Knowing your Neighborhood: Asymmetric Information in Real Estate Markets*

Pablo Kurlat
Stanford University

Johannes Stroebel
NYU Stern

HIGHLY PRELIMINARY - COMMENTS VERY WELCOME

Abstract

We study equilibrium asset market outcomes when there is heterogeneity in information about asset values among both buyers and sellers. We argue that in residential real estate markets a key source of information heterogeneity pertains to hard-to-observe characteristics of the neighborhood that are subsequently capitalized in land prices. Sellers are usually better informed about neighborhood characteristics than buyers, but there are some sellers and some buyers that are better informed than their peers. Consistent with theoretical predictions, we find that changes in the seller composition towards (i) more informed sellers and (ii) sellers with a larger supply elasticity with respect to neighborhood trends predict subsequent price declines of houses in that neighborhood. This effect is larger for homes with a bigger neighborhood-β, and smaller for homes bought by more informed home buyers.

In many markets, different 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 there exists important additional heterogeneity in the information sets of both buyers and sellers.

We argue that the residential real estate market constitutes such a market with heterogenous assets and differentially informed buyers and sellers. Transaction prices for properties include payments for both the land and the structure, both of which might be hard to value and are plausibly the source of asymmetric information between different market participants. For example, the value of a house's structure includes hard-to-observe aspects of construction quality (Stroebel, 2012). Similarly, local amenities such as crime rates or school

*June 2013. We thank Theresa Kuchler, Matteo Maggiori, Monika Piazzesi and Amit Seru, our discussant Chris Parsons, as well as seminar participants at Stanford University, Chicago Booth and the 2013 Summer Real Estate Symposium. We thank Trulia for providing data.

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). On average, home sellers are likely to have better information than potential buyers about both neighborhood and house characteristics. In addition, however, some fraction of the possible buyers or sellers might have an information advantage relative to their peers. For example, real estate agents living in a neighborhood might be particularly well informed about neighborhood trends and demographics, and buyers who have previously lived in a particular neighborhood face less of an information disadvantage relative to buyers who are moving from further away.

In this paper we show that neighborhood characteristics do, in fact, constitute a significant source of asymmetric information between market participants, and argue that this information structure has important implications for equilibrium market outcomes. To motivate our empirical procedure for identifying asymmetric information, we propose a theoretical framework that allows us to formalize the implications of different information structures. An agent's valuation of a property depends on characteristics of both the neighborhood and the structure. Current home owners can obeserve these characteristics for their own property, but the valuation of their 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. Some buyers are able to observe an informative signal about the property's value; more expert buyers get better signals. Prices and trading are determined in competitive equilibrium as in Kurlat (2012). In this setting, differentially informed buyers choose houses of differential quality, while houses that are observationally equivalent to the least informed buyers trade at the same price.

The model has several predictions. First, the characteristics of sellers in a neighborhood should predict subsequent price changes for houses in that neighborhood. The decision to sell will be more responsive to hard-to-observe neighborhood characteristics for some groups of owners than for others; consequently, the proportion of more-reponsive owner types among sellers is indicative of these characteristics. Since they are hard to observe, they are not immediately reflected in prices and thus result in subsequent appreciation. Hence, changes in the composition of sellers should predict subsequent price changes in that neighborhood. 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. We begin by showing that the share of total value assigned by the tax assessor to the land component of each property is a good measure of a property's neighborhood-β. That is, the prices of properties with a larger land share in total value follow average price changes in a neighborhood more closely. We then document that average neighborhood price appreciation responds to changes in the composition of sellers as predicted by the model. We argue that real estate professionals living in a neighborhood should be particularly well informed about changes in neighborhood quality, and should thus be particularly elastic in their decision to sell when neighbood characteristics change. Consistent with this, we find that increases in the share of real estate professionals amongst sellers in a neighborhood predicts future price declines. Second, we propose that owners of houses with higher neighborhood-β should respond more elastically to changes in neighborhood characteristics. We show that, indeed, when the average neighborhood-β of houses sold in a neighborhood increases, this is predictive of future price declines. Finally, we argue that longer-tenure residents are less elastic in their decision to move, and show that an increase in the share of sellers who had only recently moved into the neighborhood predicts negative neighborhood price changes. We also find that, consistent with the model, the impact of changes in seller composition on subsequent housing returns is significantly larger for houses with a higher neighborhood-β. In addition to determining the impact of seller composition on neighborhood level housing returns, we also test directly whether seller composition is correlated with observable changes in neighborhood characteristics. We show that the share of socioeconomically disadvantaged students in local schools moves with the composition of sellers in a neighborhood in a way that is consistent with the model predictions.

We also consider how house appreciation varies with the characteristics of buyers. We find that real estate agents and buyers who have previously owned a house nearby, are are likely to be better informed, purchase houses that outperform otherwise similar homes by an economically meaningful amount. One implication of this is that asymmetric information between buyers and sellers has important distributional implications when different buyers have different information sets. We also find that the impact of seller composition on price appreciation is smaller for houses bought by real estate agents or for houses bought by individuals who have previously lived in the same neighborhood. This is also consistent with the model's prediction that more informed buyers should be less sensitive to neighborhood characteristics than less informed ones.

1 Related Literature

This paper relates to a number of different literatures. 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 the commercial real estate market. They use regional variation in the quality of tax assessments to proxy for the importance of private information about property values 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 (2012) 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. Wong et al. (2011) argue that real estate markets in Hong Kong are more liquid when the arguably more readily observable component of land makes up a larger fraction of total property values, reducing the scope for market shut-downs due to asymmetric information. Relative to this literature, the current paper is the first to document that neighborhood characteristics provide a significant source of adverse selection in real estate markets. It also is the first to focus on the the impact on market outcomes when there are sellers and buyers that are better informed than their peers.[1]

On the theoretical side, the model is an extension of Kurlat (2012), who proposes a definition of competitive equilibrium for markets with asymmetric information where some buyers have different quality of information. This 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 to this literature, the model adds the feature that different buyers are allowed 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 and aggregate uncertainty about supply and/or payoffs, here the setting is closer to the

[1] This paper also contributes to the wide literature on testing for asymmetric information in a variety of markets such as insurance and annuities (Chiappori and Salanie, 2000; Finkelstein and Poterba, 2004; Bond, 1982).

original Akerlof (1970) model, with many heterogeneous assets but no uncertainty about the aggregate variables.

2 Model

2.1 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. 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. Let $v \equiv \beta\theta + (1-\beta)\,\eta$ be 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.

