NBER WORKING PAPER SERIES

TESTING FOR INFORMATION ASYMMETRIES IN REAL ESTATE MARKETS

Pablo Kurlat
Johannes Stroebel

Working Paper 19875
http://www.nber.org/papers/w19875

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2014

This paper was previously circulated as "Knowing your neighborhood: Asymmetric Information and Real Estate Markets." We thank Viral Acharya, Yakov Amihud, Aurel Hizmo, Andreas Fuster, Stefano Giglio, Theresa Kuchler, Stijn van Nieuwerburgh, Matteo Maggiori, Holger Mueller, Monika Piazzesi, Florian Scheuer and Amit Seru, our discussant Chris Parsons, as well as seminar participants at Stanford University, Chicago Booth, Berkeley Haas, New York University and the 2013 Summer Real Estate Symposium. We thank Trulia for providing data. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2014 by Pablo Kurlat and Johannes Stroebel. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Testing for Information Asymmetries in Real Estate Markets
Pablo Kurlat and Johannes Stroebel
NBER Working Paper No. 19875
January 2014
JEL No. D53,D82,G14,R21,R31

ABSTRACT

We study equilibrium outcomes in markets with asymmetric information about asset values among both buyers and sellers. In residential real estate markets hard-to-observe neighborhood characteristics are a key source of information heterogeneity: sellers are usually better informed about neighborhood values than buyers, but there are some sellers and some buyers that are better informed than their peers. We propose a new theoretical framework for analyzing such markets with many heterogeneous assets and differentially informed agents. Consistent with the predictions from this framework, we find that changes in the seller composition towards (i) more informed sellers and (ii) sellers with a larger supply elasticity predict subsequent house-price declines and demographic changes in that neighborhood. This effect is larger for houses whose value depends more on neighborhood characteristics, and smaller for houses bought by more informed buyers. Our findings suggest that home owners have superior information about important neighborhood characteristics, and exploit this information to time local market movements.

Pablo Kurlat
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305
and NBER
pkurlat@stanford.edu

Johannes Stroebel
New York University
Leonard N. Stern School of Business
44 West 4th Street, KCM 9-98
New York, NY 10012
johannes.stroebel@nyu.edu

In many markets, market participants have differential information about important characteristics of heterogenous assets. Akerlof (1970), for example, analyzes a situation in which sellers of used cars have superior information relative to potential buyers. In other markets, sellers are better informed than buyers on average, but not all buyers and sellers are equally well informed. Consider the residential real estate market: transaction prices include payments for both the land and the structure, each of which is hard to value and can be a source of asymmetric information between market participants.[1] On average, home sellers are likely to have better information than potential buyers about neighborhood and house characteristics. In addition, however, some of the possible buyers or sellers might have an information advantage relative to their peers. For example, real estate agents might be particularly well informed about neighborhood gentrification patterns and demographic trends, and buyers who have previously lived in the same neighborhood face less of an information disadvantage relative to buyers who are moving from further away.

In this paper we argue that such asymmetric information is substantial and has important implications for equilibrium housing market outcomes. Our empirical analysis is guided by the predictions from a new theoretical framework for analyzing markets with many heterogeneous assets and differentially informed agents. In our model, an agent's valuation of a property depends on characteristics of both the neighborhood and the structure. Current home owners can observe these characteristics for their own property, but the valuation of their current unit also includes an idiosyncratic shock that captures, for example, the need to move for job-related reasons. All potential buyers value a property identically based on characteristics of the neighborhood and the structure, both of which they do not observe perfectly. We model information as the ability to differentiate between properties of different overall value and assume that some agents can do this better than others. Differential information across buyers is not only a realistic features of many asset markets, but will generate additional predictions that allow us to cleanly identify the presence of asymmetric information in markets with some degree of price predictability, such as housing markets.

In this setting we cannot apply existing asset pricing models with heterogeneously informed agents, which typically rely on one of two standard notions of equilibrium. Models in the spirit of Akerlof (1970) assume that assets which buyers cannot distinguish are pooled at the same price. In our setting, whether or not two assets are distinguishable depends on the identity of the buyer: only some buyers can tell apart good and bad neighborhoods and houses. Therefore, the price pooling assumption cannot be made here, at least not without

[1] For example, the value of a house's structure includes hard-to-observe aspects of construction quality (Stroebel, 2013). Similarly, local amenities such as crime rates or school quality that are capitalized in the value of the property's land component are oftentimes in flux, and hard for market participants to observe (Guerrieri et al., 2010a).

further qualification. In models relying on rational expectations equilibria along the lines of Grossman and Stiglitz (1980) each asset is divisible, clearly identified and has its own separate market. Some traders may be less informed about the expected value of the asset, but they can be sure to be trading the same asset as other traders. We want to capture a stronger notion of relative informedness: some traders can tell whether two houses have different fundamentals while others cannot. This implies that an uninformed buyer cannot learn about the value of house A by looking at the equilibrium price of house B because he does not know whether it is the same kind of house.

We propose a notion of competitive equilibrium that respects traders' individual ability to tell houses apart from one another. We allow each house, in principle, to be traded at any price. Sellers choose at what prices to (try to) sell their house; buyers choose at what price to (try to) buy a house and use their information to select which house to buy. The prices at which houses actually trade are determined by a form of rationing at every possible price, as in the competitive equilibrium defined by Kurlat (2012). We derive conditions under which more informed buyers choose better houses even though all houses that are observationally equivalent to the least informed buyer endogenously trade at the same price.

The model has several implications for equilibrium outcomes. First, the characteristics of sellers in a neighborhood should predict subsequent price changes for houses in that neighborhood. This is because some owners are more likely to sell in response to changes in hard-to-observe and partially unpriced neighborhood characteristics. More responsive owners might include those with better information, or those with houses that are more affected by neighborhood trends. Consequently, the proportion of more-responsive owner types among sellers is indicative of these neighborhood characteristics. As these characteristics become more visible to all market participants, home prices will adjust towards the true property value. Hence, appreciation during an ownership period will be correlated with the composition of sellers at the time the house was purchased. Second, this effect will be stronger for houses with a high neighborhood-β, since their value is more dependent on neighborhood factors. Third, more informed buyers should obtain higher appreciation on average, because they are able to select better houses among the heterogeneous pool of houses on sale. Fourth, the appreciation obtained by more informed buyers should be less sensitive to hard-to-observe neighborhood characteristics (and hence seller composition) than that of less informed buyers. Informed buyers select which house to buy based on their combined information about both the structure and the neighborhood. Therefore they trade them off in a way that less-informed buyers do not: conditional of buying from a worse neighborhood, they are more selective on the structure, which reduces the effect of neighborhood characteristics on the value of the houses they buy.

We test these predictions empirically using nearly 20 years of transaction-level house price data from Los Angeles county, covering about 1.5 million property sales. We first document that average neighborhood price appreciation correlates with changes in the composition of sellers as predicted by the model. We focus on three measures of the composition of sellers. First, we argue that real estate professionals should be particularly well informed about changes in the quality of their neighborhood. Using data on the universe of real estate licenses issued by the California Department of Real Estate, we find that a one standard deviation increase in the share of real estate professionals amongst sellers in a neighborhood predicts a decline in future annual neighborhood appreciation of 13 basis points. Second, we argue that owners of houses whose value is more affected by neighborhood characteristics (higher neighborhood-β houses) should respond more elastically to changes in neighborhood characteristics, as the value of their house is more affected when the neighborhood changes. Since neighborhood characteristics are primarily capitalized in the land component of a property's value, we propose the share of land in the total property value assigned by the tax assessor as a proxy for the property's neighborhood-β. We verify this by showing that the prices of properties with a larger land share do in fact respond more to changes in average neighborhood prices. Consistent with the model we find that a one standard deviation increase in the average neighborhood-β of houses sold in a neighborhood is predictive of future neighborhood-level price declines of 75 basis points annually. Finally, we argue that longer-tenure residents are less elastic in their decision to move, and show that a one standard deviation increase in the share of sellers who have only recently moved to the neighborhood predicts neighborhood price declines of 47 basis points annually.

In addition to determining the impact of seller composition on neighborhood level house price changes, we also test directly whether seller composition is correlated with observable changes in neighborhood characteristics. Using data from the California Department of Education and the Home Mortgage Disclosure Act, we show that the share of socioeconomically disadvantaged students in local schools as well as the average income and racial composition of new home buyers moves with our three measure of the composition of sellers in a neighborhood in a way that is consistent with the model predictions. We also show that the impact of changes in seller composition on subsequent price changes is indeed significantly larger for houses with a higher neighborhood-β.

While this evidence is highly consistent with the importance of asymmetric information about neighborhood characteristics in housing markets, the significant autocorrelation of house price change means that a relationship between seller characteristics and subsequent return is by itself insufficient evidence for the presence of asymmetric information. For example, it could be that all market participants are equally aware of future neighborhood level price declines, but more responsive owners react more strongly to them. To rule out such alternative explanations, we control for past neighborhood level price changes in our regressions, to remove the commonly predictable component of house price changes. Our estimated correlation between seller composition and subsequent house price changes remains unchanged. In addition, we also consider how house appreciation varies with the characteristics of the buyer, and argue that these findings are uniquely explained by information asymmetries.

We find that real estate agents purchase houses that experience almost a full percentage point higher subsequent capital gains. We also identify a second set of buyers who are likely to be better informed about neighborhood characteristics. Specifically, using the geocoded address of all transacted properties combined with the identity of the transactors, we identify buyers who previously owned a house relatively close to the property they are purchasing. We argue that these buyers are likely to be better informed about neighborhood characteristics, and find that they indeed purchase houses that experience above-average subsequent capital gains. This is hard to reconcile with an explanation in which all agents are equally informed. Crucially, we also show that the impact of seller composition on price appreciation is smaller for houses bought by real estate agents and for houses bought by individuals who have previously lived closer to the house they are purchasing. This is consistent with the model's prediction that the capital gains of more informed buyers should be less sensitive to hard-to-observe neighborhood characteristics than that of less informed ones. Models of price predictability for reasons other than asymmetric information do not generate these predictions.

Asymmetric information in real estate markets has been considered in a number of different settings. Garmaise and Moskowitz (2004) examine the importance of asymmetric information about property values in commercial real estate markets. They use regional variation in the quality of tax assessments to proxy for the importance of private information and show that properties with less informative assessments attract more local buyers whose geographic proximity allows them to obtain a better valuation of the property. Stroebel (2013) shows that lenders differ in their information about the true quality of houses used as mortgage collateral, and use this to subject less informed lenders to adverse selection on collateral quality. Levitt and Syverson (2008) analyze the interaction between a home seller and her real estate agent who has better information about the value of the house. They show that agents exploit this information asymmetry to advise homeowners to sell too quickly relative to when agents sell their own home. Bayer et al. (2011) analyze the roles of speculators and middlemen in real estate markets, but find no evidence that these agents have superior information that allows them to time changes in market prices.[2] Chinco and Mayer (2013) argue that out-of-town investors are less able to time market movements relative to local speculators. Cheng et al. (2013) analyze whether managers working in mortgage securitization were aware of the housing bubble, but find no evidence that these agents had superior information that allowed them to time the market. Relative to this literature, the current paper is the first to document that neighborhood characteristics provide a significant source of information asymmetry in residential real estate markets, allowing owners to time market movements. It also is the first with an explicit empirical focus on understanding market outcomes when some sellers and buyers are better informed than their peers..

We also contribute to a large literature that has tested the predictions from trading models with asymmetrically informed agents in markets other than real estate. One important set of papers analyzes correlations between the trading behavior of firm insiders and subsequent stock returns. For example, Lorie and Niederhoffer (1968) measure the predictive properties of insiders transactions, and find that they forecast large movements in stock prices. See Jaffe (1974), Finnerty (1976), Seyhun (1986, 1992), Lin and Howe (1990) and Coval and Moskowitz (2001) for related studies. Easley et al. (2002), Kelly and Ljungqvist (2012) and Choi et al. (2013) show the empirical importance of information asymmetries in Grossman and Stiglitz (1980)-style asset pricing models applied to the stock market. In our setting, we show that the share of informed and elastic sellers predicts future neighborhood-level capital gains, suggesting that they, too, have insider information about characteristics of the neighborhood. However, in real estate markets the autocorrelation in house prices means that a relationship between sellers' behavior and subsequent price changes by itself is not sufficient evidence for asymmetric information. To rule out alternative explanations, we analyze unique predictions from a model of differentially informed buyers and sellers.

On the theoretical side, the model builds on Kurlat (2012), who proposes a definition of competitive equilibrium for markets with asymmetric information where some buyers have different quality of information. We adapt this definition to to account for the indivisibility of houses and, importantly, extend it to cases where also sellers are heterogeneous in various dimensions. This theory in turn builds on the literature that, following Akerlof (1970), has analyzed competitive equilibria in settings with asymmetric information (Wilson, 1980; Hellwig, 1987; Gale, 1992, 1996; Dubey and Geanakoplos, 2002; Guerrieri et al., 2010b). Relative

[2] In other settings, more informed investors have been shown to be able to time market movements to their advantage. Brunnermeier and Nagel (2004) conclude that hedge funds were able reduce their exposure to tech stocks before the dot-com bubble burst. Temin and Voth (2004) study a sophisticated investor, Hoare's Bank, who successfully traded on knowledge of the South Sea bubble. Cohen et al. (2008) show that portfolio managers perform better on holdings of stock when they attended school with senior management or board members. Cohen et al. (2010) estimate that analysts outperform by up to 6.60% per year on their stock recommendations when they have an educational link to the company.

to this literature, the current model allows different buyers to be differentially informed. The model also relates to the literature on markets where more- and less- informed traders coexist (Grossman and Stiglitz, 1980; Kyle, 1985). While this literature has typically focused on settings with a single asset or several clearly distinct assets and aggregate uncertainty about supply and/or payoffs, here the setting is closer to the original Akerlof (1970) model, with many heterogeneous assets but no uncertainty about aggregate variables.

In the next section we present a new competitive equilibrium framework for analyzing markets with heterogeneously informed buyers and sellers, which is mapped to the real estate market in Section 2. Section 3 describes the datasets used in our empirical application. Section 4 tests the predictions from the model, showing the significant presence of asymmetric information about neighborhood characteristics in the residential real estate market.

1 Model

1.1 Agents and preferences

There is a unit measure of houses. Current home owners decide whether to offer their homes for sale. If an owner stays in his house, he will derive utility

$$u = [\beta\theta + (1-\beta)\,\eta]\,\varepsilon \tag{1}$$

from living in it. θ is distributed according to F_θ, with support in $\left[0,\bar{\theta}\right]$. It is a common shock to all houses in a given neighborhood but takes different values in different neighborhoods. It represents neighborhood-level factors such as school quality and crime rates. η is distributed according to F_η, with support in $[0,\bar{\eta}]$. It takes a different value for each house and captures house-specific factors such as construction quality and maintenance. β is the relative weight of neighborhood-level factors in the overall value of the property. For now assume that it is a fixed parameter and the same for every house; heterogeneity in β across houses will play a role later. We refer to $v \equiv \beta\theta + (1-\beta)\,\eta$ as the total value of a house. The distribution of v, denoted F with density f and support $[0,\bar{v}]$ results from the convolution of $\beta\theta$ and $(1-\beta)\,\eta$. Values of ε are distributed according to G, with support in $\mathbb{R}^+$. This variable captures idiosyncratic shocks to the quality of the match between a house its current owner, resulting from changes in family structure or job-related relocation needs. Low values of ε (mismatches between a house and its current owner) are the source of gains from trade. The fact that ε enters equation (1) multiplicatively means that the potential gains from trade are proportional to the value of the house. θ, η and ε are independent. There is a large mass of potential buyers of houses. They all have identical preferences, and their valuation for a house is equal to its total value v, i.e. $\varepsilon = 1$ for all potential buyers.[3]

1.2 Information

Both θ and η are private information of each house's current owner. Buyers can be of two types, informed or uninformed. Uninformed buyers have no information on either θ or η for any house.[4] This need not mean that the buyer finds all houses literally indistinguishable; there could be a certain amount of public information, such as each house's number of bedrooms or widely known features of the neighborhood, which is already built into the prior distribution F. In other words, we are focusing on houses that look indistinguishable to uninformed buyers.

