLEIFER, L. H. SUMMERS, AND R. J. WALDMANN (1990): “Positive Feedback Investment Strategies and Destabilizing Rational Speculation,” *Journal of Finance*, 45(2), 379–395.
- DUFLO, E., AND E. SAEZ (2002): “Participation and Investment Decisions in a Retirement Plan: the Influence of Colleagues’ Choices,” *Journal of Public Economics*, 85(1), 121–148.
- ELLISON, G., AND D. FUDENBERG (1993): “Rules of Thumb for Social Learning,” *Journal of Political Economy*, 101(4), 612–643.
- FROOT, K. A., D. S. SCHARFSTEIN, AND J. C. STEIN (1992): “Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation,” *Journal of Finance*, 47(4), 1461–1484.
- HAUGHWOUT, A., D. LEE, J. TRACY, AND W. VAN DER KLAAUW (2011): “Real Estate Investors, the Leverage Cycle, and the Housing Market Crisis,” Federal Reserve Bank of New York Working Paper.
- HONG, H., J. D. KUBIK, AND J. C. STEIN (2004): “Social Interaction and Stock-Market Participation,” *Journal of Finance*, 59(1), 137–163.
- JACKSON, M. O. (2010): *Social and Economic Networks*. Princeton University Press.
- KINDELBERGER, C. P. (1978): *Manias, Panics and Crashes: A History of Financial Crises*. John Wiley & Sons, Inc., 2005 edn.
- KIRMAN, A. (1993): “Ants, Rationality, and Recruitment,” *Quarterly Journal of Economics*, 108(1), 137–156.
- LINDEN, L., AND J. ROCKOFF (2008): “Estimates of the Impact of Crime Risk on Property Values from Megan’s Laws,” *American Economic Review*, 98(3), 1103–1127.
- LUX, T. (1995): “Herd Behaviour, Bubbles and Crashes,” *The Economic Journal*, 105(431), 881–896.
- (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, C. (1841): *Extraordinary Popular Delusions and the Madness of Crowds*. Richard Bentley, 1932 edn.
- MORRIS, S. (2000): “Contagion,” *Review of Economic Studies*, 67(1), 57–78.
- ORLEAN, A. (1995): “Bayesian Interactions and Collective Dynamics of Opinion: Herd Behavior and Mimetic Contagion,” *Journal of Economic Behavior and Organizations*, 28(2), 257–274.
- PRASANNA, G., AND S. KAPADIA (2010): “Contagion in Financial Networks,” *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 466, 2401–2423.
- ROTHMAN, S. J. (2012): *Epidemiology: An Introduction*. Oxford University Press.
- SCHARFSTEIN, D. S., AND J. C. STEIN (1990): “Herd Behavior and Investment,” *American Economic Review*, 90(3), 465–479.
- SCHEINKMAN, J. A., AND W. XIONG (2003): “Overconfidence and Speculative Bubbles,” *Journal of Political Economy*, 111(6), 1183–1220.
- SHILLER, R. J. (1995): “Conversation, Information, and Herd Behavior,” *American Economic Review Papers and Proceedings*, 85(2), 181–185.
- SHLEIFER, A., AND L. H. SUMMERS (1990): “The Noise Trader Approach to Finance,” *Journal of Economic Perspectives*, 4(2), 19–33.
- TOPOL, R. (1991): “Bubbles and Volatility of Stock Prices: Effect of Mimetic Contagion,” *The Economic Journal*, 101(407), 786–800.