2.2 Information

Both θ and η are private information of each house's current owner. Buyers have heterogeneous information, and their types are indexed by b. A buyer of type b observes signal $x\,(v, b)$ from a house of value v, with

$$x\,(v, b) = \mathbb{I}\,(v \geq b) \tag{2}$$

The signal observed by buyer b tells him whether the total value of the house exceeds the threshold b, where b can be intrepreted as an index of expertise (higher values for b will enable buyers to find better houses on sale). Buyers with $b = 0$ are labeled "nonexperts" because they observe the same signal from every house; buyers with $b > 0$ are labeled "experts".[2] The expertise b can take on a discrete set of values in the set $[0, b^{\max}]$. We denote by $n(b)$ the measure of buyers of type b. We will focus on the case where $n(0)$ is large and $n(b)$ is relatively small for all $b > 0$, i.e. there is a relatively small number (in a sense to be made precise later) of potential expert buyers who are better informed than the rest.

2.3 Equilibrium

We study a competitive equilibrium of this economy. The challenge in defining a competitive equilibrium is that houses that look the same to nonexperts may look different to experts. Therefore the assumption made by Akerlof (1970) that indistinguishable assets are pooled at the same price cannot be made, at least not without further qualification, since whether or not two assets are indistinguishable depends on the identity of the buyer.

Kurlat (2012) proposes a notion of competitive equilibrium for this type of environment that we describe below. Instead of looking for a single market-clearing price for each type of house (or subset of types of houses), sellers are allowed to offer their houses on sale in a large set of potential "markets", which operate simultaneously. Each market is an abstract construct defined by a price and a rule for determining the order in which different buyers can trade. Sellers can offer any house on sale in any market, but depending on buyers' decisions on what to buy from which market, they may or may not be able to sell it. As in Gale (1996), the probability of being able to trade (rather than the price, which is a defining characteristic of the market) is what clears each market. Thus, transaction prices are determined by which markets/prices buyers and sellers choose to trade in and there is no presumption about whether or not different houses trade at the same price or not. We will, however, look for an equilibrium where all houses that look identical *to the least informed buyer* do trade at the same price and find conditions such that this is true.

Buyers demand a single house and can choose which market they will buy it from. In any given market, there may be houses of different qualities on sale and buyers are only partially able to distinguish between them using their signals. When placing an order for a house in a given market, they can specify a selection criterion that rejects some houses and

[2] $b = 0$ 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, which is already built into the prior distribution F. In other words, $b = 0$ is a renormalization of information to the information set of the least informed agent.

accepts others, but only to the extent that they are able to tell them apart; i.e. they can only distinguish between houses for which they observe different signals. Given the signal structure (2), a buyer with expertise b will typically reject any house where he observes a signal $x(v,b)=0$, i.e. any house for which $v<b$.

When buyers of different expertise buy houses from the same market, they will impose different selection criteria and therefore draw houses from different sub-pools. A key determinant of who ultimately gets what house is the order in which they are able to pick from the pool of houses on sale. In particular, if high-b buyers trade before low-b buyers, they will remove the higher-quality houses from the pool and low-b buyers will draw from a more-adversely-selected sample. In equilibrium, competition determines which ordering is used. Given the signal structure (2), it turns out that the equilibrium trading rule is that less-selective, lower-b buyers choose first. The reason is that lower-b buyers face more adverse selection if higher-b buyers trade before them, but not the other way around, so buyers self-select into markets where lower-b buyers trade first. As a result, any buyer buys a house that is randomly drawn from the entire pool of houses on offer that they are willing to accept.[3] Overall, an equilibrium is defined as a pattern of selling decisions by owners and buying decisions by buyers across the entire set of markets such that sellers and buyers optimize and the prices and trades that result are derived from the rules of the markets where agents choose to trade. See Kurlat (2012) for mathematical details. Under assumptions 1-3 below, the equilibrium will be such that all trades of houses that look indistinguishable to nonexperts take place at the same price p^*. Thus, even though the equilibrium construct requires many abstract markets operating simultaneously, the on-equilibrium trading behavior is actually quite simple: all houses that are observationally equivalent to nonexperts trade at the same price, but more expert buyers can pick better houses at that price.

2.4 Equilibrium Characterization

Let's focus on a set of houses that look indistinguishable to non-experts. Consider a market where the price is p and suppose that no buyers buy houses in any market where the price is higher than p, so that p is the best price an owner might be able to get. Owners put their houses on sale iff $p \geq u$ or, equivalently, iff $\varepsilon \leq \frac{p}{v}$. Hence the supply of houses of quality v that will be on sale will be

$$S(p,v)=G\left(\frac{p}{v}\right)f(v) \tag{3}$$

[3]This requires that the number of experts be sufficiently small. See Assumption 1 below.

and therefore the distribution of houses on sale at price p is given by the density

$$f(v|p) = \frac{S(p,v)}{\int S(p,\tilde{v})\,d\tilde{v}}$$

which has a mean of

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

Assume that the marginal buyer of houses is a nonexpert (which will be true under assumptions 1 and 2 below). This buyer must be indifferent to buying a random house, and hence the equilibrium price p^* must satisfy:[4]

$$p^* = \bar{v}(p^*) \tag{5}$$

Experts will not simply draw houses from the total supply but will instead reject any house with $v < b$ and therefore draw a house from a better distribution than nonexperts, given by the density

$$f(v|p,b) = \frac{S(p,v)\,\mathbb{I}(v \geq b)}{\int_{\tilde{v}\geq b} S(p,\tilde{v})\,d\tilde{v}} \tag{6}$$

which has a mean of

$$\bar{v}(p,b) = \frac{\int_{v\geq b} vS(p,v)\,dv}{\int_{v\geq b} S(p,v)\,dv} \tag{7}$$

Equation (7) implies that $\bar{v}(p,b)$ is increasing in b, i.e. more expert buyers will on average buy better houses. This implies that $\bar{v}(p^*,b) > p^*$ for all $b > 0$, so experts strictly prefer buying a house to not buying it, and therefore all experts buy houses. Adding up (6) over all experts, the fraction of houses of quality v on sale that are bought by experts is given by

$$\mu(v,p^*) = \sum_{b>0} \frac{n(b)\,\mathbb{I}(v \geq b)}{\int_{\tilde{v}\geq b} S(p^*,\tilde{v})\,d\tilde{v}} \tag{8}$$

[4]Notice that equation (5) already embeds the result that nonexperts pick houses before experts and therefore draw from the entire pool of houses offered on sale. The reason for this is that a buyer of type b is indifferent to picking houses after all buyers of types $b' < b$ (as long as these do not buy all the available houses, which will be true under assumption 1). This is because with signal structure (2) buyers with $b' < b$ do not disproportionately accept the better houses among those buyer b would accept. Conversely, buyer b strictly prefers picking a house before any trader with $b' > b$. As a result, buyers self-select into markets where the less selective buyers pick houses first. Equation (5) could have more than one solution. In that case, the equilibrium corresponds to the highest-price solution.