Informed buyers receive a signal from each house, given by:

$$s(v) = \mathbb{I}[v \geq b] \tag{2}$$

for a house of total value v, where $b > 0$. This signal allows informed buyers to determine whether any given house is a relatively good house ($v \geq b$) or a relatively bad house ($v < b$).

There is a measure n^U of uninformed buyers and a measure n^I of uninformed buyers. We will focus on the case where n^U is large and n^I is small, i.e. there is a large number of potential buyers but a relatively small number (in a sense to be made precise later) of them are informed.

1.3 Equilibrium

We study a competitive equilibrium of this economy, abstracting from transactions costs and search frictions. We propose a notion of competitive equilibrium that respects uninformed buyers' inability to tell houses apart from one another. This definition is adapted from Kurlat (2012), modified to account for the indivisibility of individual houses and for having only two types of buyers, and adjusted to allow us to analyze the role played by differentially informed sellers.

Trading takes place as follows. There exists a finite set of prices $P \subseteq [0, \bar{v}]$ at which a house may trade.[5] An auctioneer calls out the prices in P in descending order. At every price, owners decide whether to put their house on sale or not. If an owner does not put his

[3] Allowing for buyer-specific utility for each house would not affect the main predictions from the model.

[4] Extending the model to allow for an infinite number of differentially informed buyer types does not affect any of the model's predictions. Please contact the authors for details.

[5] Finiteness is assumed to avoid mathematical complications but is not essential. One can think of prices being rounded to the nearest dollar.

house on sale at price p then he withdraws from the market and cannot put it on sale at any price $p' < p$; therefore owners' decisions can be summarized by a reservation price.

At every price, potential buyers must decide whether to buy a house among those on offer. If an uninformed buyer buys a house, he picks one at random from among the houses that are offered. If an informed buyer buys a house, he can be selective and only accept houses for which he observes $s(v) = 1$, i.e. only relatively good houses; he picks a house at random from among the relatively good houses on offer.[6] Letting $S(p, v)$ denote the supply of houses of value v at price p, the expected value of the house bought by uninformed and informed buyers respectively will be:

$$\bar{v}^U(p) = \frac{\int_0^{\bar{v}} vS(p,v)\,dv}{\int_0^{\bar{v}} S(p,v)\,dv} \tag{3}$$

$$\bar{v}^I(p) = \frac{\int_b^{\bar{v}} vS(p,v)\,dv}{\int_b^{\bar{v}} S(p,v)\,dv} \tag{4}$$

as long as the denominator is positive; if the denominator in (4) is zero, then the buyer will not find an acceptable house at price p.

Note that, depending on buyers' decisions, many of the houses that are offered at any price p will remain unsold; some of these will perhaps be sold at lower prices. As in Gale (1996), we do not assume that supply equals demand at any particular price. Instead, the probability of trade is what clears the market at every price.

We denote the reservation price for owners with a house of value v and idiosyncratic shock ε by $p^R(v, \varepsilon)$ and the probability that a house of value v sells at price p by $\eta(p, v)$.

Definition 1. *A competitive equilibrium consists of:*

1. *Reservation prices $p^R(v, \varepsilon)$ for every $\{v, \varepsilon\}$;*

2. *Measures of houses $d^U(p)$ and $d^I(p)$ demanded by uninformed and informed buyers respectively at each price $p \in P \cup \emptyset$, with $\sum_{P\cup\emptyset} d^U(p) = n^U$ and $\sum_{P\cup\emptyset} d^I(p) = n^I$;*[7]

3. *Supply $S(p, v)$ for each $\{p, v\}$.*

4. *Selling probabilities $\eta(p, v)$ for each $\{p, v\}$;*

[6] Assuming buyers always receive a random house from among the acceptable ones requires assuming that the relative proportions of different acceptable houses do not depend on which other buyers buy at the same price. A sufficient condition for this is to impose that uninformed buyers pick houses first. Kurlat (2012) shows that if one allows for different possible orderings, this is indeed the one that emerges in equilibrium.

[7] $d^U(\emptyset)$ and $d^I(\emptyset)$ denote the measures of uninformed and informed buyers who choose not to buy a house.

such that

1. *Reservation prices are set optimally, i.e.* $p^R(v,\varepsilon) = v\varepsilon$

2. *Buyers choose* p *optimally, i.e. if* $d^U(p) > 0$ *then* p *solves*

$$\max_p \bar{v}^U(p) - p \tag{5}$$

   and if $d^I(p) > 0$ *then* p *solves*

$$\max_p \bar{v}^I(p) - p \tag{6}$$

   where it is understood that $\emptyset$ *maximizes (5) or (6) if there is no* p *such that the objective is strictly positive.*

3. *Supply is consistent with reservation prices and selling probabilities*

$$S(p,v) = f(v) \times G\left(\frac{p}{v}\right) \times \prod_{\tilde{p}>p} [1 - \eta(\tilde{p}, v)] \tag{7}$$

4. *The probabilities of selling are consistent with agents' decisions:*

$$\eta(p,v) = \frac{d^U(p)}{\int_0^{\bar{v}} S(p,\tilde{v})d\tilde{v}} + \frac{d^I(p)\mathbb{I}[v \geq b]}{\int_b^{\bar{v}} S(p,\tilde{v})d\tilde{v}} \tag{8}$$

Equation (7) says that the supply of houses of value v at price p will be equal to the total number of owners who (i) have a house of value v, (ii) are willing to sell at price p, i.e. have $\varepsilon \leq \frac{p}{v}$ and (iii) have tried and failed to sell their house at every price higher than p. Equation (8) is interpreted as follows: for each type of buyer, the probability of selling a house of value v is the ratio of the demand of that buyer (provided he accepts houses of value v) to the total supply of houses that buyer accepts. Adding up over uninformed and informed buyers results in (8).

Under assumptions 1-3 below, the equilibrium will be such that all trades of houses that look indistinguishable to uninformed buyers take place at the same price p^*. Thus, even though the equilibrium construct allows for trading at many possible prices simultaneously, the on-equilibrium trading behavior is actually quite simple: all houses that are observationally equivalent to uninformed buyers trade at the same price, but informed buyers can pick better houses at that price.

1.4 Equilibrium Characterization

Conjecture that in equilibrium buyers only buy houses at price at a single price p^*. Since no buyers buy at higher prices, then $\eta(v,p) = 0$ for all $p > p^*$ so (7) reduces to

$$S(p^*, v) = f(v)\, G\left(\frac{p^*}{v}\right) \tag{9}$$

Assume that $d^U(p^*) > 0$ and $d^U(\emptyset) > 0$ (which will be true under assumptions 1 and 2 below), so some uninformed buyers buy houses and some do not. Hence p^* must satisfy:[8]

$$p^* = \bar{v}^U(p^*) = \frac{\int_0^{\bar{v}} vS(p^*, v)\, dv}{\int_0^{\bar{v}} S(p^*, v)\, dv} \tag{10}$$

Informed buyers instead will obtain a house with expected value

$$\bar{v}^I(p^*) = \frac{\int_b^{\bar{v}} vS(p^*, v)\, dv}{\int_b^{\bar{v}} S(p^*, v)\, dv} \tag{11}$$

Equations (10) and (11) imply that $\bar{v}^I(p^*) > p^*$ so informed buyers strictly prefer buying a house to not buying one, and therefore $d^I(\emptyset) = 0$. It remains to show that they would rather buy at price p^* instead of at some other price, i.e. that $d^I(p^*) = n^I$.

Assumption 1. $n^I < \int_b^{\bar{v}} S(p^*, v)\, dv$

Assumption 1 says that there are sufficiently few informed buyers that they cannot buy all the good houses that are offered at price p^*.

If $d^I(p^*) = n^I$, then condition (8) says that the fraction of houses of quality v sold at price p^* must be:

$$\eta(p^*, v) = \frac{d^U(p^*)}{\int S(p^*, \tilde{v})\, d\tilde{v}} + \frac{n^I \mathbb{I}(v \geq b)}{\int_b^{\bar{v}} S(p^*, \tilde{v}) d\tilde{v}} \tag{12}$$

In equilibrium, $d^U(p^*)$ must be such that $\eta(p^*, v) = 1$ for all $v \geq b$. In other words, uninformed buyers buy just enough houses so that all the relatively good houses are sold. The reason is that if they bought fewer houses than this, then some houses, including high-v houses, would remain unsold at price p^* and their owners would also attempt to sell them at price $p^* - \epsilon$; buyers would then prefer to buy at price $p^* - \epsilon$ instead of price p^* and we would not have an equilibrium at p^*. Conversely, if uninformed buyers bought any more houses, some informed buyers would not be able to buy a house because they would run out; they

[8] Equation (10) could have more than one solution. In that case, the equilibrium corresponds to the highest-price solution.

would therefore have an incentive to preempt this by buying at price $p^* + \epsilon$ and we would not have an equilibrium at p^*. Hence, in equilibrium it must be that

$$d^U(p^*) = \int S(p^*, v)\, dv \cdot \left(1 - \frac{n^I}{\int_b^{\bar{b}} S(p^*, v)\, dv}\right) \tag{13}$$

Assumption 2. $n^U > \int S(p^*, v)\, dv \cdot \left(1 - \frac{n^I}{\int_b^{\bar{v}} S(p^*, v) dv}\right)$.

Assumption 2 formalizes the condition that the number of uninformed buyers be sufficiently large to buy d^U houses as given by equation (13). This justifies the underlying assumption of equation (10) that uninformed buyers get no surplus and some choose not to buy. For any $v < b$, the selling probability will be given by (12) and will be lower than 1, meaning that some houses will remain unsold. The owners of those houses will offer them on sale at all prices $p \in [v\varepsilon, p^*)$ in addition to offering them at price p^*. Buyers, if they wanted to, could choose to buy houses at those alternative prices. In order to establish that in equilibrium they do not do so and trade takes place at the single price p^*, we must verify that at any price $p < p^*$ the adverse selection problem is so much more severe than at p^* that buyers have no incentive to buy at p.

By (7), for $p < p^*$ the supply of houses of quality v at price p is

$$S(p, v) = G\left(\frac{p}{v}\right) f(v)\left(1 - \eta(p^*, v)\right) \tag{14}$$

Assumption 3. *For any $p < p^*$, $\bar{v}^U(p) - p \leq 0$.*

Assumption 3 ensures that uninformed buyers do not wish to buy from the residual supply at a lower price that by buying from the original supply at price p^* (informed buyers would never buy at prices below p^* because all the houses that they would accept have been sold at p^*). Assumption 3 will hold if the left tail of the distribution of house qualities is sufficiently fat and/or the left tail of the distribution of idiosyncratic shocks for owners is sufficiently thin that even at prices approaching 0 the adverse selection effect is strong enough to prevent trade. If assumptions (1)-(3) hold then the equilibrium for houses that appear identical to the least informed buyer can be characterized by:

1. An equilibrium price given by (10).
2. Average qualities obtained by buyer b given by (10) and (11).
3. Selling probabilities for house quality v given by (12).
4. Demand from uninformed buyers given by (13).

Notice that one significant feature of the equilibrium is that uninformed buyers cannot learn about the quality of individual houses by observing equilibrium prices since (i) they cannot infer the quality of house B from the price of house A because to do so would require establishing that the houses are truly similar, which is precisely what they cannot do and (ii) in equilibrium, all observationally equivalent houses trade at the same price anyway.

Appendix A.1 provides an example to illustrate the features of the equilibrium as well as the content of assumptions (1)-(3).

2 Predictions of the model

2.1 Mapping the model to the data

We assume that owners in our data have bought houses in a market that is well described by the above model; by the time they resell their house, v has become public information, so the sale price is $v + u$, where u is a mean-zero independent random variable that captures any unexpected shocks that take place during the owner's tenure. A buyer's expected appreciation during his tenure is therefore $v - p^*$.

Our main object of interest in the empirical analysis will be the price appreciation experienced by different owners, which we relate to the model's predictions for $v - p^*$. One of the main tests of the model is to see how appreciation differs across neighborhoods. We will assume that within a neighborhood, θ is the same for all houses for a given period and the law of large numbers applies with respect to η; therefore any average effects for a neighborhood are due to shocks to θ.

2.2 Composition of sellers

Suppose that owners of houses belong to one of two possible groups, A and B, with probabilities π_A and π_B. These two groups differ with respect to the conditional distribution of the ε shock, which we denote by G_A and G_B (the unconditional distribution is still G).

Proposition 1. *Suppose that* $\frac{g_B(\varepsilon)}{G_B(\varepsilon)} \geq \frac{g_A(\varepsilon)}{G_A(\varepsilon)}$ *for every* $\varepsilon \geq \frac{p^*}{\bar{v}}$. *Then the proportion of sellers who belong to group* A *among sellers in neighborhood* j *is increasing in* θ_j.

Proof. See Appendix A □

Marginal sellers are those for whom $\varepsilon = \frac{p^*}{v}$, while those with lower ε are infra-marginal. The condition $\frac{g_B(\varepsilon)}{G_B(\varepsilon)} \geq \frac{g_A(\varepsilon)}{G_A(\varepsilon)}$ says that group B has more marginal sellers (as a fraction of infra-marginal sellers) than group A and therefore its supply is more elastic with respect to changes in $\frac{p^*}{v}$. Proposition 1 says that in neighborhoods that experience negative shocks, sellers should include a relatively higher share from more elastic groups in the population, because the more elastic groups respond to low v (high $\frac{p^*}{v}$) by putting their houses on sale. The following prediction is a direct implication of Proposition 1:

Prediction 1. Group composition of sellers in a neighborhood should predict subsequent appreciation.

A high proportion of sellers from inelastic groups means that θ (and therefore average v in the neighborhood) is high but this has not yet been captured in the price. When this shock is revealed, at least in part, during the following owner's holding period, he will experience higher-than-average house appreciation. The model so far is agnostic regarding which groups of owners have more or less elastic supply, so in principle any predictability of house appreciation on the basis of the group composition of sellers can be considered evidence in favor of the model. A stricter test of the model is to see whether high returns correlate with a high proportion of sellers from groups where there are a priori theoretical reasons to believe that they have more elastic supply. In the following, we consider three groups of potentially more elastic sellers: those who are better informed about neighborhood characteristics, those who own houses whose value changes more with neighborhood characteristics, and those who have only recently moved to the neighborhood.

2.2.1 More and less informed sellers

The model so far has considered owners that are perfectly informed about the neighborhood quality θ. Suppose we instead allowed some owners to be better informed than others. In particular, assume a group of owners are informed and observe θ perfectly while others are less informed and only observe a noisy signal x, and assume that whether an owner is informed is independent of the realization of θ. Denote the conditional expectation of θ given x by $\hat{\theta}(x)$. The informed will sell their house if $\varepsilon \leq \frac{p^*}{\beta\theta+(1-\beta)\eta}$ while the uninformed will sell theirs if $\varepsilon \leq \frac{p^*}{\beta\hat{\theta}(x)+(1-\beta)\eta}$.

Proposition 2. *The proportion of informed among sellers is higher in the worst neighborhood ($\theta = 0$) than in the best neighborhood $\left(\theta = \bar{\theta}\right)$*

Proof. See Appendix A □

Proposition 2 says that in the lowest-θ neighborhood we should expect to see a high proportion of informed owners among sellers, while in the best neighborhood the proportion should be lower. Proposition (2) does not necessarily imply that the fraction of informed sellers decreases monotonically with θ, but this is true for many common cases; for instance it is true if θ and x are Normally distributed. Overall, the logic behind Proposition 2 is that informed owners' selling decisions react more strongly to θ, simply because they know about its realization. This leads to the following prediction:

Prediction 1.a. The fraction of informed sellers in a neighborhood should be negatively associated with subsequent appreciation of houses in that neighborhood.

2.2.2 Neighborhood-β of transacted homes

Suppose that different houses within a neighborhood have different loadings on neighborhood and idiosyncratic factors so the value of a house is $v = \beta_h \theta + (1 - \beta_h)\, \eta$, where β_h is different for different houses. Assume that the distribution of β_h within a neighborhood is independent of the realization of the neighborhood-quality shock θ. This shock affects different houses in the same neighborhood differently depending on their value for β_h and we would therefore expect the supply response to θ to depend on β_h.[9]

Proposition 3.