A Placebo Tests

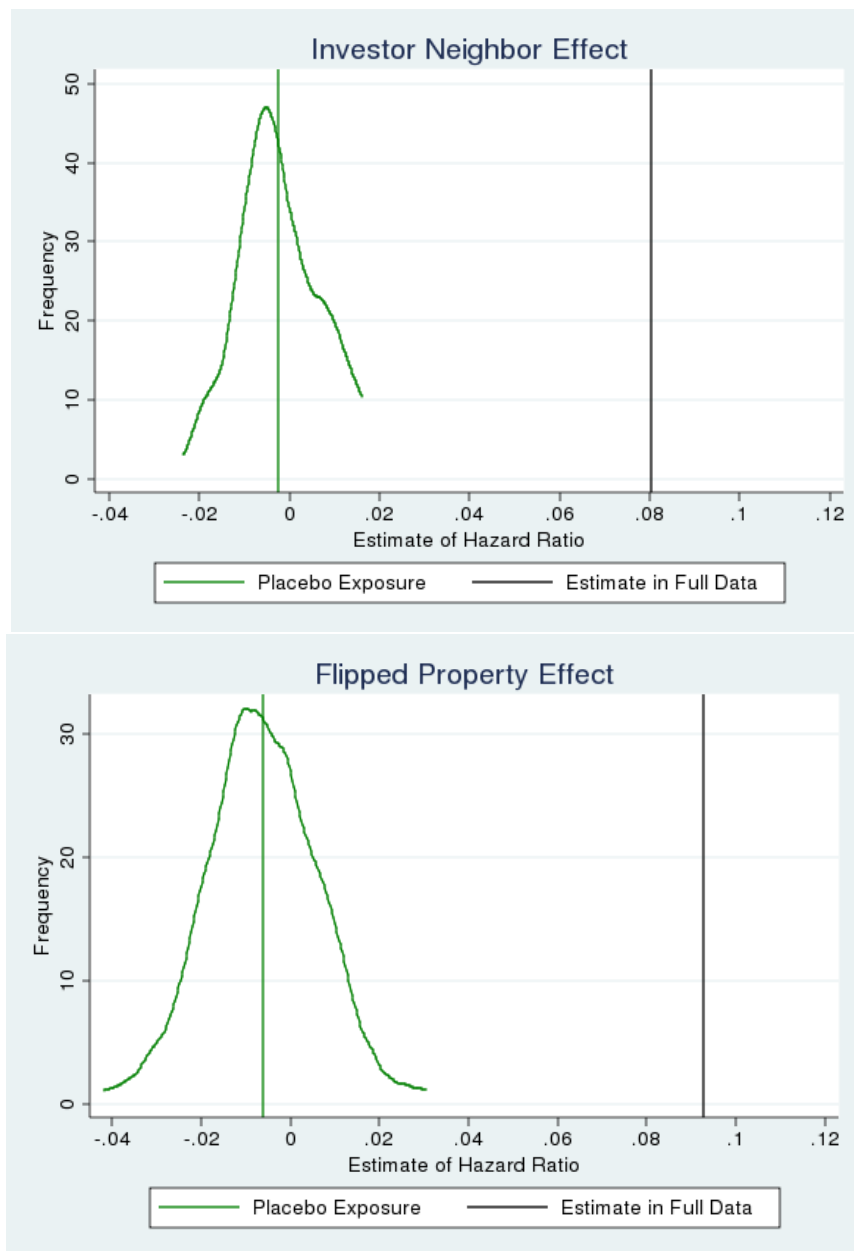
As a falsification test of our main results, we calculate a false exposure measure by randomly drawing an at-risk tenure’s inner ring (0.1 mile) exposure to both flips and investor neighbors conditional on their actual outer ring exposure. For example, if an at-risk tenure is exposed to 3 flips within 0.5 miles, we randomly draw how many of these are within 0.1 mile according to the actual distribution of inner/outer exposure in the data. We then estimate comparable hazard regressions using this false-exposure data set.

In Figure 6, we report the distribution of the inner hazard effect from 100 draws of the false-exposure data set. The placebo effect is centered tightly around zero, with the

actual estimated effect far higher than even the largest placebo estimate. Alternatively, Figure 7 presents the full distribution of both the false and actual data hazard estimates for 250 draws of one-tenth subsets of the at-risk tenure data. Here too, the placebo estimate is approximately centered at zero, showing that our research design produces an unbiased estimator. Moreover, the actual data produces a distribution that is clearly to the right of the false exposure for both types of investing activity, indicating that homes actually exposed to very nearby investing activity are significantly more likely to become investors themselves.