Notice that $\mu(v, p^*)$ is (weakly) increasing in v because higher-quality houses are acceptable to more buyers than lower-quality houses.

Assumption 1. $\mu(b^{\max}, p^*) < 1$.

Assumption 1 formalizes the condition that the number of expert buyers be sufficiently small. In particular, it states that the number of experts buyers is sufficiently small that there aren't enough of them to buy all the houses of quality $v = b^{\max}$. Given (8), this means that there there is no quality v such that all houses of that quality are bought by experts. Nonexperts, on the other hand, draw houses randomly from the entire pool of houses supplied. If they buy a total q_0 houses, this means that they buy a fraction

$$\mu_0(p^*) = \frac{q_0}{\int S(p^*, v)\, dv}$$

of houses the houses on sale of any given quality level. In equilibrium, they are indifferent to how many houses they buy and end up buying just enough houses so as that all houses with $v \geq b^{\max}$ are sold. If they bought any less, then some houses, including high-v houses, would remain unsold in market p^* and their owners would also attempt to sell them in market $p^* - \epsilon$; buyers would then prefer to buy from market $p^* - \epsilon$ instead of market p^* and we would not have an equilibrium at p^*. Conversely, if nonexperts bought any more houses, some expert buyers would not be able to buy a house because they would run out; they would therefore have an incentive to preempt this by buying from market $p^* + \epsilon$ and we would not have an equilibrium at p^*. Hence, in equilibrium it must be that $\mu_0(p^*) + \mu(b^{\max}, p^*) = 1$ or

$$q_0 = \int S(p^*, v)\, dv \cdot \left(1 - \sum_{b>0} \frac{n(b)}{\int_{v \geq b} S(p^*, v)\, dv}\right) \tag{9}$$

Assumption 2. $n(0) > q_0$.

Assumption 2 formalizes the condition that the number of nonexperts be sufficiently large. It must be large enough to buy q_0 houses as given by equation (9). This justifies the underlying assumption of equation (5) that nonexpert buyers get no surplus. For any quality level $v < b^{\max}$, the fraction of houses sold will be

$$\mu(v, p^*) + \mu_0(p^*) = 1 - [\mu(b^{\max}, p^*) - \mu(v, p^*)] < 1 \tag{10}$$

which means 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 don't do so and trade takes place in a single market with price p^*, one 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'.

The supply of houses of quality v that owners would be willing to sell at a price $p' < p$ is $S(p', v)$. But a fraction $1 - [\mu(b^{\max}, p^*) - \mu(v, p^*)]$ of those will have been sold at price p^*, so the residual supply of quality v houses offered at price p' is

$$S^R(p', p^*, v) = S(p', v) \cdot [\mu(b^{\max}, p^*) - \mu(v, p^*)] \tag{11}$$

Because $\mu(v, p^*)$ is increasing in v, equation (11) says that the residual supply supply is more adversely selected than the original supply. Higher-quality houses are more likely to have been picked by expert buyers in market p^* and therefore they are less likely to be on sale in market $p' < p^*$. The average quality obtained by buyer b in market p' is

$$\bar{v}^R(p', p^*, b) = \frac{\int_{v \geq b} v S^R(p', p^*, v)\, dv}{\int_{v \geq b} S^R(p', p^*, v)\, dv}$$

and therefore the surplus from buying in that market is $\bar{v}^R(p', p^*, b) - p'$.

Assumption 3. *For any b and any $p' < p^*$, $\bar{v}^R(p', p^*, b) - p' \leq \bar{v}(p^*, b) - p^*$.*

Assumption 3 ensures that no buyer is capable of obtaining a higher surplus by buying from the residual supply at a lower price that by buying from the original supply at price p^*. This assumption will hold if (i) experts buy a sufficient number of houses at p^* that the residual supply is significantly more adversely selected than the original supply and (ii) the left tail of the distribution of house qualities is sufficiently fat 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 (5).
2. Average qualities obtained by buyer b given by (7).
3. Selling probabilities for house quality v given by (10).

2.5 Example

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

Paremeter	Value		
β	0.1		
Distribution of θ	0	with probability	0.5
	1	with probability	0.5
Distribution of η	0	with probability	0.1
	1	with probability	0.9
Distribution of v	0	with probability	0.05
(resulting from	0.1	with probability	0.05
distributions of θ and η)	0.9	with probability	0.45
	1	with probability	0.45
Distribution of ε	$\varepsilon \sim U\,[0.5, 1]$		
Number of buyers per type	$n\,(0) = 1$		
	$n\,(0.1) = 0.2$		
	$n\,(0.9) = 0.2$		

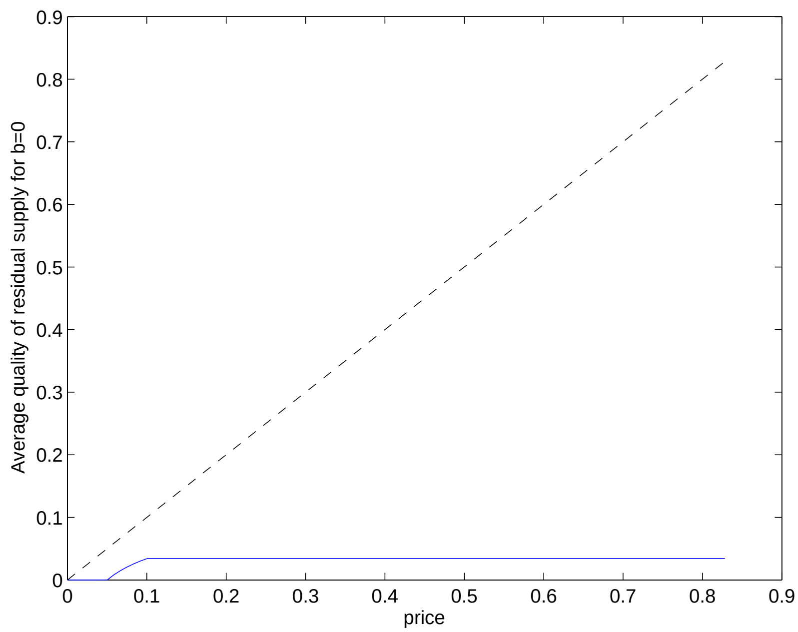
For this example, the equilibrium price given by (5) is $p^* = 0.83$, and quantities supplied and demanded are as follows:

House quality	$S\,(p^*, v)$	Quantity bought per buyer type			
		$b = 0$	$b = 0.1$	$b = 0.9$	Total
$v = 0$	0.05	0.02	0	0	0.02
$v = 0.1$	0.05	0.02	0.01	0	0.03
$v = 0.9$	0.38	0.16	0.10	0.11	0.38
$v = 1$	0.30	0.13	0.08	0.09	0.30
Total	0.78	0.33	0.2	0.2	0.73
Average v	0.83	0.83	0.89	0.94	

In the $p = 0.83$ market, all the $v = 0$ and $v = 0.1$ houses are put on sale, but only some of the higher-quality houses. Nonexperts buy 0.33 houses, drawn at random. The average quality they obtain is equal to the price, so they are indifferent to how many houses they buy. Since $n\,(0) = 1 > 0.33$, assumption 2 holds. Buyers of type $b = 0.1$ and $b = 0.9$ are more selective and obtain an average quality that exceeds the price. Thus, they all buy houses. They demand a combined 0.01 out of 0.05, 0.21 out of 0.38 and 0.17 out of 0.30 houses of qualities 0.1, 0.9 and 1 respectively, less than 100% in all cases. Thus, assumption 1 holds.

The 0.33 houses bought by nonexperts are exactly enough so that all houses acceptable to the most expert buyer (those with $v = 0.9$ and $v = 1$) get sold. Instead, some $v = 0.1$ and $v = 0$ 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 no buyer prefers to buy in those markets instead of in market p^*. For $b = 0.1$ buyers, they could obtain houses of $v = 0.1$ in any market where sellers are willing to offer them; since the lowest value of ε is 0.5, the lowest price at which anyone is willing to sell a $v = 0.1$ house is $p = 0.05$. Buying at that price, the buyer would obtain a surplus of 0.05, whereas he can obtain a surplus of $0.89 - 0.83 = 0.06$ in the $p = 0.83$ market. For nonexperts, they will obtain a mixture of $v = 0$ and $v = 0.1$ houses at any price below p^*. Figure 1 shows the average quality that nonexperts would obtain if they tried to buy from markets with $p < p^*$. 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 1: Would nonexperts want to buy from markets with $p < p^*$?



Note: This figure shows the average quality of houses that owners would be willing to sell at prices below 0.83 in the example. In all cases, the average quality is below the price so nonexperts would not be willing to buy.

3 Predictions of the model

3.1 Mapping the model to the data

The residential real estate market has many of the features described in the set up above, and below we will test the key implications for the equilibrium given the information structure. Our main object of interest in the empirical analysis will be the price appreciation experienced by different owners. We will relate that object to the model's predictions for $v - p$. The assumption is that the owners in our data have bought houses in a market that is well described by the above model; by the time they sell their house, the value v has been revealed and thus they obtain a price of v (plus or minus any subsequent shocks) upon resale. One of the main tests of the model is to see how $v - p$ 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.

3.2 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 $\frac{g_B(\mathcal{E})}{G_B(\varepsilon)} \geq \frac{g_A(\varepsilon)}{G_A(\varepsilon)}$ for every $\varepsilon \geq \frac{p^*}{\bar{v}}$. Then the proportion of sellers who belong to group A among sellers in neighbourhood 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 inframarginal. The condition $\frac{g_B(\mathcal{E})}{G_B(\varepsilon)} \geq \frac{g_A(\varepsilon)}{G_A(\varepsilon)}$ says that group B has more marginal sellers (as a fraction of inframarginal 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 empirical implication is that group composition of sellers in a neighborhood should predict subsequent appreciation. A high proportion of sellers from inelastic groups means that θ is high but this has not yet been captured in the price; if 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.

3.2.1 More and less informed sellers

The model above assumes that owners are perfectly informed about the neighborhood quality θ. Suppose instead we allowed for 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 distribution of x by $F_{x|\theta}(x|\theta)$ and 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. Without imposing more sturcture on the distributions of η, ε and x, it is not possible to ensure that the fraction of informed sellers decreases monotonically with θ, but this is true in many common cases; for instance, it's true under Normal distributions. Overall, the prediction that follows from Proposition 2 is that informed owners' selling decisions react more strongly to θ, simply because they know about its realization. Therefore we should expect the fraction of informed sellers in a neighborhood to be negatively associated with subsequent appreciation.

3.2.2 Neighborhood-β of sellers

In the model above, all houses were assumed to have the same weighting β on neighborhood level factors. Suppose instead 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.[5]

[5]In general, changing the weights on θ and η means that the distribution of v will be different for houses of different β_h. If β_h is publicly observable, equations (3)-(5) imply that houses of different β_h should be priced differently. We will instead maintain that all trade takes place at the same price. There are two ways to justify this. The first is to assume that β_h is not observable by potential buyers, so the distribution $F(v)$ already incorporates uncertainty about β_h. The second is to assume that house-specific and neighborhood-level shocks are drawn from the same distribution (i.e. $F_\theta = F_\eta$) and only consider local deviations of β_h

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 withdraw their houses from the market in response to high θ (or choose to put them on the market in response to low θ) to a greater extent than low-β_h owners. 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; without making more assumptions about distributions, it is not possible to ensure that $\frac{d^2 \Pr[\text{Sell}|\theta,\beta]}{d\theta d\beta} < 0$, i.e. that the effect is always monotonic. However, the implication of Proposition 3 is that, overall, higher-β_h owners to react more strongly to θ. Therefore we should expect a negative association between the β_h of sellers in a neighborhood and the subsequent appreciation of houses in that neighborhood.

3.2.3 Sellers of different tenure

One dimension of heterogeneity among home owners is how long ago they have bought their houses; 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

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.

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 very well matched with their house when they buy it 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.