1. *Assume $\bar{\theta} \geq \bar{\eta}$. Then the proportion of owners who choose to sell is increasing in β_h in the worst neighborhood ($\theta = 0$) and decreasing in β_h in the best neighborhood ($\theta = \bar{\theta}$).*

2. *The proportion of owners who choose to sell in a neighborhood of quality θ does not change with θ for houses with $\beta_h = 0$ and decreases with θ for houses with $\beta_h = 1$.*

Proof. See Appendix A □

A neighborhood-level shock has a larger impact on high-β_h houses than on low-β_h houses. Therefore, high-β_h owners will put their house on the market in response to low θ (or withdraw them from the market in response to high θ) to a greater extent than low-β_h owners.

[9] In general, changing the weights on θ and η means that the distribution of v, and therefore possibly the equilibrium price, will be different for houses of different β_h. One case where this does not happen is if house-specific and neighborhood-level shocks are drawn from the same distribution (i.e. $F_\theta = F_\eta$) and we only consider local deviations of β_h around $\beta_h = \frac{1}{2}$. In this case,

$$f(v) = \frac{1}{\beta(1-\beta)} \int f_\eta\left(\frac{v-x}{1-\beta}\right) f_\theta\left(\frac{x}{\beta}\right) dx$$

$$\frac{df(v)}{d\beta} = -\frac{1-2\beta}{[\beta(1-\beta)]^2} \int f_\eta\left(\frac{v-x}{1-\beta}\right) f_\theta\left(\frac{x}{\beta}\right) dx + \int \left[f'_\eta\left(\frac{v-x}{1-\beta}\right) \frac{v-x}{(1-\beta)^2} f_\theta\left(\frac{x}{\beta}\right) - f_\eta\left(\frac{v-x}{1-\beta}\right) \frac{x}{\beta^2} f'_\theta\left(\frac{x}{\beta}\right)\right] dx$$

so if $f_\theta = f_\eta$, then $\left.\frac{df(v)}{d\beta}\right|_{\beta=\frac{1}{2}} = 0$ for all v.

The comparison is unambiguous for the extreme cases of comparing $\beta_h = 0$ and $\beta_h = 1$, or for comparing the propensity to sell in the best and worst neighborhoods. With more assumptions about distributions it is possible to make stronger statements. For instance, if $\varepsilon \sim U[a, 1]$, $\eta \sim U[0, \bar{\eta}]$, $\theta < \bar{\eta}$ and $\beta = 0.5$, then $\frac{d^2 \Pr[Sell|\theta,\beta]}{d\theta d\beta} < 0$. This says that for uniform distributions, as long as the weight of neighborhood and structure factors in the total value of the house is close to even, then one can ensure that the effect holds locally, not just in comparing the extremes of the distribution. Overall, the logic behind Proposition 3 is that higher-β_h owners' selling decisions react more strongly to θ. This leads to the following prediction:

Prediction 1.b. The average neighborhood-β of houses sold in a neighborhood should be negatively associated with subsequent appreciation of houses in that neighborhood.

2.2.3 Sellers of different tenure

A further dimension of heterogeneity among home owners is how long ago they have bought their house; the distribution of ε could be different depending on the owner's tenure. In general, it is possible to think of reasons why $\frac{g(\varepsilon)}{G(\varepsilon)}$ might be either increasing or decreasing in the owner's tenure.

Suppose owners receive idiosyncratic shocks every period and ε results from the sum of all these shocks over time. This will generate three effects. First, ε will have higher variance for long-tenure owners who have received more shocks. Other things being equal, a more dispersed distribution means more owners are either very well or very poorly matched with their house and thus have inelastic decisions. Second, if each period's shocks have non-zero mean, then the distribution of ε will shift over time. Arguably, it makes sense to assume that the shocks are likely to have a negative mean, as owners are initially well matched with their house and match quality tends to deteriorate thereafter. This induces a downward drift in ε, which increases $G(\varepsilon)$ over time, reinforcing the prediction that long-tenure owners should have lower elasticity. Third, long tenure owners are a selected sample, since they are the ones that chose not to sell their house in the past. The selection into non-selling eliminates the left tail of the distribution each period, potentially introducing a positive drift, which lowers $G(\varepsilon)$ and possibly makes long-tenure owners more elastic. The more selective the non-seller sample (i.e the higher the likelihood of selling the house), the stronger this countervailing effect.

In Appendix A.2 we work through a numerical example that suggests that long-tenure owners are likely to have less elastic supply. This implies that when θ is high, the relatively elastic short-tenure owners should be less likely to sell, leaving a high proportion of long-tenure sellers among sellers. If the example is correctly calibrated, this suggests the following prediction:

Prediction 1.c. The proportion of long-tenure sellers in a neighborhood should be positively associated with subsequent appreciation of houses in that neighborhood.

However, because the prediction comes from a stylized two-period model and relies on functional form assumptions, the relative elasticity of different owner groups is ultimately an empirical question more than an unambiguous theoretical prediction. Therefore, in our empirical analysis we focus on whether the composition of sellers by tenure is a consistent indicator of θ across the various specifications where the model indicates that it should be.

2.3 Differential effect by neighborhood-β

Suppose, as in section 2.2.2, that different houses within a neighborhood have different β_h. The sensitivity of house values to neighborhood level shocks will depend on each house's β_h.

Proposition 4. *The response of a house's value to a shock to the quality of its neighborhood is increasing in β_h.*

Proof.

$$v = \beta_h \theta + (1 - \beta_h)\,\eta$$

$$\frac{\partial^2 v}{\partial \theta \partial \beta_h} = 1 > 0$$

□

Proposition 4 implies that high-β_h houses appreciate more than low-β_h houses after a high neighborhood shock θ. The shock itself is unobservable, but one empirical implication of this is that when the seller composition in a neighborhood suggests that θ is high, then one should expect to see, not just higher subsequent appreciation overall but a disproportionate effect on high-β_h houses.

Prediction 2. Seller composition in a neighborhood should predict more subsequent appreciation (in absolute value) for high-β houses.

2.4 Differential effects by buyer information

A straightforward implication of the model is that, by being able to select better houses at the same price, more informed buyers experience higher subsequent price appreciation.

Proposition 5. *The expected value of a house conditional on being bought by an informed buyer is higher than its expected value conditional on being bought by an uninformed buyer.*

Proof. Immediate from equations (10) and (11). □

Proposition 5 immediately implies the following prediction:

Prediction 3. More informed buyers should obtain higher appreciation.

A second implication of the model is that the expected appreciation of a house bought by an informed buyer is less sensitive to the neighborhood quality than that of a house bought by an uninformed buyers.

Proposition 6. *Assume $\beta\bar{\theta} > b$. Then, conditional on buying in a sufficiently good neighborhood, the expected value of houses bought by informed and uninformed buyers is the same.*

Proof. See Appendix A □

Conditional on buying a house in a neighborhood that turns out to be bad, the information of the buyer will have a strong impact on the expected subsequent appreciation of the house. The reason is that informed buyers will only buy high-η houses in such a neighborhood while uninformed buyers might buy any house. In a good neighborhood, however, most houses will be acceptable to informed buyers so, conditional on buying in such a neighborhood, they will be drawing from a sample that is not so different from the one that uninformed buyers draw from; therefore buyer information has less impact on expected subsequent appreciation. In the limit of a sufficiently good neighborhood, then all houses are acceptable to all buyers and buyer information has no impact. Proposition 6 implies the following prediction:

Prediction 4. The differential appreciation obtained by informed buyers should be negatively associated with seller compositions that predict high neighborhood appreciation.

In other words, both being bought by an informed buyer and being located in a neighborhood where few of the sellers are from elastic groups should predict a high appreciation for a given house, but the interaction of these two variables should be negative.

3 Data description

To conduct the empirical analysis, we combine a number of datasets. The first dataset contains information on the universe of ownership-changing housing deeds in Los Angeles county between June 1994 and the end of 2011. We observe approximately 7.15 million deeds covering such transactions. Properties are uniquely identified via their Assessor Parcel Number (APN). Variables in this dataset include property address (including latitude and longitude of each property), contract date, transaction price, type of deed (e.g. Intra-Family Transfer Deed, Warranty Deed, Foreclosure Deed) and the identity of the buyer and seller. It also reports the amount and duration of the mortgage and the identity of the mortgage lender. Figure 1 shows the location of each of the properties with transactions in our dataset. From this dataset, we extract all arms-length transactions for which transaction prices reflect the true market value of a property. This procedure, which excludes, amongst others, intra-family transfer deeds and foreclosure deeds, is described in Appendix B. There are about 1.45 million arms-length transactions.

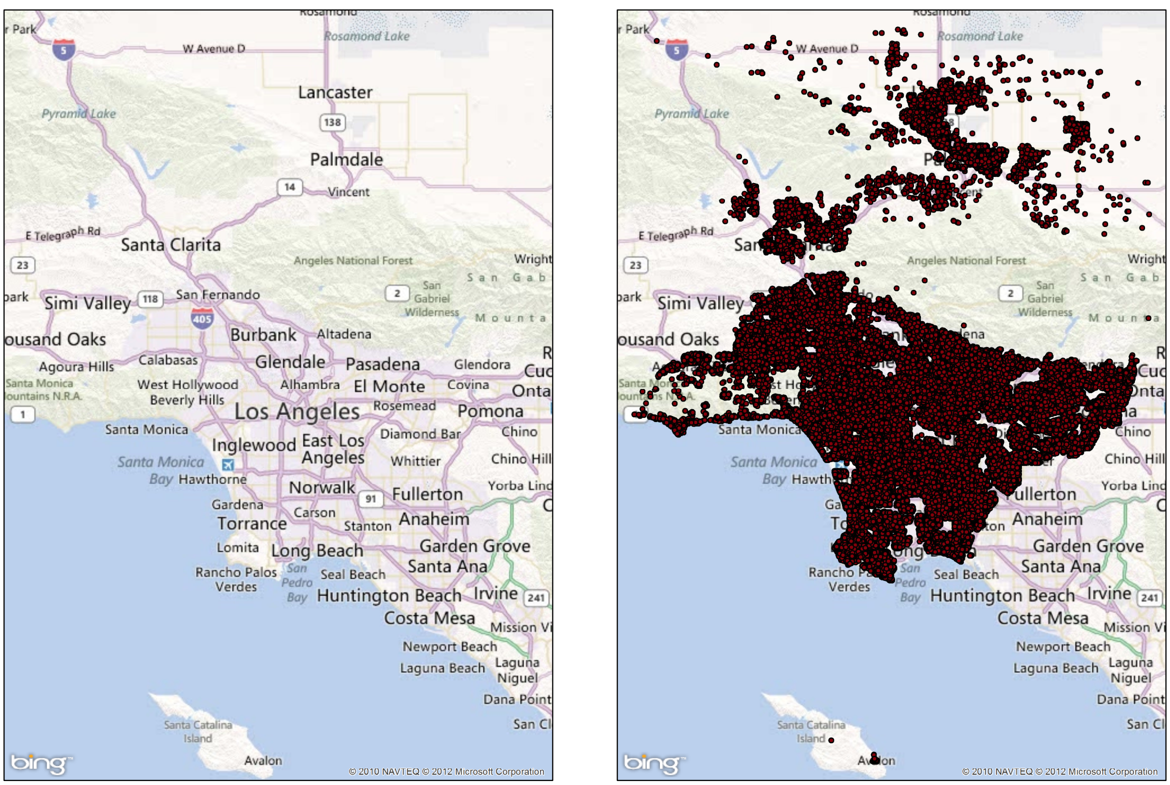


Figure 1: Transaction Sample

Note: This map shows the location of all houses for which we observe a transaction over between June 1994 and December 2011 for the Los Angeles area.

The second dataset contains the universe of residential tax-assessment records for the year 2010. This dataset includes information on property characteristics such as construction year, owner-occupancy status, lot size, building size, and the number of bedrooms and bathrooms. The tax assessment records also include an estimate of the market value of the property in January 2009, split into a separate assessment for the land and the structure. This will be important, since the price of properties with a larger share of total value constituted by land should change more in response to neighborhood characteristics than the price of properties with a smaller land share. In other words, we propose that the land share in total value might be a good proxy for neighborhood-β. Section 4.1 shows empirically that this is indeed the case. Figure 6 in Appendix B.3 provides an example of two properties and their associated land share. As a check whether the relative values assigned to land and property by the tax assessors appear realistic, Figure 2 shows how the fraction of total value that is constituted by land varies across Los Angeles county. As one might expect, land is more valuable relative to the structure in the downtown area and near the coast. Importantly for our purposes, there is also significant variation in the land share measure for houses that are relatively close to each other (i.e. in the same "neighborhood").

We also use data from the California Department of Real Estate on the universe of real estate agent and broker licenses issued in California since 1969. We propose that such real estate professionals may be particularly well informed about changes in neighborhood characteristics relative to other buyers and sellers. We merge this license data to the housing transaction data using the name of the transactors reported in the property deeds. This allows us to identify properties that have been bought or sold by a real estate professional. In particular, we classify a property as having been bought or sold be a real estate professional if at the time of sale there was an active real estate agent or broker license issued in Los Angeles county to somebody of that name.[10]

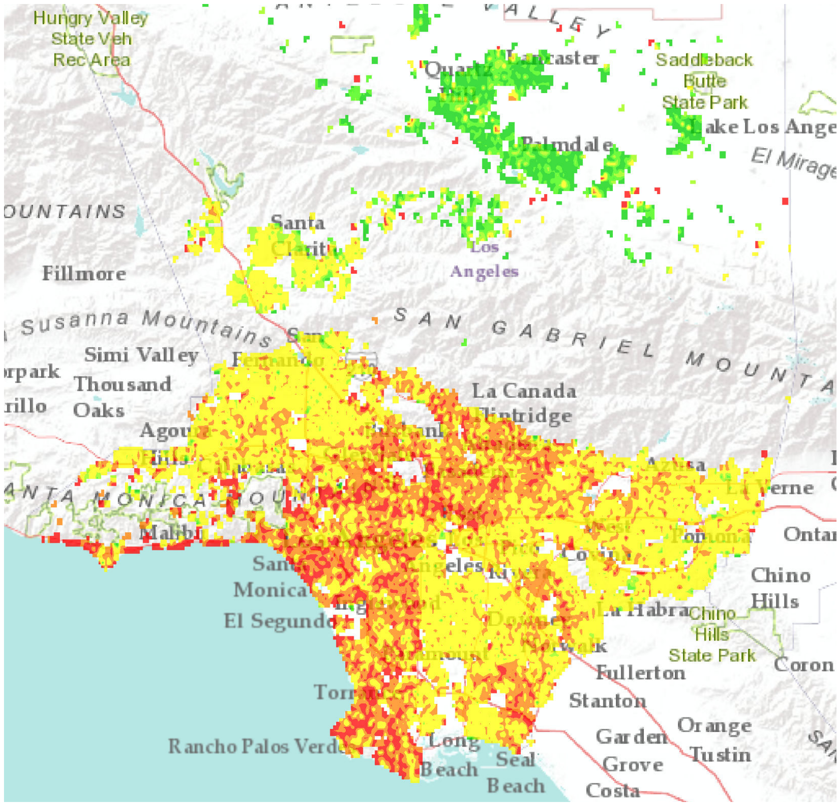
4 Results

4.1 Measuring Neighborhood - β

One of the key characteristics that differentiated houses in the model described in section 1 was the neighborhood-β of the individual homes. That is, houses differed in how much

[10] This will introduce measurement error, since we misclassify people with common names to be real estate agents. However, this should not introduce systematic bias into our analysis other than attenuation bias. For houses jointly bought or sold by more than one individual (for example, by a married couple), we assign the property to have been bought or sold by a real estate professional if one of the transactors' names matches with an active real estate agent or broker license.