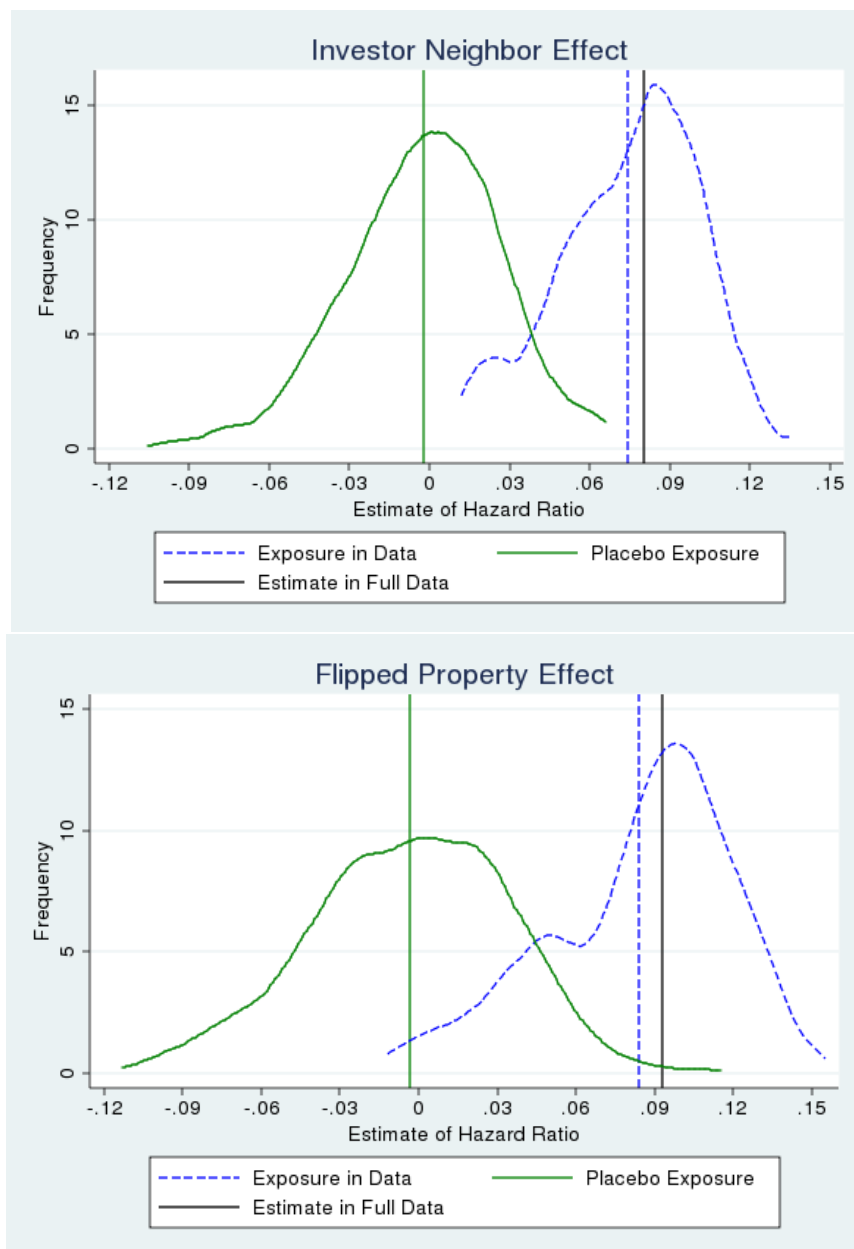
While this cannot identify the actual mechanisms at work—information sharing, word-of-mouth influence, etc.—the results of the placebo test are strong evidence of causal effects occurring within the narrow ring, city-block level of geography that induces homeowners to engage in investing activity.

Figure 6: Placebo Test of Contagion Mechanism: Full Data Sample with Placebo Exposure.



NOTES: The figure displays the distribution of the hazard estimator for the false (placebo) exposure to investment activity. The distribution comes from 100 random draws of placebo inner ring (0.1 mi.) exposures over the full data set. The vertical lines indicate the means, while the solid black vertical line is the estimate from the full data sample (Table 6).

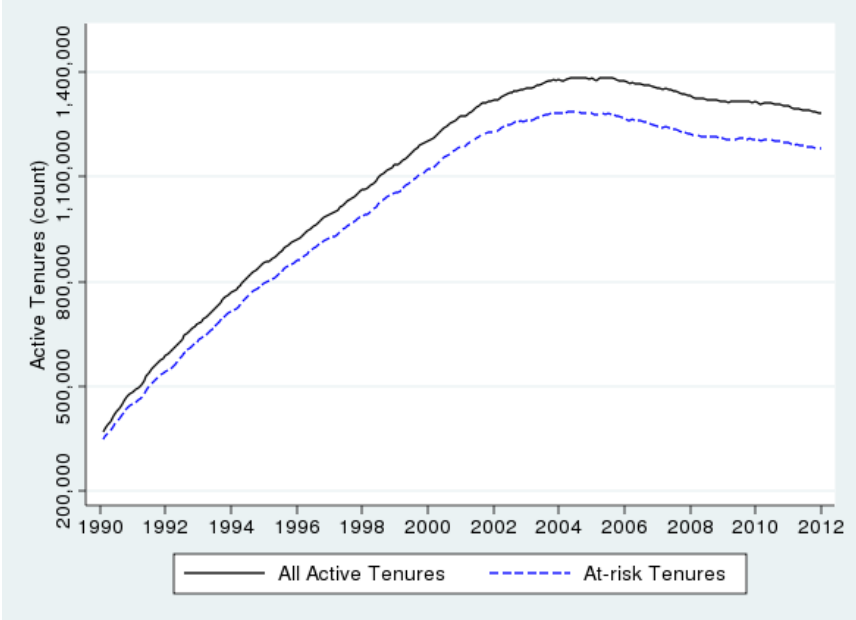
Figure 7: Placebo Test of Contagion Mechanism: Repeated Subsamples With Actual Data and Placebo Exposure.