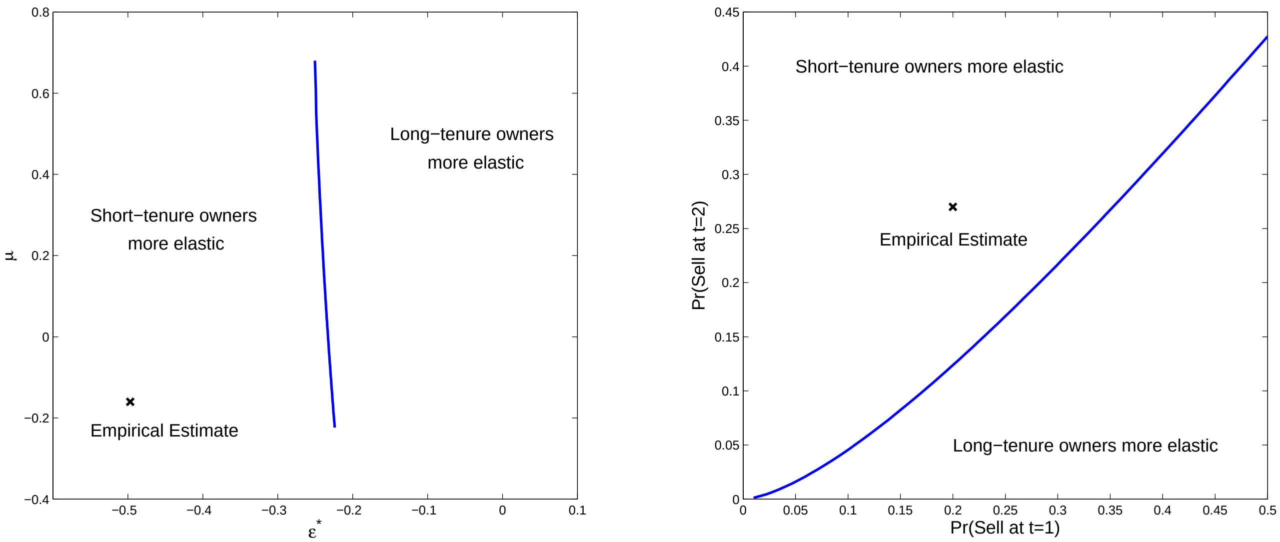
Example Suppose an owner's potential tenure consists of two periods. In the first period, he receives a shock $x_1 \sim N(\mu, 1)$ and his match-quality is $\varepsilon = m(x_1)$, where $m(\cdot)$ is any continuous monotonic function. Owners sell their house if $\varepsilon \leq \varepsilon^*$.[6] If 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 $\varepsilon \leq \varepsilon^*$.

Figure 2 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.

[6] Abstracting from heterogeneity in house quality.

Figure 2: 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.

Based on this example, we tentatively predict 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. Hence the proportion of long-tenure owners among sellers should be positively associated with subsequent appreciation. 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, more than on any single specification.

3.3 Differential effect by neighborhood-β

Suppose, as in section 3.2.2, 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.

3.4 Differential effects by buyer expertise

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 a buyer of type b is increasing in b.*

Proof. Immediate from equation (7). □

A less immediate implication of the model is that the expected appreciation of a house bought by an expert is less sensitive to the neighborhood quality than that of a house bought by a nonexpert.

Proposition 6. *Assume $\bar{\theta} > \frac{b^{\max}}{\beta}$. Then the expected house price appreciation obtained by a buyer of type b is increasing in b conditional on buying in relatively bad neighborhoods but does not depend on b in sufficiently good neighborhoods.*

Proof. See Appendix A □

Conditional on buying a house in a neighborhood that turns out to be bad, the expertise of the buyer will have a strong impact on the expected subsequent appreciation of the house. The reason is that expert buyers will only buy high-η houses in such a neighborhood while nonexperts might buy any house. In a good neighborhood, however, most houses will be acceptable to experts so, conditional on buying in such a neighborhood, they will be drawing from a sample that is not so different from the one that nonexperts draw from; therefore buyer expertise 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 expertise has no impact. Proposition 6 implies that we should see that the differential appreciation of houses bought by expert buyers should be negatively associated with neighborhood characteristics, or predictors thereof such as the composition of the pool of sellers. In other words, both being bought by an expert 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.

4 Data description

To conduct the empirical analysis, we combine a number of datasets. The first dataset contains the universe of ownership-changing housing deeds in Los Angeles county between June 1994 and the end of 2011. There are 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), the type of property (e.g. Apartment, Single-Family Residence) and the name of buyer and seller. It also reports the amount and duration of the mortgage and the identity of the mortgage lender. Figure 3 shows the location of each of the properties with transactions in our dataset. From this dataset, we extract all armslength transactions for which transaction prices reflect a true market value of the property. This prodecure, which excludes, amongst others, intra-family transfer deeds and foreclosure deeds, is described in more detail in Stroebel (2012). There are about 1.45 million armslength transactions.

A 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 up 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, the land share in total value might be a good proxy for neighborhood-β. As a first check whether the relative values assigned to land and property by the tax assessors appear realistic, Figure 4 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