Figure 2: Heat Map of Land Share in Property



Note: This figure shows the distribution of the fraction of total property values that is made up from land, as reported in the assessment records. Land share in total value is increasing from green to red.

their value varies as neighborhood characteristics change. To test the model, we must first determine a measure of each house's neighborhood-β. As suggested above, neighborhood characteristics have a larger effect on the land value component of a property than on the structure value component. This is because in the long-run it is the land rather than the structure that capitalizes neighborhood amenities (e.g. Arnott and Stiglitz, 1979; Davis and Heathcote, 2007; Albouy, 2009).[11]

In this section we show that the land share in total value of each house as identified by the tax assessor is indeed a good proxy for the neighborhood-β of that house. We consider a zip code as the neighborhood of interest. For each pair of arms-length transactions of house i located in zip code n with first sale in quarter q_1 and second sale in quarter q_2 we calculate the annualized capital gain of the house between the two transactions. In addition, we measure average price movements in that zip code over the same period, $ZipCapGain_{n,q_1,q_2}$. We do this by determining the annualized change in the median transaction price. In addition,

[11] We might expect the value of the structure to also be affected in the short-run, but less so than the value of the land.

we construct a measure of the land share in total value for each house, $LandShare_i$, by exploiting that the assessor records provide each house with a separate valuation of the land and the structure component. We then run the regression specified in equation (15) for all repeat sales between June 1994 and December 2011.

$$CapGain_{i,n,q_1,q_2} = \alpha + \beta_1 ZipCapGain_{n,q_1,q_2} + \beta_2 ZipCapGain_{n,q_1,q_2} \times LandShare_i + \epsilon_i \quad (15)$$

The results are presented in Table 1. In column (1) we drop the interaction between *ZipCapGain* and *LandShare*. The coefficient on *ZipCapGain* shows that, reassuringly, on average house price movements closely track movements of the zip code median price. In column (2) we include the interaction. The positive coefficient β_2 shows that houses with a larger land share in total value move more in the direction of the market, both when prices increase and when prices decrease. This suggests that the land share of a house is indeed an appropriate proxy for the neighborhood-β of that house. In column (3) we only include transaction pairs from zip codes with at least 5,000 transactions between June 1994 and and December 2011. For those zip codes the measurement of average neighborhood level price changes is more precise. The results are unchanged when looking at this subsample.

Table 1: Land Share as Neighborhood-β

	(1) Capital Gain	(2) Capital Gain	(3) Capital Gain
Zip Code Capital Gain	0.997***	0.955***	0.966***
	(0.004)	(0.013)	(0.015)
Zip Code Capital Gain × Land Share		0.068***	0.061***
		(0.017)	(0.021)
R-squared	0.793	0.793	0.808
N	391,533	391,531	286,140

Note: This table shows the results from regression 15. We include all sales pairs in the June 1994 to December 2011 period. In column (3) we restrict the sales pairs to be from zip codes with at least 5,000 transactions observed between 1994 and 2011. Standard errors are clustered at the zip code level.

4.2 Changes in Seller Composition Predict Price Changes

In this section we test Prediction 1, which says that if sellers have superior information about neighborhood characteristics, then changes in the composition of sellers in a neighborhood should be predictive of future price changes of homes in that neighborhood. We regress the annualized capital gain of houses between two arms-length transactions, $CapGain_{i,n,q_1,q_2}$, on control variables and the composition of sellers in neighborhood n and quarter q_1. We focus on three measures of seller composition, suggested by Predictions 1.a., 1.b., and 1.c. respectively: (a) the fraction of sellers that are real estate professionals, and are thus particularly well informed about neighborhood characteristics, (b) the average land share of transacted houses and (c) the average time sellers have lived in their home.[12] Table 2 shows summary statistics of the seller composition variables for two definitions of a neighborhood: a zip code and a 4-digit census tract. We show both the sample-wide standard deviation, as well as the within-neighborhood standard deviation.

Table 2: Summary Statistics Seller Composition

Variable	Neighborhood	Mean	Standard Deviation	
			Unconditional	Conditional
Share Informed Sellers	Zip Code	0.043	0.031	0.027
	Census Tract	0.043	0.052	0.050
Average Seller Land Share	Zip Code	0.594	0.114	0.045
	Census Tract	0.594	0.122	0.056
Seller Share Tenure > 3	Zip Code	0.789	0.079	0.068
	Census Tract	0.789	0.120	0.111

Note: This table shows summary statistics for the seller composition by quarter and neighborhood for two different definitions of neighborhood: zip code and 4-digit census tract. Standard deviations are shown both unconditionally and conditional on the particular neighborhood (i.e. showing the within-neighborhood standard deviation). The sample period for share of informed sellers and average seller land share is June 1994 to December 2011; for the share of sellers with tenure exceeding 3 years the sample period is July 1997 to December 2011.

We then run regression 16 using different geographies as our definition of a neighborhood. The regression includes neighborhood fixed effects as well sales quarter pair fixed effects, to remove aggregate (Los Angeles-wide) market movements in house prices over time.

$$CapitalGain_{i,n,q_1,q_2} = \alpha + \beta_1 SellerComposition_{n,q_1} + X_i'\beta_2 + \xi_n + \phi_{q_1,q_2} + \epsilon_i \tag{16}$$

Table 3 shows the results from regression 16 when we consider a neighborhood to be a zip

[12]Our measurement of home tenure is censored, since for sellers who initially bought a property before the beginning of our sample period (June 1994) we cannot observe actual tenure, but only know that it must have been longer than the time since the beginning of the sample. To deal with this, we define a long-tenure seller to be someone who moved into the neighborhood more than 3 years ago. We then consider the impact of the share of long-tenure sellers amongst the total population of sellers, and only look at the return between transaction pairs where $q_1 >$ Q2 1997. Results are not sensitive to the choice of 3 years as the cut-off value; please contact authors for those results.

code. Standard errors are clustered at the quarter by zip code level. Column (1) analyzes the impact of the share of real estate agents amongst home sellers on the subsequent return of homes without controlling for home and buyer characteristics. A one conditional standard deviation increase in the share of sellers that are real estate professionals is associated with a 13 basis points decline in the annualized return of houses. In column (2) we add a large set of control variables X_i, including information on the property (age, building size, number of bedrooms and bathrooms, information on pool and air conditioning, property type), the buyers (whether they are married, Asian or Latino) and the mortgage financing (the loan-to-value ratio, the mortgage duration, and whether it is a VA, FHA or jumbo mortgage). Appendix B describes these control variables in more detail, and provides summary statistics. The estimated correlation between changes in the seller composition and subsequent returns is unchanged by the addition of these control variables. This suggests that the correlation is not driven by observable differences in the composition of houses or buyers that might confound our estimates of the impact of the composition of sellers (see Altonji et al., 2005). This is comforting, since we argue that the correlation is driven by hard-to-observe information that current inhabitants have about neighborhood characteristics.

In columns (3) - (4) we consider the effect of changes in the composition of transacted houses towards those with a higher land share in total value. We argued that an increase in the average land share of transacted homes should predict future declines in neighborhood prices since the owners of homes with a higher neighborhood-β should be more elastic in their response to sell upon hard-to-observe negative neighborhood shocks. The results in column (4) suggest that a one conditional standard deviation increase in the average land share of houses sold is indeed associated with a 75 basis points decline in subsequent annualized capital gains in that neighborhood.

In columns (5) and (6) we analyze the impact of a change in the share of long-tenured sellers. The results in column (6) suggest that a one conditional standard deviation decrease in the share of sellers with tenure of more than three years is associated with a decline in annualized capital gains of houses in that neighborhood by about 47 basis points. We argue that this is consistent with owners that have only recently moved into the neighborhood being more elastic in their decision to sell when neighborhood characteristics change. In column (7) we jointly include all three measures of neighborhood composition. The magnitude of the estimated contribution of each of the three measures falls somewhat, as one would expect if each is a noisy measure of the same underlying neighborhood characteristics.

The results presented in Table 3 suggest that the characteristics of sellers within a zip code are correlated with subsequent neighborhood price changes. However, there might be additional relevant information about the immediate neighborhood of a particular property

Table 3: Effect of Seller Composition in Zip Code on Capital Gains

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share Informed Sellers	-4.415***	-4.701***					-4.497***
	(0.883)	(0.865)					(0.909)
Average Seller Land Share			-16.76***	-17.64***			-14.15***
			(0.770)	(0.767)			(0.892)
Share in Zip of Tenure > 3					6.781***	6.808***	5.828***
					(0.383)	(0.377)	(0.372)
Fixed Effects	✓	✓	✓	✓	✓	✓	✓
House and Buyer Controls	·	✓	·	✓	·	✓	✓
R-squared	0.626	0.636	0.628	0.638	0.647	0.658	0.659
$\bar{y}$	12.58	12.56	12.58	12.56	13.72	13.70	13.70
N	394,801	391,835	394,801	391,835	302,568	300,106	300,106

Note: This table shows results from regression 16. The dependent variable is the annualized capital gain of the house between the two sequential arms-length sales. The seller composition variables are measured at the quarter × zip code level. All specifications include sales quarter pair fixed effects and zip code fixed effects. Columns (2), (4), (6) and (7) control for characteristics of the property (property size, property age, property type, number of bedrooms and bathrooms, whether the property has a pool or air conditioning), characteristics of the financing (mortgage type, mortgage duration and loan-to-value ratio), and characteristics of the buyer (married, Asian, Latino). Standard errors are clustered at the quarter × zip code level. Columns (1) - (4) include sales pair where the first sale was after June 1994, columns (5) - (7) include sales pairs where the first sale was after June 1997. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

that is not reflected in the composition of all sellers in a zip code, but only in the composition of sellers in the more immediate vicinity of a property. Table 4 reports results from regression 16 with a neighborhood being defined as a four-digit census tract. While there are 293 unique zip codes in our sample, there are 1,255 unique 4-digits census tracts. The results in columns (1) - (3) include census tract fixed effects in addition to the sales quarter pair fixed effects. As before, increases in the share of informed sellers and the average land share of transacted homes predict subsequent declines in neighborhood level capital gains, while an increase in the average tenure of sellers predicts increases in neighborhood level capital gains. The magnitude of the estimated effect is smaller than the ones estimated at the zip code level, probably due to more noise in the measures of seller composition and the resulting attenuation bias. In addition, columns (4) - (6) include an interaction of zip code fixed effects with the sales quarter pair fixed effects in addition to census tract fixed effects. This allows the time movement of house prices to differ by zip code. Here, the identification comes from differential variation of seller composition across census tracts within the same zip code. Since this removes neighborhood characteristics that are common for different census tracts within the same zip code, the estimated coefficients are unsurprisingly smaller.

In Appendix B we provide various robustness checks to this analysis. In particular, we show that the results are not driven by selection into the sample of repeat sales, that our tenure results are not driven by the presence of "flippers", and that the results extend to considering subsequent ownership periods of the house.

Table 4: Effect of Seller Composition in Census Tract on Capital Gains

	(1)	(2)	(3)	(4)	(5)	(6)
Share Informed Sellers	-1.492***			-0.323		
	(0.319)			(0.420)		
Average Seller Land Share		-8.137***			-3.869***	
		(0.379)			(0.660)	
Share in CT of Tenure > 3			2.960***			1.939***
			(0.163)			(0.378)
Fixed Effects	$q_1 \times q_2$, Census Tr.	$q_1 \times q_2$, Census Tr.	$q_1 \times q_2$, Census Tr.	$q_1 \times q_2 \times$ zip, Census Tr.	$q_1 \times q_2 \times$ zip, Census Tr.	$q_1 \times q_2 \times$ zip, Census Tr.
House and Buyer Controls	✓	✓	✓	✓	✓	✓
R-squared	0.638	0.639	0.660	0.687	0.687	0.705
$\bar{y}$	12.56	12.56	13.70	12.56	12.56	13.70
N	391,800	391,800	300,082	391,800	391,800	300,082

Note: This table shows results from regression 16. The dependent variable is the annualized capital gain of the house between the two sequential arms-length sales. The seller composition variables are measured at the quarter × 4-digit census tract level. Columns (1) - (3) includes sales quarter pair fixed effects and census tract fixed effects, while columns (4) - (6) include sales quarter pair × zip code fixed effects in addition to census tract fixed effect. All specifications include characteristics of the property (property size, property age, property type, number of bedrooms and bathrooms, whether the property has a pool or air conditioning), characteristics of the financing (mortgage type, mortgage duration and loan-to-value ratio) and characteristics of the buyer (married, Asian, Latino). Standard errors are clustered at the quarter × census tract level. For the results in column (4) clustered standard errors could not be produced, and robust standard errors are reported. Columns (1), (2), (4) and (5) include sales pairs where the first sale was after June 1994, columns (3) and (6) include sales pair where the first sale was after June 1997. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

One important question is why less informed home buyers do not condition their choice of house, or at least their choice of neighborhood, on the composition of sellers if it is truly informative about neighborhood characteristics. One reason is that in practice this information is unavailable or extremely hard to obtain in real time. For example, it usually takes months before deed records are updated and accessible to the public. In addition, the bulk-level transaction-level deeds information that would be required to analyze changes in the seller composition, is not directly provided by Los Angeles county, but only accessible through commercial data providers at costs that are prohibitive to individual home buyers. In addition, the very significant transaction costs in the housing market (about 6% of purchase price) make this market unattractive to arbitrageurs who might have the resources to purchase real time data access.

4.3 Predictability in House Prices

In markets like the stock market, for which we have strong theoretical and empirical reasons to believe that they are relatively efficient and frictionless, an uninformed marginal investor should not be able to predict price changes. In such markets, the fact that some traders' behavior predicts price changes constitutes strong evidence that they are better informed than the marginal investor. In housing markets, however, it is a well established empirical fact that aggregate price changes are at least somewhat predictable (see Ghysels et al., 2013, and the references therein). This predictability complicates the interpretation of our results up to now as tests for asymmetric information: finding a correlation between seller composition and subsequent returns is a necessary, but not a sufficient condition to detect asymmetric information. One alternative explanation could be that seller composition predicts appreciation because more elastic groups of owners simply respond more to commonly anticipated changes in neighborhood-level house prices, rather than to private information.[13]

We argue in two ways that such alternative explanations that do not rely on asymmetric information are unable to explain our findings. First, in this section we explicitly control for what the literature has found to be the main source of house price predictability, by showing that the correlation between seller composition and subsequent returns remains unchanged after conditioning on past price changes. Second, in section 4.6 we provide strong evidence for Prediction 4, which is unique to a model with asymmetric information. These tests consider the interaction of buyer and seller informedness, and show that the impact of seller composition on subsequent returns are particularly big for houses purchased by uninformed buyers. This is inconsistent with a story in which seller composition is driven by price movements that are predictable by all market participants.

Case and Shiller (1989, 1990) and a long subsequent literature find that house price appreciation in the short-run is positively serially correlated. Hence it could be that the only reason why the composition of sellers predicts appreciation is that elastic groups of owners react more strongly to changes that everyone can predict on the basis of past appreciation. If this were the case, then controlling for past capital gains in regression 16 should significantly reduce the correlation between seller composition and returns. To show that this is not the case, we run regression 17, where we control for past capital gains of houses in the zip code over the past 12 and 24 months.[14]

[13] Any Los Angeles-wide price predictability is already controlled for through the ϕ_{q_1,q_2} fixed effects.

[14] This uses the same median sales price index as Section 4.1. Similar results are achieved when controlling for price changes over the past 3, 6 and 36 months. Please contact the authors for details.

$$CapGain_{i,n,q_1,q_2} = \alpha + \beta_1 SellerComposition_{n,q_1} + \beta_2 PastCapGain_{n,q_1} + X_i'\beta_3 + \xi_n + \phi_{q_1,q_2} + \epsilon_i \tag{17}$$

The results in Table 5 show that, indeed, past capital gains in a zip code have strong predictive power for future capital gains. However, importantly, the inclusion of past returns as a control variable does not affect the magnitude or statistical significance of the estimated relationship between seller composition and future returns. This suggests that the predictive power of seller composition for future returns is not just driven by the autocorrelation of returns. In other words, sellers are reacting to information beyond what is contained in past returns.[15]

4.4 Seller Composition and Neighborhood Demographics

So far our evidence has shown that changes in seller composition can predict the future capital gains of houses, and can do so over and above what would be predictable from past house price changes. This is a prediction from a model in which sellers are better informed and reacting to ongoing changes in neighborhood characteristics that are difficult for potential buyers to observe. In this section we test for whether changes in seller composition actually predict changes in neighborhood characteristics that are observable at the zip code level.