NOTES: The figure displays the distribution of the hazard estimator for the false (placebo) and actual exposure to investment activity. The distribution comes from 250 one-tenth subsamples of the at-risk tenures drawn with replacement. The vertical lines indicate the means, while the solid black vertical line is the estimate from the full data sample (Table 6).

B Additional Exhibits

Figure 8: Active Tenures and At-Risk Tenures Over Time, Monthly.



NOTES: The figure displays the count of active tenures and at-risk tenures (i.e. not, or not yet, investors) identified using the transaction data.

Table 15: Overlap in Measures of Investment Activity at 0.1 Mile Radius.

Investor's residences ↓	Flipped properties →	0	1	≥2	Total
0	Count	72,531,908	8,548,771	1,352,215	82,432,894
	<i>cell pct</i>	69.30	8.17	1.29	78.76
	<i>row pct</i>	87.99	10.37	1.64	100.00
	<i>col pct</i>	81.09	68.84	48.35	78.76
1	Count	14,080,376	2,766,344	754,047	17,600,767
	<i>cell pct</i>	13.45	2.64	0.72	16.82
	<i>row pct</i>	80.00	15.72	4.28	100.00
	<i>col pct</i>	15.74	22.28	26.96	16.82
≥2	Count	2,837,826	1,102,407	690,699	4,630,932
	<i>cell pct</i>	2.71	1.05	0.66	4.42
	<i>row pct</i>	61.28	23.81	14.91	100.00
	<i>col pct</i>	3.17	8.88	24.69	4.42
Total	Count	89,450,110	12,417,522	2,796,961	104,664,593
	<i>cell pct</i>	85.46	11.86	2.67	100.00
	<i>row pct</i>	85.46	11.86	2.67	100.00
	<i>col pct</i>	100.00	100.00	100.00	100.00

NOTES: The table reports the joint distribution of the two exposure variables for the estimation sample in the inner ring (0.1 mile). Counts refer to tenure-month level observations in the main sample period of 2000-2007.

Table 16: Robustness of Baseline Results to Functional Specification.

	1		2		3		4	
	Baseline (Qtrly Effects)	Indicator for > 0 at 0.1mi, Count Controls	RHS Indicator Instead of Counts Indicator for > 0 at 0.1mi, Indicator Controls	Baseline (No Qtrly Effects)	Probit	Nonlinear Models Logit		
Investor Neighbor Coefficient (w/i. 0.1 mi)	0.347 (0.053)***	0.344 (0.069)***	0.677 (0.066)***	0.324 (0.052)***	0.017 (0.003)***	0.062 (0.01)***		
Hazard (w/i. 0.1mi)	0.085 (0.013)***	0.085 (0.017)***	0.174 (0.017)***	0.080 (0.013)***	0.064 (0.011)***	0.064 (0.011)***		
Flipped Properties: Coefficient (w/i. 0.1 mi)	0.412 (0.066)***	0.490 (0.084)***	0.910 (0.081)***	0.378 (0.065)***	0.025 (0.003)***	0.089 (0.012)***		
Hazard (w/i. 0.1mi)	0.101 (0.016)***	0.122 (0.021)***	0.235 (0.021)***	0.093 (0.016)***	0.093 (0.013)***	0.093 (0.013)***		
Qtrly effects	Y	Y	Y					

NOTES: The table presents robustness checks relating to functional form specification. See the text for details and variable descriptions. The indicator variable specifications use dummies for whether exposure is greater than zero. Standard errors in parentheses are clustered at the at-risk tenure level. Coefficients have been multiplied by 10,000 for readability. *** p<0.01, ** p<0.05, * p<0.1, +p<0.2.

Table 17: Returns on Properties With Observed Sale, Pre-2000.

	1	2	3
	Total Returns $\frac{p_{sale} - p_{buy}}{p_{buy}}$	Buyer Discount $\ln(p_{buy}) - \ln(\hat{p}_{buy})$	Seller Premium $\ln(p_{sell}) - \ln(\hat{p}_{sell})$
Purchases 1988-1999			
<i>N</i>	61,215	61,215	61,215
Mean	0.3969	-0.0058	0.0103
SD	0.5646	0.1852	0.1841
Investor Neighbor <i>w/i</i> 0.1mi	0.0657 (0.0672)	0.0092 (0.0239)	0.0242 (0.0238)
Flip <i>w/i</i> 0.1mi	0.1542 (0.0749)**	-0.0076 (0.0267)	0.0194 (0.0265)
Controls			
Time of Entry	Y	Y	Y
Outer Ring (<i>w/i</i> 0.5 mi)			
Exposure	Y	Y	Y

NOTES: *** p<0.01, ** p<0.05, * p<0.1, +p<0.2.