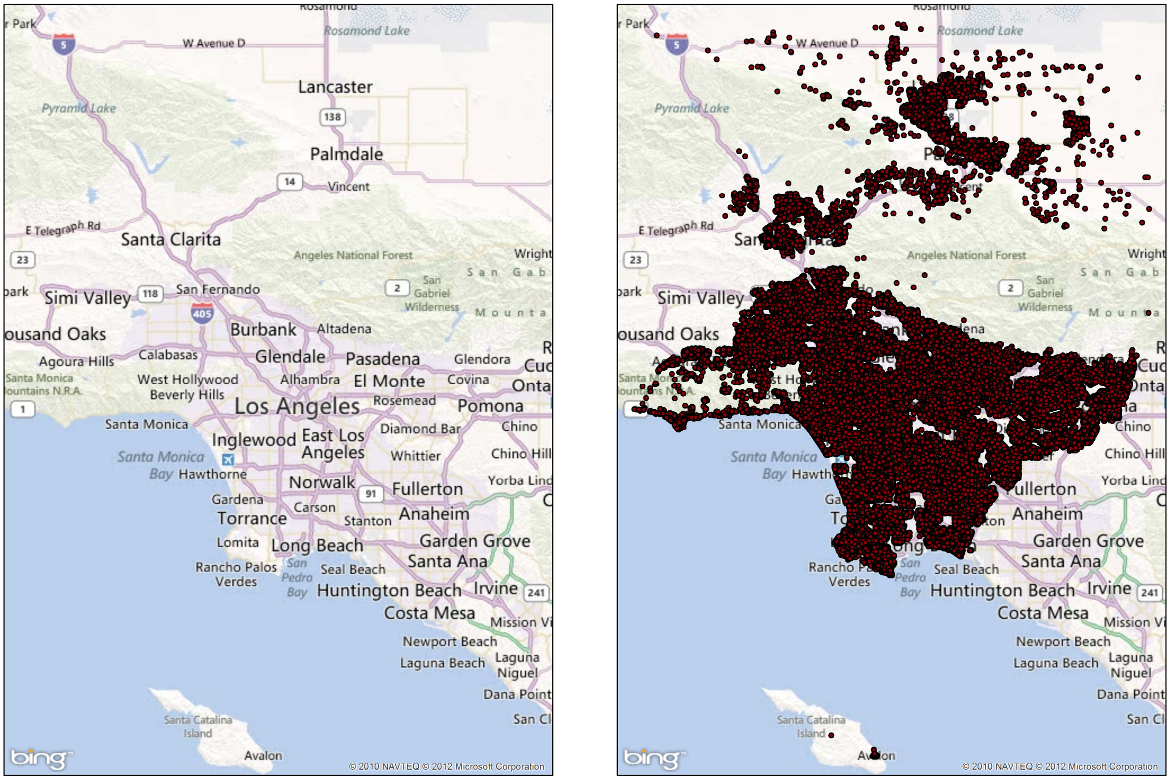


Figure 3: Transaction Sample

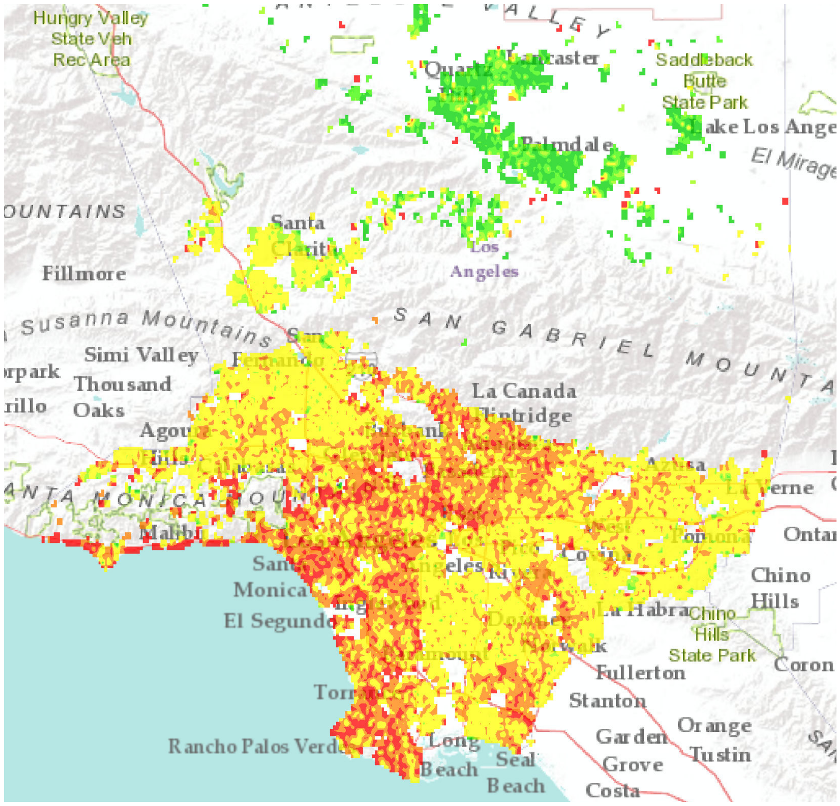
Note: This map shows the location of all houses for which we observe a transaction over between 1994 and 2012 for the Los Angeles area.

coast. Importantly for our purposes, there is also significant variation in this 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 informed about changes in neighborhood characteristics relative to other buyers and sellers. We merge this data to the deeds data using the name of the transactors to identify properties that have been bought or sold by a real estate professional. In particular, we classify a property as having been bought/sold be a real estate professional if at the time of sale there was an active real estate broker license issued in Los Angeles county to somebody of that name.[7]

[7] There will introduce some measurement error, since we will classify people with more common names to be real estate agents more regularly. However, we do not believe that this should introduce systematic bias into our analysis.

Figure 4: 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.

5 Results

5.1 Measuring Neighborhood - β

One of the key characteristics that differentiated houses in the model described in section 2 was the neighborhood-β of the individual homes. That is, houses differed in how much their value varies as neighborhood characteristics change. We argued that owners of houses with a higher neighborhood-β should respond particularly elastically in their decision to sell when they observe changes in neighborhood characteristics that are not yet factored into current prices. If this were indeed true, we would expect the share of high neighborhood-β homes amongst all transacted homes to have predictive power for average future neighborhood price changes. Similarly, we showed that if there is asymmetric information about neighborhood characteristics, the return of houses with a higher neighborhood-β should be particularly predictable using changes in the seller composition. To test these predictions of 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).[8]

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 armslength 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 market price movements in that zip code over the same period, $ZipReturn_{n,q_1,q_2}$. We do this by determining the annualized change in the median transaction price. In addition, 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 (12) for all repeat sales pairs with transactions between June 1994 and December 2011.

$$Return_{i,n,q_1,q_2} = \alpha + \beta_1 ZipReturn_{n,q_1,q_2} + \beta_2 ZipReturn_{n,q_1,q_2} \times LandShare_i + \epsilon_i \qquad (12)$$

The results are presented in Table 1. In column (1) we drop the interaction between *ZipReturn* and *LandShare*. The coefficient on *ZipReturn* shows that, reassuringly, on average house prices movements closely track movements of the zip code median. 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.

5.2 Changes in Seller Composition Predict Price Changes

In this section we test the predictions of Proposition 1, which suggested that changes in the composition of sellers should be predictive of future price changes of homes in that neighborhood. We regress the annualized capital gain of houses between two armslength transactions, $Return_{i,n,q_1,q_2}$, on control variables and the composition of sellers in neighbor-

[8] 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.

Table 1: Land Share as Neighborhood-β

	(1) Return	(2) Return	(3) Return
Zip Code Return	0.997***	0.956***	0.966***
	(0.004)	(0.013)	(0.015)
Zip Code Return × Land Share		0.068***	0.061***
		(0.017)	(0.021)
R-squared	0.793	0.793	0.807
N	390,972	390,970	285,663

Note: This table shows the results from regression (12). 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 over that period. Standard errors are clustered at the zip code level.

hood n and quarter q_1. We focus on three measures of seller composition, suggested by Propositions 2, 3 and section 3.2.3 respectively: (1) the fraction of sellers that are real estate professionals, and are thus particularly well informed about neighborhood characteristics, (2) the average land share of transacted houses and (3) the average tenure of sellers. Our measurement of 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 short-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.[9] Table 2 shows summary statistics on 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.

We then run regression (13) 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 market movements in house prices over time. The vector of control variables X_i includes information on the property (age, lot size, 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).

$$Return_{i,n,q_1,q_2} = \alpha + \beta_1 SellerComposition_{n,q_1} + X_i'\beta_2 + \xi_n + \phi_{q_1,q_2} + \epsilon_i \tag{13}$$

[9]Results are not sensitive to the choice of 3 years as the cut-off value.

Table 2: Summary Statistics Seller Composition

Variable	Neighborhood	Mean	Standard Deviation	
			Unconditional	Conditional
Share Informed Sellers	Zip Code	0.029	0.024	0.022
	Census Tract	0.029	0.041	0.040
Average Seller Land Share	Zip Code	0.59	0.114	0.045
	Census Tract	0.59	0.122	0.056
Seller Share Tenure > 3	Zip Code	0.78	0.079	0.068
	Census Tract	0.79	0.120	0.111

Note: This table shows summary stastistics for the seller composition by quarter and neighborhood for two different defintions 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 share of sellers with tenure exceeding 3 years is July 1997 to December 2011.