We employ two datasets that contain information about annual zip code level demographic information. It is important to notice that none of these datasets was available to home buyers at the time of purchasing the house; this means that demographic shifts measured in these data were not easily observable in real time. The first dataset contains information from the Home Mortgage Disclosure Act's (HMDA) Loan Application Registry, which provides details on the near universe of mortgage application in major Metropolitan Statistical Areas. It includes details on the year of mortgage application, the census tract of the house and the applicant's income and race. We use this data to construct an annual zip code level measure of (i) the share of African-American mortgage applicants and (ii) the average income of all mortgage applicants. We then run regression 18, where we regress these demographic measures on the seller composition in that year. As in Section 4.3 we

[15] Controlling for past capital gains accounts for a main source of common predictability but not necessarily for every possible source. In other words, the evidence in Table 5 does not rule out the possibility that, rather than reacting to private information, owners are reacting to other commonly-known information, not observable to the econometrician, that also predicts appreciation. This story, however, would not generate Prediction 4, i.e. the negative interaction between the effect of seller composition and buyer informedness. If seller composition were just a proxy for commonly known information that is not observed by the econometrician, this would not mean that the advantage of informed buyers is especially great in neighborhoods that are predicted to underperform.

Table 5: Effect of Seller Composition on Capital Gains - Control for Past Capital Gains

	(1) Return	(2) Return	(3) Return	(4) Return	(5) Return	(6) Return
Share Informed Sellers	-6.147***			-4.623***	-4.670***	-4.671***
	(0.902)			(0.909)	(0.910)	(0.910)
Average Seller Land Share		-17.95***		-14.00***	-13.99***	-14.04***
		(0.809)		(0.896)	(0.898)	(0.898)
Share in Zip of Tenure > 3			6.795***	5.957***	5.944***	5.940***
			(0.377)	(0.372)	(0.375)	(0.374)
Capital Gain Past Year	0.754***	0.660***	0.646***	0.645***		0.499**
	(0.192)	(0.187)	(0.198)	(0.194)		(0.228)
Capital Gain Past Two Years					0.423***	0.208
					(0.152)	(0.178)
Fixed Effects	✓	✓	✓	✓	✓	✓
House and Buyer Controls	✓	✓	✓	✓	✓	✓
R-squared	0.635	0.636	0.658	0.659	0.659	0.659
MeanDepVar	12.98	12.98	13.70	13.70	13.70	13.70
N	367632	367632	300064	300064	300009	299993

Note: This table shows results from regression 17. The dependent variable is the annualized capital gain of the house between the two sequential arms-length sales. The seller composition and the past capital gains variables are measured at the quarter × zip code level. All specifications include sales quarter pair fixed effects, zip code fixed effects, and control for characteristics of the property (property size, property age, property type, number of bedrooms and bathrooms, whether the property has a pool or air conditioning), characteristics of the financing (mortgage type, mortgage duration and loan-to-value ratio), and characteristics of the buyer (married, Asian, Latino). Standard errors are clustered at the quarter × zip code level. Columns (1) - (4) include sales pair where the first sale was after June 1994, columns (5) - (7) include sales pairs where the first sale was after June 1997. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

include the capital gain of houses in the zip code over the past year to make sure we are not just capturing differential elasticity to persistent but observable demographic shifts.[16] We also include fixed effects for the calendar year and the zip code, and cluster standard errors at the zip code level.

$$ZipCode\_Demographics_{n,y} = \alpha + \beta_1 SellerComposition_{n,y} + \beta_2 PastCapGain_{n,q_1} + \xi_n + \phi_y + \epsilon_i \quad (18)$$

The results are shown in Table 6. A one conditional standard deviation increase in the share of informed sellers is associated with a 0.2 percentage point increase in the share of African-American mortgage applicants (off a base of 5.7 percent), and a $1,400 decline in the average income of mortgage applicants in that zip code (off a base of $123,900). A one

[16] Results are very similar when excluding this variable, and are available from the authors.

conditional standard deviation increase in the average land share of sellers is also corresponds to a 0.2 percentage point increase in the share of African-American mortgage applicants and a $2,600 decline in the average income reported in mortgage applications. A change in the share of short-tenured sellers is not statistically related to the share of African-American mortgage applicants, but a one standard deviation increase in this value corresponds to a $2,800 decrease in applicant income. This evidence suggests that sellers do in deed react to changes in neighborhood demographics which are hard to observe in real time.

Table 6: Seller Characteristics and Mortgage Applicant Demographics

	Share of African-American Applicants				Average Applicant Income			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Appreciation Past Year	-0.012** (0.005)	-0.011** (0.005)	-0.010* (0.005)	-0.013** (0.006)	-19.85** (8.703)	-19.55** (9.010)	-20.91** (8.593)	-18.17** (8.959)
Share Informed Sellers	0.089*** (0.032)			0.085*** (0.031)	-53.37* (31.18)			-37.23 (32.53)
Average Seller Land Share		0.038** (0.017)		0.036** (0.016)		-58.00** (27.78)		-47.67* (27.98)
Share in Zip of Tenure > 3			0.010 (0.009)	0.015 (0.010)			42.15*** (9.164)	37.26*** (9.450)
Fixed Effects (Zip Code and Year)	✓	✓	✓	✓	✓	✓	✓	✓
R-squared	0.974	0.974	0.974	0.974	0.930	0.930	0.930	0.930
$\bar{y}$	0.057	0.057	0.057	0.057	123.9	123.9	123.9	123.9
N	3,618	3,618	3,618	3,618	3,618	3,618	3,618	3,618

Note: This table shows results from regression 18 for the years 1996 - 2011. The unit of observation is zip code by year. The dependent variable is the share of African-American mortgage applicants (columns 1 - 4), and the average income of all mortgage applicants (columns 5-8). Each specification includes zip code and year fixed effects. Standard errors are clustered at the zip code level. Significance Levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

We also use a second dataset to provide us with zip code level demographic information. In particular, we obtain annual data from the California Department of Education on the demographics of the student population between 2000 and 2011 at the school level. From this data we construct for each zip code a student-population weighted measure of demographics of all schools in that zip code, and then measure the share of students that are classified as socioeconomically disadvantaged.[17] Table 7 shows results from regression 18 replacing $ZipCode\_Demographics_{n,y}$ with the share of socio-economically disadvantaged students. The results show that a one conditional standard deviation increase in the share

[17] A "socioeconomically disadvantaged" student is defined as (i) a student neither of whose parents have received a high school diploma or (ii) a student who is eligible for the free school lunch program.

of informed sellers coincides with a 0.67 percentage point shift in the demographics of the student population towards socioeconomically disadvantaged students (off a base of 60%). Similarly, a one conditional standard deviation increase in the average land share of transacted homes corresponds to an increase in the share of socio-economically disadvantaged students by 0.5 percentage points. Finally, a one conditional standard deviation increase in the share of sellers with only a short ownership-tenure is associated with a 0.64 percentage point increase in the share of children that are economically disadvantaged.

Table 7: Seller Characteristics and "Share of Socio-economically disadvantaged students"

	(1)	(2)	(3)	(4)
Appreciation Past Year	-1.628	-1.219	-1.151	-2.013
	(1.832)	(1.822)	(1.861)	(1.823)
Share Informed Sellers	24.96***			22.88***
	(6.550)			(6.478)
Average Seller Land Share		10.79***		7.869**
		(3.946)		(3.997)
Share in Zip of Tenure > 3			-9.438***	-8.175***
			(2.480)	(2.462)
R-squared	0.968	0.968	0.968	0.969
$\bar{y}$	60.33	60.33	60.33	60.33
N	3,087	3,087	3,087	3,087

Note: This table shows results from regression 18 for the years 2000 - 2011. The unit of observation is zip code by year. The dependent variable is the share of all students that are classified as socio-economically disadvantaged. Each specification includes zip code and year fixed effects. Standard errors are clustered at the zip code level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

These results show that the composition of sellers in a neighborhood is indeed related to changes in neighborhood demographics in the way predicted by the model. While these demographic shifts might be hard for buyers to observe in real time (for example, because the relevant data is usually only released with significant delay), sellers are likely to have better information. As these demographic shifts become common knowledge they will be reflected in prices. This explains why current seller composition predicts future capital gains.[18]

[18] In addition, in appendix B.7 we show that current seller composition also has some predictive power for future demographics, over and above what is predictable using current demographics. This suggest that current sellers do not only have an information advantage in detecting current demographic shifts, but might also have an insight into predicting future demographic shifts.

4.5 Importance of Neighborhood-β

In this section we consider to what extent the impact of neighborhood seller composition varies across different houses within the same neighborhood, testing Prediction 2. Since neighborhood amenities are capitalized in the land value of properties, we would expect the impact of seller composition on price changes to be larger for houses with a larger land share component in total value. To measure whether this is indeed the case, we run regression 19, where $LandShare_i$ is the house-specific share of total value made up by land, as reported in the assessor data. The coefficient of interest is β_3, which measures the increase in the responsiveness of capital gains to seller composition when the house has a larger land share.

$$\begin{aligned} CapGain_{i,n,q_1,q_2} &= \alpha + \beta_1 \times SellerComposition_{n,q_1} + \beta_2 \times LandShare_i + \\ &\quad \beta_3 \times SellerComposition_{n,q_1} \times LandShare_i + X_i'\beta_2 + \xi_n + \phi_{q_1,q_2} + \epsilon_i \end{aligned} \tag{19}$$

The results of this regression are presented in Table 8, for neighborhoods defined as both zip codes and 4-digit census tracts. The effect of all three measures of average seller characteristics is larger for houses with a larger land share. In column (1) we can see that a move from the 25^{th} to the 75^{th} percentile in the land share distribution (i.e. 47% land share to 75% land share) increases the response of annualized capital gain to a one conditional standard deviation increase in the share of informed sellers in a zip code by about 8 basis points. A similar move in the land share distribution will increase the response of capital gains to an increase in the average land share of sellers by 19 basis points (column 2). Finally, moving from the 25th to the 75th percentile in land share distribution will increase the response of capital gains to a one conditional standard deviation change in the share of sellers who have lived in their house for more than 3 years by 8 basis points (column 3). Columns (4) - (6) show similar effects when we consider a census tract to be a neighborhood.

4.6 Relative Informedness of Buyers

The model also suggests that more informed buyers should obtain higher average appreciation (Prediction 3) and that this advantage should be especially strong conditional on buying houses from bad neighborhoods (Prediction 4). To test these results, we construct three measures of better-informed buyers. Our first measure presumes that real estate professionals are more informed about the true value of houses on sale, and tests the predictions by replacing $InformedBuyer_i$ in regression 20 with a dummy variable for whether or not the

Table 8: Effect of Seller Composition by Land Share

	(1)	(2)	(3)	(4)	(5)	(6)
Land Share	-0.607***	7.966***	-4.537***	-0.682***	7.672***	-3.291***
	(0.107)	(0.481)	(1.170)	(0.0979)	(0.435)	(0.731)
Share Informed Sellers	-1.600			1.385		
	(1.153)			(0.435)		
Land Share × Share Informed Sellers	-9.956***			-4.439***		
	(2.332)			(0.869)		
Average Seller Land Share		-7.602***			1.652***	
		(0.919)			(0.637)	
Land Share × Average Seller Land Share		-15.43***			-14.68***	
		(0.852)			(0.783)	
Share in NH of Tenure > 3			4.277***			1.197*
			(1.024)			(0.634)
Land Share × Share in NH of Tenure > 3			4.124***			2.746***
			(1.476)			(0.918)
Neighborhood	Zip Code	Zip Code	Zip Code	Census Tr.	Census Tr.	Census Tr.
Fixed Effects, House and Buyer Controls	✓	✓	✓	✓	✓	✓
R-squared	0.637	0.638	0.659	0.638	0.640	0.660
$\bar{y}$	12.56	12.56	13.70	12.56	12.56	13.70
N	391,835	391,835	300,106	391,800	391,800	300,082

Note: This table shows results from regression 19. The dependent variable is the annualized capital gain of the house between the two repeat sales. The seller composition variables are measured at the quarter × zip code level in columns (1) - (3), and at the quarter × 4-digit census tract level in the other columns. Columns (1) - (3) include sales quarter pair fixed effects and zip code fixed effects, while columns (3) - (6) include sales quarter pair × census tract fixed effect. All specifications include characteristics of the property (property size, property age, property type, number of bedrooms and bathrooms, whether the property has a pool or air conditioning), characteristics of the financing (mortgage type, mortgage duration and loan-to-value ratio) and characteristics of the buyer (married, Asian, Latino). Standard errors are clustered at the . Columns (3) and (6) include sales pairs where the first sale was after June 1997, all other columns include sales pair where the first sale was after June 1994. Standard errors are clustered at the quarter × zip code level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

buyer was a real estate agent.

$$\begin{aligned} CapGain_{i,n,q_1,q_2} \quad = \quad & \alpha + \beta_1 \times SellerComposition_{n,q_1} + \beta_2 \times InformedBuyer_i + \\ & \beta_3 \times SellerComposition_{n,q_1} \times InformedBuyer_i + X_i'\beta_2 + \xi_n + \phi_{q_1,q_2} + \epsilon_i \end{aligned} \tag{20}$$

Table 9 shows the results from this regression. In column (1), which tests Prediction 3, we do not include the measure of seller composition or its interaction with the informed buyer measure. Those buyers who are real estate agents purchase houses that outperform by about

75 basis points annually relative to otherwise observationally similar houses purchased by agents that are not real estate agents. This is consistent with real estate agents being better at picking good deals from the set of homes on offer. In columns (2) - (4) we show that, consistent with Prediction 4, the difference in the capital gain of houses purchased by real estate agents and other individuals is particularly big in neighborhoods that are predicted to underperform. As discussed above, the reason for this is that in good neighborhoods, informed buyers find most houses to be a good deal, and thus behave similarly to uninformed buyers, who cannot tell good and bad houses apart. In bad neighborhoods, however, informed buyers use their information to only select homes that are a particularly good bargain, while uninformed buyers continue to be unable to tell good and bad houses apart.

Table 9: Effect of Buyer Characteristics - Real Estate Professionals

	(1)	(2)	(3)	(4)
Real Estate Prof.	0.737***	0.531***	0.390	1.812***
	(0.0456)	(0.0733)	(0.270)	(0.629)
Share Informed Sellers		-5.105***		
		(0.873)		
Real Estate Prof. × Share Informed Sellers		4.708***		
		(1.496)		
Average Land Share			-17.69***	
			(0.766)	
Real Estate Prof. × Average Land Share			0.593	
			(0.461)	
Share in Zip of Tenure > 3				6.908***
				(0.378)
Real Estate Prof. × Share in Zip of Tenure > 3				-1.373*
				(0.790)
R-squared	0.637	0.637	0.638	0.659
$\bar{y}$	12.56	12.56	12.56	13.70
N	391,835	391,835	391,835	300,106

Note: This table shows results from regression 20. The dependent variable is the annualized capital gain of the house between the two repeat sales. In columns (1) - (3) sales pairs are included when the first sale was after June 1994, in columns (4) when the first sale was after June 1997. All specifications include sales quarter pair fixed effects, zip code fixed effects,characteristics of the property (property size, property age, property type, number of bedrooms and bathrooms, whether the property has a pool or air conditioning), characteristics of the financing (mortgage type, mortgage duration and loan-to-value ratio) and characteristics of the buyer (married, Asian, Latino). Standard errors are clustered at the quarter × zip code level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

We construct two additional measures of the relative informedness of buyers, which exploit that we can observe the names of all buyers and sellers in the deeds records. As our first measure of "informed buyer," we check whether we observe someone with the same name to have purchased or sold a different house in the same zip code in the past year. Having lived in the same zip code should provide buyers with better information relative to buyers who have not done so.[19] About 5% of all houses are bought by individuals who have previously lived in the same zip code. The results are presented in Table 10.

Table 10: Effect of Buyer Characteristics - Same Zip Code

	(1)	(2)	(3)	(4)
Same Zip	1.105***	0.907***	0.162	3.936***
	(0.0618)	(0.106)	(0.309)	(0.800)
Share Informed Sellers		-4.938***		
		(0.873)		
Same Zip × Share Informed Sellers		4.769**		
		(2.248)		
Average Land Share			-17.70***	
			(0.767)	
Same Zip × Average Land Share			1.580***	
			(0.521)	
Share in Zip of Tenure > 3				7.025***
				(0.377)
Same Zip × Share in Zip of Tenure > 3				-3.550***
				(1.020)
R-squared	0.637	0.637	0.638	0.659
$\bar{y}$	12.56	12.56	12.56	13.70
N	391,835	391,835	391,835	300,106

Note: This table shows results from regression 20. The dependent variable is the annualized capital gain of the house between the two repeat sales. In columns (1) - (3) sales pairs are included when the first sale was after June 1994, in columns (4) when the first sale was after June 1997. All specifications include sales quarter pair fixed effects, zip code fixed effects,characteristics of the property (property size, property age, property type, number of bedrooms and bathrooms, whether the property has a pool or air conditioning), characteristics of the financing (mortgage type, mortgage duration and loan-to-value ratio) and characteristics of the buyer (married, Asian, Latino). Standard errors are clustered at the quarter × zip code level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Column (1) shows that those buyers who previously owned a house in the same zip code purchase homes that have a 1.1 percentage point higher annual capital gain. This is consistent with Prediction 3. Columns (2) - (4) test Prediction 4. We can see that the effect of seller composition on the capital gains of houses bought by neighborhood insiders

[19] We only observe the previous location for individuals who were previous owners in a neighborhood. This means that we will assign a value of "0" to current buyers who have previously rented, even if they lived in the same neighborhood. This will downward-bias our estimate of $SameZip$ and its interaction with the seller composition.

is significantly lower.

We also generate a second, more continuous measure of buyer informedness. For those houses bought by people that we observe selling a house anywhere in Los Angeles county within 12 months of the purchase, we construct a measure of the log-distance in kilometers between the house they sold and the house they bought to proxy for $InformedBuyer_i$ in regression 20.[20] This variable has a mean of about 2.01, and a standard deviation of 1.48. We conjecture that the further these buyers previously lived from the house they are now purchasing, the less likely they are to have information about neighborhood trends. The results are presented in Table 11. The sample size is smaller than for our other regressions, because we do not always find a previous seller with the same name. The balance of homes is bought either by people who were previously renters, or by people moving from outside of Los Angeles county.

Column (1) shows that buyers who previously lived further away buy houses that underperform otherwise similar houses bought by people that lived closer by. This is consistent with agents that lived closer having superior information about characteristics of the neighborhood that allow them to pick better deals. Again, columns (2) to (4) show that the capital gains difference between those houses bought by neighborhood insiders and outsiders is particularly big in bad neighborhoods, i.e. those where the share of informed sellers and the average land share of sold homes is high, and the share of long-tenured sellers is lower.

In this section we provided evidence that buyers who have had past experience in the same zip code, who lived closer by and who are real estate professionals purchase houses that subsequently outperform otherwise similar homes bought by less informed agents. Their superior information seems to allow them to pick better properties, though in the case of real estate agents they might also be better at selling the house and achieve a higher resale price (Levitt and Syverson, 2008). This is consistent with Prediction 3. In addition, and consistent with Prediction 4, this advantage is lower when the predicted neighborhood quality is higher. This provides further evidence for our interpretation of the correlation between seller composition and subsequent price changes to be driven by superior information of the sellers. If seller composition were just a proxy for commonly known information that is not observed by the econometrician, this would not explain why better informed buyers outperform, or why seller composition is less predictive of returns bought by more informed buyers.

[20]For houses bought by an individual with a name that shows up more than once as a seller in the previous 12 months we take the distance to the geographically closest sale. The results are very similar when we pick the average across all observed sales.

Table 11: Effect of Buyer Characteristics - Distance to previous home

	(1)	(2)	(3)	(4)
Log(Distance)	-0.281***	-0.272***	0.0223	-1.090***
	(0.0180)	(0.0299)	(0.0849)	(0.234)
Share Informed Sellers		-6.166***		
		(1.839)		
Log(Distance) × Share Informed Sellers		-0.239		
		(0.626)		
Average Land Share			-19.56***	
			(1.240)	
Log(Distance) × Average Land Share			-0.515***	
			(0.144)	
Share in Zip of Tenure > 3				4.871***
				(0.926)
Log(Distance) × Share in Zip of Tenure > 3				1.060***
				(0.302)
R-squared	0.630	0.630	0.631	0.679
$\bar{y}$	12.82	12.82	12.82	13.73
N	99,472	99,472	99,472	68,972

Note: This table shows results from regression 20. The dependent variable is the annualized capital gain of the house between the two repeat sales. In columns (1) - (3) sales pairs are included when the first sale was after June 1994, in columns (4) when the first sale was after June 1997. All specifications include sales quarter pair fixed effects, zip code fixed effects,characteristics of the property (property size, property age, property type, number of bedrooms and bathrooms, whether the property has a pool or air conditioning), characteristics of the financing (mortgage type, mortgage duration and loan-to-value ratio) and characteristics of the buyer (married, Asian, Latino). Standard errors are clustered at the quarter × zip code level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

5 Conclusion

In many markets, sellers of an asset are better informed than buyers about the true value of the asset. In addition, there might also be information heterogeneity among both buyers and sellers. We argue that residential real estate is an example of this type of market. Sellers are better informed than buyers about both neighborhood characteristics and structural attributes of a house, but among both buyers and among sellers some are better informed than others. We propose a new theoretical framework for empirically analyzing such markets with many heterogeneous assets and differentially informed agents. We then analyze the universe of housing transactions in Los Angeles county between 1994 and 2011 to quantify the impact of this type of asymmetric information on equilibrium market outcomes. We find that changes in the seller composition towards (i) more informed sellers and (ii) sellers with a larger supply elasticity predict subsequent price declines of houses in that neighborhood. This effect is unaffected by the inclusion of past price changes as a control variable, and is larger for houses whose value depends more on neighborhood characteristics.

Importantly, our model and associated equilibrium concept allows us to consider the role of differentially informed buyers. This generates a set of additional predictions that are unique to a model with asymmetric information, and allows us to reject alternative explanations that rely on differential elasticity of reacting to commonly known information, for example due to differential transaction costs. We find that more informed buyers buy houses that experience higher ex post appreciation. Importantly, we also find that the correlation between seller composition and subsequent returns is smaller for houses bought by more informed buyers. Our findings suggest that home owners have superior information about important neighborhood characteristics, and exploit those to time local market movements.

It is well known that asymmetric information can severely undermine the liquidity of markets. Many markets deal with this problem through some combination of regulations such as laws against insider trading and contractual practices such as seller warranties. In real estate markets, legal disclosure requirements and the involvement of real estate agents are intended in part to mitigate the natural information advantage of sellers over buyers. Our results suggest that there remains substantial information asymmetry, involving hard-to-observe features of both neighborhoods and houses, that is immune to these remedies. Furthermore, the differential information is not limited to a difference between buyers and sellers but exists within each of these groups, which creates an advantage for those who are more informed relative to their peers.

While our empirical analysis focuses on the residential real estate market, the information structure we consider is similar in other important financial markets. For example, in the venture capital market the success of start-up firms is a combination of both the promise of their particular industry (e.g. mobile payments, social gaming), as well as the skills of the individual entrepreneurs. Some venture capitalists are better at identifying promising companies (either in promising industries, or with skilled entrepreneurs) than others (Hochberg et al., 2007). Since investment term sheets are usually not publicly disclosed and venture capital investments are indivisible (i.e. it is not possible for other firms to automatically co-invest with more informed VCs at the same terms) less informed venture capitalists cannot learn about the value of individual companies by observing prices paid by more informed investors. Similar empirical tests could be conducted to test for the magnitude of asymmetric information both between venture capital investors as well as between entrepreneurs.

A Theoretical Appendix

A.1 Example

The following example illustrates the features of the equilibrium as well as the content of assumptions (1)-(3). Parameters take the following values:

Parameter	Value
β	0.5
b	0.6
Distribution of θ	0 with probability 0.4 1 with probability 0.6
Distribution of η	0 with probability 0.3 1 with probability 0.7
Distribution of v (resulting from distributions of θ and η)	0 with probability 0.12 0.5 with probability 0.46 1 with probability 0.42
Distribution of ε	$\varepsilon \sim U\,[0.35, 1]$
Number of buyers per type	$n^U = 1$ $n^I = 0.05$

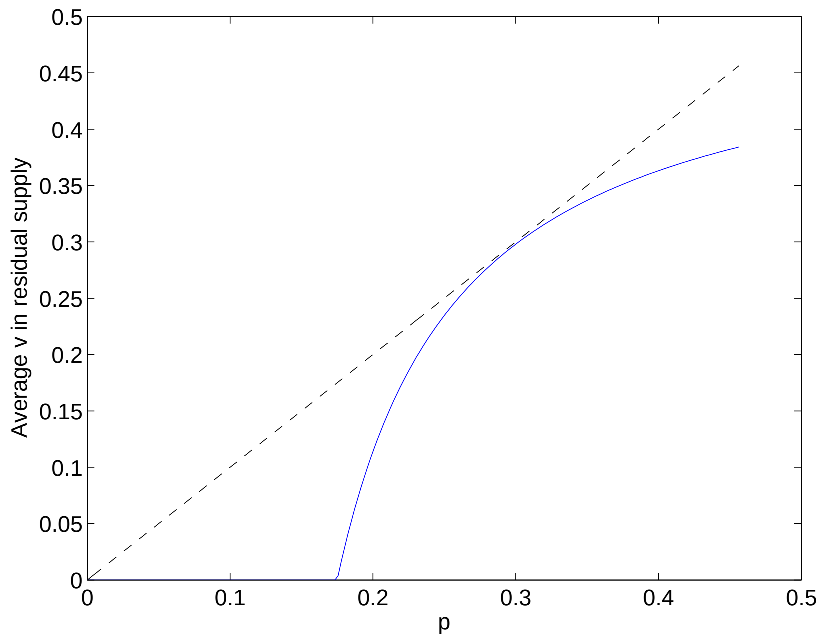
For this example, the equilibrium price given by (10) is $p^* = 0.46$, and quantities supplied and demanded are as follows:

House quality	$S\,(p^*, v)$	Quantity bought		
		Uninformed	Informed	Total
$v = 0$	0.12	0.1	0	0.1
$v = 0.5$	0.4	0.34	0	0.34
$v = 1$	0.07	0.06	0.01	0.07
Total	0.59	0.50	0.01	0.51
Average v	0.46	0.46	1	

At p^*, all the $v = 0$ houses are put on sale, but only some of the higher-quality houses (those whose owners have a low realization of ε). Uninformed buyers buy 0.5 houses, drawn at random. The average quality they obtain is equal to p^*, so they are indifferent to how many houses they buy. Since $n^U = 1 > 0.5$, assumption 2 holds. Since $b = 0.6$, informed buyers only accept houses of $v = 1$, and they all buy houses. Since the supply of good houses is $0.07 > n^I$, assumption 1 holds.

The 0.5 houses bought by uninformed buyers are exactly enough so that all $v = 1$ houses are sold. Instead, some $v = 0$ and $v = 0.5$ houses remain unsold and will be offered on sale at prices below p^* by owners for whom ε is sufficiently low. It remains to check that uninformed buyers do not prefer to buy at those prices instead of at p^*. Figure 3 shows the average quality that uninformed buyers would obtain if they tried to buy at prices $p < p^*$. In all cases, the average quality would be below the price, so these buyers cannot obtain a surplus. Hence, assumption 3 holds and we have an equilibrium.

Figure 3: Would uninformed buyers want to buy from markets with $p < p^*$?



Note: This figure shows the average value of houses that owners would be willing to sell at prices below p^* in the example. In all cases, the average quality is below the price so uninformed buyers would not be willing to buy.

A.2 Example: Tenure and Elasticity

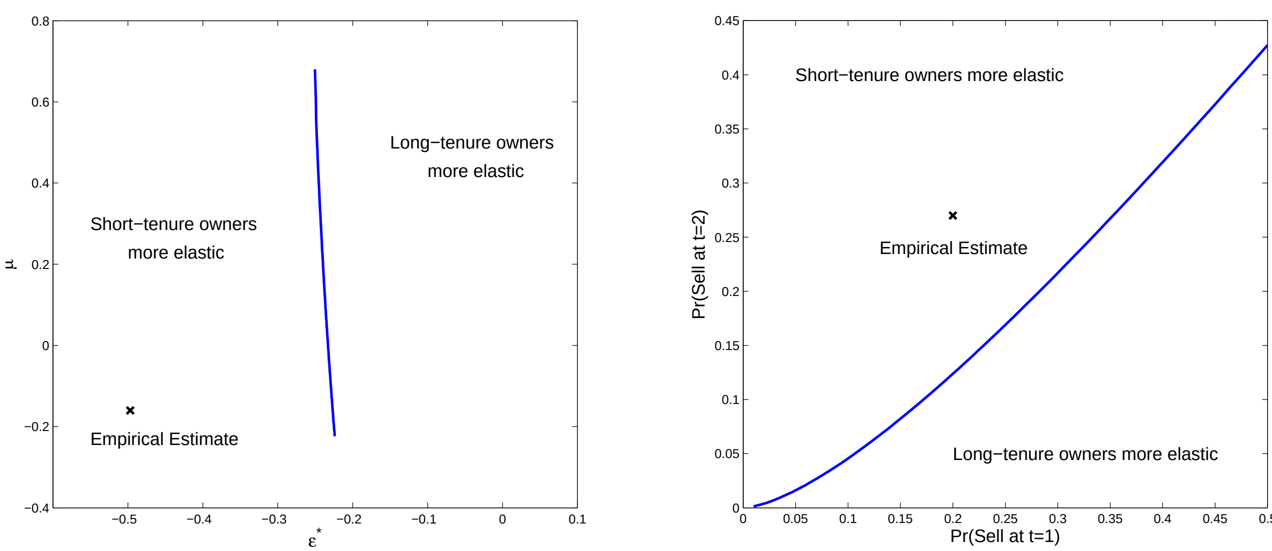
Suppose an owner's potential tenure consists of two periods. In the first period, he receives a shock $x_1 \sim N(\mu, 1)$ and his match-quality is $\varepsilon = m(x_1)$, where $m(\cdot)$ is any continuous monotonic function. Owners sell their house if $\varepsilon \leq \varepsilon^*$.[21] If the owner keeps his house and becomes a long-tenure owner, he receives a second, independent, shock $x_2 \sim N(\mu, 1)$ in the second period. The match quality of a long-term owner is $\varepsilon = m(x_1 + x_2)$. Again, he sells if $\varepsilon \leq \varepsilon^*$.

[21]Abstracting from heterogeneity in house quality.

Figure 4 shows the regions of the parameter space where each group of owners is more elastic. Long-tenure owners are less elastic than short-tenure owners for sufficiently low ε^* (which makes the selection effect weak) and for sufficiently low μ (which leads to a downward drift). A low ε^* implies a low probability of selling in the first period while a low μ leads to higher probabilities of selling in the second period relative to the first. This makes it possible to delimit the frontiers between the two regions in terms of the probability with which owners sell in each of the periods, as shown in the right panel.

In our empirical section, we divide owners into short and long tenure depending on whether they have been in their house longer than three years. In our sample, the proportion of owners who sell their house within three years is 0.2. For longer-tenure owners the match between the model is less clear, since the model has a finite horizon and in reality ownership is open-ended. However, we can compute the relative hazard rate of selling for owners who have been in their house more or less than three years. In a two-period model, this translates directly into a relative probability of selling. Since in our sample the selling hazard of long-tenure owners is 1.35 times that of short-tenure owners, this would correspond to a selling probability of 0.27. Probabilities of selling of 0.2 and 0.27 respectively, which result from $\varepsilon^* = -0.5$ and $\mu = -0.16$ are well within the region where the long-tenure owners are predicted to have less elastic supply.