Table 3 shows the results from regression (13) when we consider a neighborhood to be a zip code. 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 10 basis points decline in the annualized return of houses. In column (2) we add a large set of control variables for characteristics of the house, the financing and the buyer. The estimated correlation between changes in the seller composition and subsequent returns remains unchanged. This suggests that the correlation is not driven by observable differences in the composition of houses or buyers that confounds our estimates of the impact of the composition of sellers. 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 associated with a 70 basis points decline in subsequent annualized returns 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 returns of houses in that neighborhood by about 45 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 columnn (7) we jointly include all three measures of neighborhood composition. The magnitude of the estimated contribution of each of the three measured compositions does not change much.

Table 3: Effect of Seller Composition in Zip Code on Returns

	(1) Return	(2) Return	(3) Return	(4) Return	(5) Return	(6) Return	(7) Return
Share Informed Sellers	-5.203***	-5.205***					-6.420***
	(1.108)	(1.091)					(1.125)
Average Seller Land Share			-17.05***	-17.92***			-14.07***
			(0.773)	(0.770)			(0.753)
Share in Zip of Tenure > 3					6.844***	6.899***	6.061***
					(0.383)	(0.377)	(0.371)
Fixed Effects	✓	✓	✓	✓	✓	✓	✓
House and Buyer Controls	·	✓	·	✓	·	✓	✓
R-squared	0.626	0.636	0.627	0.638	0.647	0.658	0.659
N	394,235	391,279	394,235	391,279	302,115	299,662	299,662

Note: This table shows results from regression (13). The dependent variable is the annualized return of the house between the two sequential armslength 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).

While average tenure in the zip code can capture sellers' information about neighborhood level characteristics, zip codes are a somewhat large geographic unit, and there might be additional relevant information about the immediate neighborhood of a particular property, that is not reflected in the composition of all sellers in a zip code. In Table 4 we report results from regression (13) with a neighborhood being defined as a four-digit census tract. While there are 293 unique zip codes in the 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 returns, while an increase in the average tenure of sellers predicts increases in neighborhood level returns. The magnitude of the estimated effect is smaller than the ones estimated at the zip code level, possibly due to more measurement error and 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, all the identification comes from differential variation of seller composition across census tracts within the same zip codes. Since this estimate partials out neighborhood characteristics that are common for different census tracts within the same zip code, the estimated coefficients are unsurprisingly smaller.[10]

Table 4: Effect of Seller Composition in Census Tract on Returns

	(1) Return	(2) Return	(3) Return	(4) Return	(5) Return	(6) Return
Share Informed Sellers	-1.809*** (0.393)			-0.700 (0.973)		
Average Seller Land Share		-8.312*** (0.363)			-3.952*** (0.660)	
Share in CT of Tenure > 3			2.974*** (0.160)			1.922*** (0.375)
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	0.636	0.637	0.658	0.683	0.683	0.702
N	391,279	391,279	299,662	391,279	391,279	299,662

Note: This table shows results from regression (13). The dependent variable is the annualized return of the house between the two sequential armslength 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. 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).

[10] In an additional robustness check we calculate our measures of seller composition in an area of 1/4 kilometer radius around each particular house. The results, which are very similar to the ones presented above, are available from the authors.

5.3 Future Neighborhood Characteristics

So far our evidence has shown that changes in seller characteristics can predict the future returns of houses in a way that is consistent with our model. We argued that this was because sellers were 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. In particular, we analyze 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, which is not available to buyers in real time, we construct for each zip code a student-population weighted measure of demographics of all schools in that zip code. We measure the share of students that are (i) classified as socioeconomically disadvantaged, (ii) qualify for free or reduced-price school meals under the National School Lunch Program or (iii) are classified as a non-native English Learner.[11] We then run regression (14), where we regress the demographic measures on the seller composition in that year. We also include fixed effects for the calendar year and the zip code.

$$Demographics_{n,y} \quad = \quad \alpha + \beta_1 \times SellerComposition_{n,y} + \xi_n + \phi_y + \epsilon_i \tag{14}$$

Table 5 shows results from this regression. The results show that in years with a higher share of informed sellers the demographics of the student population shifts more towards socioeconomically disadvantaged students, and students that qualify for free school meals. Similarly, and increase in the average land share of transacted homes and an increase in the share of sellers with only a short tenure of their home are associated with an increase in the the fraction of children that are economically disadvantaged.

5.4 Ownership Period of Second Buyer

The model in section 2 has only two periods and assumes that all the private information has been 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 armslength transactions of the house. We calculate the appreciation between the last two sales, as shown in Figure 5, and determine to what degree this is predicted by the composition of sellers contemporaneous

[11] The first two outcome variables are highly correlated. In fact, one is a strict subset of the other. "Socioeconomically Disadvantaged" 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.

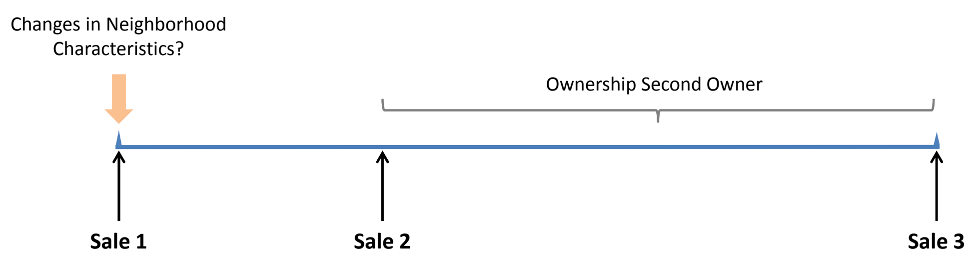
Table 5: Seller Characteristics and Demographics

	Socioeconomically Disadvantaged				Free Meal	English Learner
	(1)	(2)	(3)	(4)	(5)	(6)
Share Informed Sellers	258.1***			235.7***	17.32**	21.18***
	(73.53)			(73.45)	(8.281)	(7.643)
Average Seller Land Share		105.5***		76.00**	18.94***	12.17***
		(33.87)		(34.29)	(3.863)	(3.566)
Share in Zip of Tenure > 3			-92.54***	-81.82***	-11.91***	-0.0162
			(21.06)	(21.32)	(2.402)	(2.217)
Fixed Effects (Zip Code and Year)	✓	✓	✓	✓	✓	✓
R-squared	0.968	0.968	0.968	0.969	0.958	0.904
N	3,087	3,087	3,087	3,087	3,088	3,088

Note: This table shows results from regression (14) for the years 2000 - 2011. Each specification includes zip code and year fixed effects. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

to the first sale. In other words, we run regression (15), where q_1, q_2 and q_3 represent the calendar quarters of the first, second and third sale.