Figure 4: Do longer-tenure owners have less elastic supply?



Note: This figure shows the regions of the parameter space where longer-tenure owners have less elastic supply, together with the combination of parameters that best matches the empirical data.

A.3 Proofs

Proposition 1. *Suppose that* $\frac{g_B(\varepsilon)}{G_B(\varepsilon)} \geq \frac{g_A(\varepsilon)}{G_A(\varepsilon)}$ *for every* $\varepsilon \geq \frac{p^*}{\bar{v}}$. *Then the proportion of sellers who belong to group* A *among sellers in neighborhood* j *is increasing in* θ_j.

Proof. Conditional on house quality v, house supply from group A and group B respectively are

$$S_A(p^*, v) = \pi_A G_A\left(\frac{p^*}{v}\right) f(v) \quad \text{and } S_B(p^*, v) = \pi_B G_B\left(\frac{p^*}{v}\right) f(v)$$

so the proportion of group A among sellers of houses of quality v is

$$\pi_{A|Sell}(v) = \frac{\pi_A G_A\left(\frac{p^*}{v}\right)}{\pi_A G_A\left(\frac{p^*}{v}\right) + \pi_B G_B\left(\frac{p^*}{v}\right)}$$

Taking the derivative with respect to v:

$$\begin{aligned}
\frac{d\pi_{A|Sell}(v)}{dv} &= \frac{-\pi_A g_A\left(\frac{p^*}{v}\right)\frac{p^*}{v^2}\left[\pi_A G_A\left(\frac{p^*}{v}\right) + \pi_B G_B\left(\frac{p^*}{v}\right)\right] + \left[\pi_A g_A\left(\frac{p^*}{v}\right)\frac{p^*}{v^2} + \pi_B g_B\left(\frac{p^*}{v}\right)\frac{p^*}{v^2}\right]\pi_A G_A\left(\frac{p^*}{v}\right)}{\left[\pi_A G_A\left(\frac{p^*}{v}\right) + \pi_B G_B\left(\frac{p^*}{v}\right)\right]^2} \\
&= \frac{-\pi_A g_A\left(\frac{p^*}{v}\right)\frac{p^*}{v^2}\left[\pi_B G_B\left(\frac{p^*}{v}\right)\right] + \left[\pi_B g_B\left(\frac{p^*}{v}\right)\frac{p^*}{v^2}\right]\pi_A G_A\left(\frac{p^*}{v}\right)}{\left[\pi_A G_A\left(\frac{p^*}{v}\right) + \pi_B G_B\left(\frac{p^*}{v}\right)\right]^2} \\
&= \frac{\pi_A \pi_B \frac{p^*}{v^2} G_A\left(\frac{p^*}{v}\right) G_B\left(\frac{p^*}{v}\right)}{\left[\pi_A G_A\left(\frac{p^*}{v}\right) + \pi_B G_B\left(\frac{p^*}{v}\right)\right]^2}\left[\frac{g_B\left(\frac{p^*}{v}\right)}{G_B\left(\frac{p^*}{v}\right)} - \frac{g_A\left(\frac{p^*}{v}\right)}{G_A\left(\frac{p^*}{v}\right)}\right] > 0 \qquad (21)
\end{aligned}$$

The proportion of group A among sellers in a neighborhood where the neighborhood shock is θ will be

$$\pi_{A|Sell}(\theta) = \int \pi_{A|Sell}(\beta\theta + (1-\beta)\eta)\, dF_\eta(\eta)$$

Taking the derivative with respect to θ and using (21):

$$\frac{d\pi_{A|Sell}(\theta)}{d\theta} = \int \beta\left(\left.\frac{d\pi_{A|Sell}(v)}{dv}\right|_{v=\beta\theta+(1-\beta)\eta}\right) dF_\eta(\eta) > 0$$

□

Proposition 2. *The proportion of informed among sellers is higher in the worst neighborhood* $(\theta = 0)$ *than in the best neighborhood* $\left(\theta = \bar{\theta}\right)$

Proof. Given θ and η, the fraction of informed owners who choose to sell is

$$\Pr\left[\varepsilon \leq \frac{p^*}{\beta\theta + (1-\beta)\eta}|\theta, \eta\right] = G\left(\frac{p^*}{\beta\theta + (1-\beta)\eta}\right)$$

so integrating across η, the fraction of informed owners who choose to sell in a neighborhood of quality θ is

$$\Pr\left[\text{Sell}|\theta, \text{Informed}\right] = \int G\left(\frac{p^*}{\beta\theta + (1-\beta)\eta}\right) dF_\eta(\eta) \tag{22}$$

Similarly, for uninformed sellers,

$$\Pr\left[\text{Sell}|\hat{\theta}(x), \text{Uninformed}\right] = \int G\left(\frac{p^*}{\beta\hat{\theta}(x) + (1-\beta)\eta}\right) dF_\eta(\eta)$$

so integrating across realizations of x:

$$\Pr\left[\text{Sell}|\theta, \text{Uninformed}\right] = \int\int G\left(\frac{p^*}{\beta\hat{\theta}(x) + (1-\beta)\eta}\right) dF_\eta(\eta)\, dF_{x|\theta}(x) \tag{23}$$

For any non-degenerate distribution $F_{x|\theta}$, $0 < \hat{\theta}(x) < \bar{\theta}$ for all x. Equations (22) and (23) then imply that $\Pr\left[\text{Sell}|0, \text{Informed}\right] > \Pr\left[\text{Sell}|0, \text{Uninformed}\right]$ and $\Pr\left[\text{Sell}|\bar{\theta}, \text{Informed}\right] < \Pr\left[\text{Sell}|\bar{\theta}, \text{Uninformed}\right]$, which gives the result. □

Proposition 3. *1. Assume $\bar{\theta} \geq \bar{\eta}$. Then the proportion of owners who choose to sell is increasing in β_h in the worst neighborhood ($\theta = 0$) and decreasing in β_h in the best neighborhood ($\theta = \bar{\theta}$).*

2. The proportion of owners who choose to sell in a neighborhood of quality θ does not change with θ for houses with $\beta_h = 0$ and decreases with θ for houses with $\beta_h = 1$.

Proof. Owners sell their house if $\varepsilon \leq \frac{p^*}{\beta_h\theta + (1-\beta_h)\eta}$, so the proportion of owners of houses with β_h who sell in a neighborhood of quality θ is:

$$\Pr\left[\text{Sell}|\theta, \beta_h\right] = \int G\left(\frac{p^*}{\beta_h\theta + (1-\beta_h)\eta}\right) dF_\eta(\eta) \tag{24}$$

□

1. Taking the derivative of (24) with respect to β_h:

$$\frac{d\Pr\left[Sell|\theta, \beta_h\right]}{d\beta_h} = \int g\left(\frac{p^*}{\beta_h\theta + (1-\beta_h)\eta}\right) \frac{p^*}{\left[\beta_h\theta + (1-\beta_h)\eta\right]^2} (\eta - \theta)\, dF_\eta(\eta)$$

For $\theta = 0$ and $\theta = \bar{\theta}$ respectively, this reduces to

$$\left.\frac{d\Pr\left[Sell|\theta,\beta_h\right]}{d\beta_h}\right|_{\theta=0} = \int g\left(\frac{p^*}{\left(1-\beta_h\right)\eta}\right)\frac{p^*}{\left[\left(1-\beta_h\right)\eta\right]^2}\eta dF_\eta\left(\eta\right) > 0$$
$$\left.\frac{d\Pr\left[Sell|\theta,\beta_h\right]}{d\beta_h}\right|_{\theta=\bar{\theta}} = \int g\left(\frac{p^*}{\beta_h\bar{\theta}+\left(1-\beta_h\right)\eta}\right)\frac{p^*}{\left[\beta_h\bar{\theta}+\left(1-\beta_h\right)\eta\right]^2}\left(\eta-\bar{\theta}\right)dF_\eta\left(\eta\right) < 0$$

2. *Taking the derivative of (24) with respect to θ:*

$$\frac{d\Pr\left[Sell|\theta,\beta_h\right]}{d\theta} = -\int g\left(\frac{p^*}{\beta_h\theta+\left(1-\beta_h\right)\eta}\right)\frac{\beta_h p^*}{\left[\beta_h\theta+\left(1-\beta_h\right)\eta\right]^2}dF_\eta\left(\eta\right)$$

For $\beta_h = 0$ and $\beta_h = 1$ respectively, this reduces to

$$\left.\frac{d\Pr\left[Sell|\theta,\beta_h\right]}{d\theta}\right|_{\beta_h=0} = 0$$
$$\left.\frac{d\Pr\left[Sell|\theta,\beta_h\right]}{d\theta}\right|_{\beta_h=1} = -g\left(\frac{p^*}{\theta}\right)\frac{p^*}{\theta^2} < 0$$

Proposition 6. *Assume $\beta\bar{\theta} > b$. Then, conditional on buying in a sufficiently good neighborhood, the expected value of houses bought by informed and uninformed buyers is the same.*

Proof. The expected house quality obtained by an informed buyer conditional on buying a house in a neighborhood of quality θ is

$$\bar{v}^I\left(\theta\right) = \frac{\int_b^{\bar{v}} vG\left(\frac{p^*}{v}\right)f_{v|\theta}\left(v\right)dv}{\int_b^{\bar{v}} G\left(\frac{p^*}{v}\right)f_{v|\theta}\left(v\right)dv}$$

For $\theta > \frac{b}{\beta}$, $f_{V|\theta}(v) = 0$ for all $v < b$ and therefore

$$\bar{v}^I\left(\theta\right) = \frac{\int_0^{\bar{v}} vG\left(\frac{p^*}{v}\right)f_{v|\theta}\left(v\right)dv}{\int_0^{\bar{v}} G\left(\frac{p^*}{v}\right)f_{v|\theta}\left(v\right)dv} = \bar{v}^U\left(\theta\right)$$

□

B Empirical Appendix

B.1 Data Cleaning

Arms-length Transactions: The procedure to identify arms-length transactions follows Stroebel (2013). We identify all deeds that contain information about arms-length transactions in which both buyer and seller act in their best economic interest. This ensures that transaction prices reflect the market value of the property. We include all deeds that are one of the following: "Grant Deed," "Condominium Deed," "Individual Deed," "Warranty Deed," "Joint Tenancy Deed," "Special Warranty Deed," "Limited Warranty Deed" and "Corporation Deed." This excludes intra-family transfers and foreclosures. We drop all observations that are not a Main Deed or only transfer partial interest in a property. We also drop properties with transaction prices of less than $25,000 and more than $10,000,000.

Death of Owner: We identify those repeat sales pairs for which we observe a death of the owners up to twelve months before the second sale ("forced moves"). The death of an owner is identified if either (i) the seller on a deed is classified as an "estate", "executor", "deceased" or "surviving joint owner" or (ii) if we observe one of the following: "Affidavit of Death of Joint Tenant" or "Executor's Deed."

B.2 Control Variables

Table 5 shows summary statistics for the control variables used in the regressions. Most of these controls are not included linearly in the regression, but by splitting them into groups of values represented by dummy variables. This allows for a more flexible functional form. The results are not sensitive to the exact definition of groups.

House Characteristics: Building size is controlled for by adding dummy variables for 10 equally sized groups. To control for the number of bedrooms and bathrooms, we add a dummy variable for each possible value. We construct the age of the property by subtracting the construction year of the house from the year of sale. Controls for age are included by including four equally sized buckets. We include an "investment property" dummy for properties that are identified as such in the assessor data.

Buyer and Financing Characteristics: We control for whether the buyer is a single individual or a married couple, which is reported in the deeds data. We also control for whether the buyer is Asian or Latino. While this information is not provided in the deeds records, we do observe the names of the buyers. We match the surnames of buyers to the 1000 most common Asian and Latino surnames from the 2000 U.S. Census to build an "Asian" and "Latino" indicator variable. The loan-to-value (LTV) ratio is included by

Figure 5: Summary Statistics - Control Variables

	Mean	Standard Deviation	P10	P50	P90
Property Characteristics					
Condo (binary)	0.26	0.44	0	0	1
Building Area (sqft)	1687.9	885.6	924	1465	2710
Bedrooms (#)	3.07	1.20	2	3	4
Bathrooms (#)	2.31	1.06	1	2	3
Age of Building (years)	36.8	23.7	6	38	72
Pool (binary)	0.24	0.43	0	0	1
AC (binary)	0.45	0.50	0	0	1
Investment Property (binary)	0.07	0.25	0	0	0
Buyer Characteristics					
Buyers Married (binary)	0.48	0.50	0	0	1
Buyers Asian (binary)	0.10	0.30	0	0	1
Buyers Latino (binary)	0.25	0.43	0	0	1
Financing Characteristics					
Loan-To-Value Ratio	0.84	0.14	1	1	1
Mortgage Duration (years)	29.5	3.5	30	30	30
VA Mortgage (binary)	0.01	0.10	0	0	0
FHA Mortgage (binary)	0.17	0.38	0	0	1

Note: This table shows summary statistics for the control variables included in our regressions.

dummy variables for mortgages with an LTV $\leq$ 80%, between 80% and 90%, between 90% and 97%. and $>$ 97%. We also control for the duration of the mortgage, and whether it is a VA or FHA-insured mortgage.

B.3 Measuring Land Share

We calculate the share of each property's total value made up of land from data in the tax assessment records. In particular, tax assessors report a separate valuation of both the land and "improvements". Improvements include all assessable buildings and structures on the land. Figure 6 shows an example of two neighboring homes in Los Angeles county. The two houses sit on identically-sized lots. The southern-most house, however, has a larger structure built on it, and thus the share of land in overall value as reported in the tax assessment records is lower.

B.4 Robustness I: Non-random selection into observing repeat sales

One might be worried that the subsample of houses for which we observe a resale is not representative of all homes in a particular neighborhood, and that such a selection might lead us to incorrectly estimate the true correlation between seller composition and the average

Figure 6: Land Share - Example

Note: This figure shows an example of the land share calculated for two properties in Los Angeles county.

capital gain of homes in the neighborhood.[22] To address such concerns, in Table 12 we show results from regression 16 similar to Table 3, but restrict the sample to sales pairs where the second sale is precipitated by a plausibly exogenous event. In particular, we only look at those repeat sales pairs where we observe the death of the original owners in the 12 months preceding the resale.[23] We argue that such sales are more plausibly prompted by the observed death than by other factors such as the value of the house. The results show that the correlation between neighborhood seller composition and subsequent capital gains is of the same magnitude in the sample of forced moves as it is in the entire sample, and even somewhat larger for the share of informed sellers. This suggests that selection into observing repeat sales does not significantly bias our estimates.

[22]For example, this might be a problem if owners who experience high idiosyncratic capital gains on their house are more likely to sell and thus enter our sample more often than other owners in the same neighborhood. If seller composition affected the probability of these idiosyncratic capital gains events, this would lead us to overestimate the correlation between seller composition and average house price movements.

[23]These events can be identified in the deeds data as described in Appendix B.

Table 12: Effect of Seller Composition on Capital Gains - Forced Moves

	(1)	(2)	(3)	(4)
Share Informed Sellers	-8.970**			-6.607*
	(2.868)			(3.549)
Average Seller Land Share		-19.90***		-19.18***
		(2.592)		(3.329)
Share in Zip of Tenure > 3			5.697***	4.419***
			(1.446)	(1.443)
Fixed Effects	✓	✓	✓	✓
House and Buyer Controls	✓	✓	✓	✓
R-squared	0.601	0.602	0.621	0.623
$\bar{y}$	13.98	13.98	15.15	15.15
N	17,605	17,605	13,451	13,451

Note: This table shows results from regression **??** for those transactions where the resale was preceded in the 12 months before by a death of the owner. The dependent variable is the annualized capital gain of the house between the two repeat sales. All specifications include sales quarter pair fixed effects and zip code fixed effects and control for characteristics of the property (property size, property age, property type, number of bedrooms and bathrooms, whether the property has a pool or air conditioning), characteristics of the financing (mortgage type, mortgage duration and loan-to-value ratio) and characteristics of the buyer (married, Asian, Latino). Standard errors are clustered at the quarter × zip code level. Columns (1) - (2) include sales pair where the first sale was after June 1994, column (3) includes sales pairs where the first sale was after June 1997. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

B.5 Robustness II: Presence of "Flippers"

One concern might be that our measure of average seller tenure does not, in fact, pick up owners that are moving out of the neighborhood, but instead picks up the share of "flippers" among the sellers, who buy houses to resell them quickly at a profit, often after substantial remodeling or renovation. This could bias our results in either direction, depending on what kinds of neighborhoods tend to attract more flippers. If flippers are more active in overpriced neighborhoods (perhaps because they are trying to time the housing market), then this could drive the correlation we observe in the data: high flipper activity would show up as a larger share of sellers that have a short ownership-tenure and would predict low subsequent appreciation.

To rule out that this is the main driver of the observed correlation, we identify a set of transactors that we classify as flippers, and then repeat the analysis above by only calculating the average tenure amongst those sellers not identified as flippers. Similar to Bayer et al. (2011), we use the fact that the deeds data records the name of buyers and sellers to classify transactors as flippers. We apply three classification rules. Our first two rules classify an agent as a flipper if someone with that name has engaged in at least 3 transactions over the sample period, with more than 30% (40%) of them being bought and resold within 2 years. Our third rule excludes all transactors that are classified as companies, since some flippers might buy and sell homes through incorporated entities. Table 13 shows that the results are very robust to only considering the average tenure of sellers that are not classified as flippers.

Table 13: Remove Possible Flippers from Tenure

	(1)	(2)	(3)	(4)
Share in Zip of Tenure > 3	6.808***	5.455***	5.466***	6.528***
	(0.377)	(0.348)	(0.348)	(0.377)
Restriction	None	> 2 Trans. > 30% within 2y	> 2 Trans. > 40% within 2y	No companies
Fixed Effects, House and Buyer Controls	✓	✓	✓	✓
R-squared	0.658	0.658	0.658	0.658
$\bar{y}$	13.70	13.70	13.70	13.70
N	300,106	300,104	300,104	300,076

Note: This table shows results from regression ??. The dependent variable is the annualized capital gain of the house between the two repeat sales. In column (1) tenure is measured amongst all sellers, in column (2) we exclude sellers that have more than two transactions and at least 30% of them were resales within 2 years, in column (3) we exclude sellers with at least two transactions of which at least 40% are resold within 2 years. In column (4) we exclude all sales by sellers identified as companies. All specifications include sales quarter pair fixed effects, zip code fixed effects, characteristics of the property (property size, property age, property type, number of bedrooms and bathrooms, whether the property has a pool or air conditioning), characteristics of the financing (mortgage type, mortgage duration and loan-to-value ratio) and characteristics of the buyer (married, Asian, Latino). Standard errors are clustered at the quarter × zip code level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

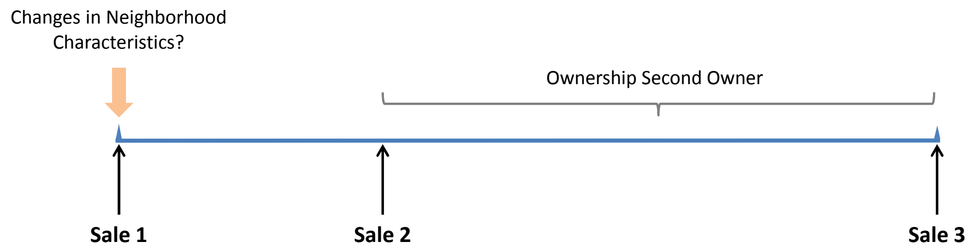
B.6 Robustness III: Ownership Period of Second Buyer

The model in section 1 has only two periods and assumes that all the private information is revealed by the time a buyer sells the house. In reality, information is likely to be revealed gradually over time. If so, then one should expect some of it to be revealed only after the buyer has resold the house and thus affect the appreciation experienced by subsequent owners. To test for this, we need to observe at least three arms-length transactions of the house. We calculate the appreciation between the last two sales, as shown in Figure 7, and determine to what degree this is predicted by the seller composition at the time of the first sale. In other words, we run regression 25, where q_1, q_2 and q_3 represent the calendar quarters of the first, second and third sale.

$$CapitalGain_{i,n,q_2,q_3} = \alpha + \beta_1 SellerComposition_{n,q_1} + \chi_{q_1} + \phi_{q_2,q_3} + \xi_n + X_i'\beta_2 + \epsilon_i \quad (25)$$

Table 14 shows the results separately for samples where we allow up to fours years and up to six years between sale one and sale two. For longer time horizons between the first and

Figure 7: Measures of Second Buyer's capital gain



second sale more of the initially unobservable neighborhood characteristics will have been revealed, which leaves less scope for initial seller composition to predict additional differential capital gain.

We can see that following an increase in the share of informed sellers, houses in that neighborhood continue to underperform, even during the ownership period of subsequent. The effect is smaller the more time has passed between the first and second transaction, and the more of the initially private information of the first owners will have been revealed and taken into account by the second buyer. Similar effects can be seen for changes in our other two measures of seller composition, the average land share of sold homes and the average tenure of sellers.

B.7 Predicting Future Neighborhood Characteristics

In Tables 6 and 7 we showed that the seller composition did not only predict future capital gains, but also contemporaneous demographic developments that might not be perfectly observable to many buyers. In this section we test whether the seller composition in addition predicts future demographics, controlling for current demographics and past capital gains, as well as zip code and year fixed effects.

$$
\begin{aligned}
ZipCode\_Demo_{n,y+1} &= \alpha + \beta_1 SellerComposition_{n,y} + \beta_2 PastCapGain_{n,q_1} + \\
&\quad \beta_3 ZipCode\_Demo_{n,y} + \xi_n + \phi_y + \epsilon_i
\end{aligned} \tag{26}
$$

Table 15 presents the results for the zip code level demographic measures introduced in section 4.4: The share of African Americans among new mortgage applicants, the average income of new mortgage applicants and the share of socio-economically disadvantaged children in schools in that zip code. Unsurprisingly, current demographics are correlated with

Table 14: Effect of Seller Composition on Capital Gains during Second Ownership Period

	(1)	(2)	(3)	(4)	(5)	(6)
Share Informed Sellers	-8.813***	-9.363***				
	(2.895)	(2.253)				
Land Share			-21.65***	-19.04***		
			(2.652)	(2.033)		
Share in Zip of Tenure > 3					3.425**	2.456**
					(1.237)	(1.030)
Fixed Effects	✓	✓	✓	✓	✓	✓
House and Buyer Controls	✓	✓	✓	✓	✓	✓
Max. Time between Sales 1&2	4 Years	6 Years	4 Years	6 Years	4 Years	6 Years
R-squared	0.557	0.563	0.558	0.563	0.597	0.607
$\bar{y}$	10.99	11.17	10.99	11.17	10.83	10.51
N	58747	82996	58747	82996	48161	63521

Note: This table shows results from regression 25. The dependent variable is the annualized capital gain of the house between the two repeat sales. The seller composition variables are measured at the quarter × zip code level. All specifications include fixed effects for the sales quarter pair, the quarter of initial sale and the zip code. Standard errors are clustered at the initial quarter × zip code level. All specifications include characteristics of the property (property size, property age, property type, number of bedrooms and bathrooms, whether the property has a pool or air conditioning), characteristics of the financing (mortgage type, mortgage duration and loan-to-value ratio) and characteristics of the first buyer (married, Asian, Latino). Columns (3) and (6) include sales pairs where the first sale was after June 1997, all other columns include sales pair where the first sale was after June 1994. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

future demographics. In addition, we can see that current seller composition is related to future demographics in the zip code, even if not all specifications are statistically significant.

Table 15: Seller Characteristics and Future Demographics

	African-American Applicants		Average Applicant Income		Socio-Economically Disadvantaged Students	
	$t+1$	$t+2$	$t+1$	$t+2$	$t+1$	$t+2$
	(1)	(2)	(3)	(4)	(5)	(6)
Appreciation Past Year	-0.012**	-0.001	-21.82***	0.875	-0.996	-3.301
	(0.005)	(0.005)	(5.845)	(6.552)	(1.968)	(2.423)
Appreciation Past Two Years	0.012***	0.008***	-5.00	-17.83***	-0.816	-0.183
	(0.003)	(0.003)	(3.689)	(4.099)	(1.213)	(1.521)
Share of African-American Applicants	0.453***	0.311***				
	(0.015)	(0.017)				
Average Applicant Income			0.509***	0.320***		
			(0.014)	(0.0154)		
Share of Socio-Economically Disadvantaged Students					0.581***	0.382***
					(0.017)	(0.020)
Share Informed Sellers	0.069***	0.041***	-27.45	1.221	3.086	-3.613
	(0.014)	(0.015)	(16.71)	(18.96)	(5.417)	(6.253)
Average Seller Land Share	0.017**	0.050***	-22.31**	-11.47	-3.837	-0.334
	(0.008)	(0.009)	(9.596)	(11.41)	(3.247)	(3.811)
Share in Zip of Tenure > 3	-0.002	-0.000	8.158	2.641	-4.198**	-5.059**
	(0.005)	(0.006)	(5.927)	(7.065)	(1.920)	(2.210)
Fixed Effects (Zip Code and Year)	✓	✓	✓	✓	✓	✓
R-squared	0.977	0.975	0.959	0.954	0.979	0.975
$\bar{y}$	0.0526	0.0526	131.8	134.7	60.41	60.50
N	3,872	3,607	3,872	3,607	2,828	2,571

Note: This table shows results from regression 26. Instead of using contemporaneous zip code demographics as the dependent variable (see Tables 6 and 7), we use the zip code demographics one year (odd columns) or two years (even columns) in the future. The dependent variables are the share of African-American mortgage applicants (columns 1 - 2), and the average income of all mortgage applicants (columns 3-4) and the share of all students that are classified as socio-economically disadvantaged. The unit of observation is zip code by year. Columns (1) and (3) cover the years 1996 - 2010, columns (2) and (4) the years 1996 - 2009. Columns (5) covers 2000 - 2010, column (6) covers 2000 - 2009. Each specification includes zip code and year fixed effects. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

References

Akerlof, G.A., "The market for "lemons": Quality uncertainty and the market mechanism," *Quarterly Journal of Economics*, 1970, pp. 488–500.

Albouy, D., "What are Cities Worth? Land Rents, Local Productivity, and the Value of Amenities," *NBER Working Paper*, 2009, *14981.*

Altonji, J.G., T.E. Elder, and C.R. Taber, "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Arnott, R.J. and J.E. Stiglitz, "Aggregate land rents, expenditure on public goods, and optimal city size," *Quarterly Journal of Economics*, 1979, *93* (4), 471–500.

Bayer, Patrick, Christopher Geissler, and James W Roberts, "Speculators and middlemen: The role of flippers in the housing market," Technical Report, National Bureau of Economic Research 2011.

Brunnermeier, Markus and Stefan Nagel, "Hedge funds and the technology bubble," *The Journal of Finance*, 2004, *59* (5), 2013–2040.

Case, Karl E and Robert J Shiller, "The Efficiency of the Market for Single-Family Homes," *The American Economic Review*, 1989, *79* (1), 125–137.

_ and _ , "Forecasting prices and excess returns in the housing market," *Real Estate Economics*, 1990, *18* (3), 253–273.

Cheng, Ing-Haw, Sahil Raina, and Wei Xiong, "Wall Street and the Housing Bubble," Technical Report, National Bureau of Economic Research 2013.

Chinco, Alex and Chris Mayer, "Distant Speculators and Asset Bubbles in the Housing Market," 2013.

Choi, James J, Li Jin, and Hongjun Yan, "Informed trading and expected returns," Technical Report 2013.

Cohen, Lauren, Andrea Frazzini, and Christopher James Malloy, "The Small World of Investing: Board Connections and Mutual Fund Returns," *Journal of Political Economy*, 2008, *116* (5), 951–979.

_ , _ , and Christopher Malloy, "Sell-Side School Ties," *The Journal of Finance*, 2010, *65* (4), 1409–1437.

Coval, Joshua D and Tobias J Moskowitz, "The geography of investment: Informed trading and asset prices," *Journal of Political Economy*, 2001, *109* (4), 811–841.

Davis, Morris A. and Jonathan Heathcote, "The price and quantity of residential land in the United States," *Journal of Monetary Economics*, 2007, *54* (8), 2595–2620.

Dubey, Pradeep and John Geanakoplos, "Competitive Pooling: Rothschild-Stiglitz Reconsidered," *Quarterly Journal of Economics*, 2002, *117* (4), 1529–1570.

Easley, David, Soeren Hvidkjaer, and Maureen Ohara, "Is information risk a determinant of asset returns?," *The Journal of Finance*, 2002, *57* (5), 2185–2221.

Finnerty, Joseph E, "Insiders activity and inside information: A multivariate analysis," *Journal of Financial and Quantitative Analysis*, 1976, *11* (2), 205–215.

Gale, Douglas, "A Walrasian Theory of Markets with Adverse Selection," *Review of Economic Studies*, April 1992, *59* (2), 229–55.

_ , "Equilibria and Pareto Optima of Markets with Adverse Selection," *Economic Theory*, 1996, *7* (2), 207–235.

Garmaise, Mark J. and Tobias J. Moskowitz, "Confronting information asymmetries: Evidence from real estate markets," *Review of Financial Studies*, 2004, *17* (2), 405–437.

Ghysels, Eric, Alberto Plazzi, Rossen Valkanov, and Walter Torous, "Chapter 9 - Forecasting Real Estate Prices," in Graham Elliott and Allan Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 2, Part A of *Handbook of Economic Forecasting*, Elsevier, 2013, pp. 509 – 580.

Grossman, Sanford J and Joseph E Stiglitz, "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, June 1980, *70* (3), 393–408.

Guerrieri, Veronica, David Hartley, and Erik Hurst, "Endogenous gentrification and housing price dynamics," *NBER Working Paper*, 2010, *16237.*

_ , Robert Shimer, and Randall Wright, "Adverse Selection in Competitive Search Equilibrium," *Econometrica*, November 2010, *78* (6), 1823–1862.

Hellwig, Martin, "Some recent developments in the theory of competition in markets with adverse selection," *European Economic Review*, 1987, *31* (1-2), 319–325.

Hochberg, Yael V, Alexander Ljungqvist, and Yang Lu, "Whom you know matters: Venture capital networks and investment performance," *The Journal of Finance*, 2007, *62* (1), 251–301.

Jaffe, Jeffrey F, "Special information and insider trading," *The Journal of Business*, 1974, *47* (3), 410–428.

Kelly, Bryan and Alexander Ljungqvist, "Testing asymmetric-information asset pricing models," *Review of Financial Studies*, 2012, *25* (5), 1366–1413.

Kurlat, Pablo, "Asset Markets with Heterogeneous Information," 2012. Stanford University Working Paper.

Kyle, Albert S, "Continuous Auctions and Insider Trading," *Econometrica*, November 1985, *53* (6), 1315–35.

Levitt, S.D. and C. Syverson, "Market distortions when agents are better informed: The value of information in real estate transactions," *Review of Economics and Statistics*, 2008, *90* (4), 599–611.

Lin, Ji-Chai and John S Howe, "Insider trading in the OTC market," *The Journal of Finance*, 1990, *45* (4), 1273–1284.

Lorie, James H and Victor Niederhoffer, "Predictive and statistical properties of insider trading," *JL & Econ.*, 1968, *11*, 35.

Seyhun, H Nejat, "Insiders' profits, costs of trading, and market efficiency," *Journal of Financial Economics*, 1986, *16* (2), 189–212.

_ , "Why does aggregate insider trading predict future stock returns?," *The Quarterly Journal of Economics*, 1992, *107* (4), 1303–1331.

Stroebel, Johannes, "Asymmetric information about collateral values," *Working Paper*, 2013.

Temin, Peter and Hans-Joachim Voth, "Riding the south sea bubble," *American Economic Review*, 2004, *94* (5).

Wilson, Charles, "The Nature of Equilibrium in Markets with Adverse Selection," *The Bell Journal of Economics*, 1980, *11* (1), 108–130.