Figure 5: Measures of Second Buyer's return



$$Return_{i,n,q_2,q_3} = \alpha + \beta_1 SellerComposition_{n,q_1} + \chi_{q_1} + \phi_{q_2,q_3} + \xi_n + X_i'\beta_2 + \epsilon_i \tag{15}$$

Table 6 shows the results separately for samples where we allow up to 4 and up to 6 years between sale one and sale two. For longer time horizons between the first and 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 return. We can see that after that following an increase in the share of informed sellers houses in that neighborhood continue to underperform, even during the ownership period of people who buy subsequently. 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

Table 6: Ownership Period of Second Buyer

	(1) Return	(2) Return	(3) Return	(4) Return	(5) Return	(6) Return
Share Informed Sellers	-9.528***	-4.629*				
	(3.370)	(2.593)				
Land Share			-14.86***	-10.52***		
			(2.364)	(1.676)		
Share in Zip of Tenure > 3					3.139**	1.743*
					(1.237)	(0.988)
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	0.605	0.595	0.608	0.597	0.637	0.628
N	76,998	90,395	76,998	90,395	59,525	66,352

Note: This table shows results from regression (15). The dependent variable is the annualized return 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: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

for changes in our other two measures of seller composition, the average land share of sold homes and the average tenure of sellers.

5.5 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 misstate the effect of seller composition on the return of homes. To address such concerns, in Table 7 we 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 either observe the death the original owners or a divorce of the original owners in the 12 months preceeding the resale.[12] arguing that such sales are more plausibly prompted by

[12] These events can be identified in the deeds data. We classify divorces through the presence of an "Intra-Family Transfer & Dissolution" deed that transfers property rights from initially joint ownership to one of the initial owners. 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 I observe one of the following: "Affidavit of Death of Joint Tenant" or "Executor's Deed."

the death or divorce rather than by any factors correlated with seller composition at the purchase date or with price changes. In Table 7 we show results from regression (13) similar to Table 3 where we measure average seller characteristics at the zip code level, but limit our sample only to those forced moves. The results show that the correlation between seller characteristics and subsequent housing returns is of the same magnitude in the sample of forced moves as it is in the entire sample. This suggests that selection into observing repeat sales does not bias our estimates.

Table 7: Effect of Seller Composition on Returns - Forced Moves

	Return	Return	Return
Share Informed Sellers	-7.949**		
	(3.542)		
Average Seller Land Share		-19.50***	
		(2.617)	
Share in Zip of Tenure > 3			6.063***
			(1.452)
Fixed Effects	✓	✓	✓
House and Buyer Controls	✓	✓	✓
R-squared	0.600	0.602	0.621
N	17,575	17,575	13,422

Note: This table shows results from regression (13) for those transactions where the resale was preceeded in the six months before by a death of the owner. The dependent variable is the annualized return 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).

5.6 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 rather, 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 large fraction of short-tenure owner and would predict low subsequent appreciation.

To rule out that this is the main driver, we attempt to identify in our data a set of individuals whom 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 8 shows that the results are very robust to only considering the average tenure of sellers that are not classified as flippers.

Table 8: Remove Possible Flippers from Tenure

	(1) Return	(2) Return	(3) Return	(4) Return
Share in Zip of Tenure > 3	6.899*** (0.377)	5.405*** (0.348)	5.400*** (0.348)	6.617*** (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
N	299,662	299,655	299,655	299,633

Note: This table shows results from regression (13). The dependent variable is the annualized return 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 $\times$ zip code level. Significance Levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

5.7 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 the predictions of Proposition 4. 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 (16), where $LandShare_i$ is the house-specific share of total value made up by land, as reported in the assesor values. The coefficient of interest is β_3, which measures the increase in the responsiveness of return to seller composition when the house has a larger land share.

$$\begin{aligned} Return_{i,n,q_1,q_2} \quad = \quad & \alpha + \beta_1 \times SellerComposition_{n,q_1} + \beta_2 \times LandShare_i + \\ & \beta_3 \times SellerComposition_{n,q_1} \times LandShare_i + X_i'\beta_2 + \xi_n + \phi_{q_1,q_2} + \epsilon_i \end{aligned} \tag{16}$$

The results of this regression are presented in Table 9, 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 columns (1) and (2) 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 return to a one conditional standard deviation increase in the share of informed sellers by about 6 basis points when considering the zip code as the neighborhood, and by about 12 basis points when considering the census tract. A similar move in the land share distribution will increase the response of returns to an increase in the average land share of sellers by 19 basis points for zip codes as neighborhoods, and about 24 basis points for census tracts as neighborhoods. Finally, moving from the 25th to the 75th percentile in land share distribution will increase the response of return to a one conditional standard deviation change in the share of sellers who have lived in their house for more than 3 years by 7 basis points (10.2 basis points) for zip codes (census tracts) as neighborhoods.

5.8 Relative Informedness of Buyers

The model predicts that more informed buyers should obtain higher average appreciation (Proposition 5) and that this advantage should be especially strong conditional on buying houses from bad neighborhoods (Proposition 6). 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 (17) with a dummy variable for whether or not the buyer was a real estate agent.

$$\begin{aligned} Return_{i,n,q_1,q_2} \quad = \quad & \alpha + \beta_1 \times SellerComposition_{n,q_1} + \beta_2 \times InformedBuyer_i + \\ & \beta_3 \times SellerComposition_{n,q_1} \times InformedBuyer_i + X_i'\beta_2 + \xi_n + \phi_{q_1,q_2} + \epsilon_i \end{aligned} \tag{17}$$

Table 9: Effect of Seller Composition by Land Share

	(1) Return	(2) Return	(3) Return	(4) Return	(5) Return	(6) Return
Land Share	-0.777***	-0.649***	7.852***	7.551***	-4.186***	-3.792***
	(0.143)	(0.107)	(0.487)	(0.424)	(1.175)	(0.727)
Share Informed Sellers	0.951	4.170***				
	(3.177)	(1.583)				
Land Share × Share Informed Sellers	-9.662**	-9.196***				
	(4.444)	(2.293)				
Average Seller Land Share			-7.973***	1.354**		
			(0.925)	(0.623)		
Land Share × Average Seller Land Share			-15.26***	-14.48***		
			(0.863)	(0.767)		
Share in NH of Tenure > 3					4.650***	0.396
					(1.030)	(0.612)
Land Share × Share in NH of Tenure > 3					3.659**	3.160***
					(1.483)	(0.900)
Neighborhood	Zip Code	Census Tr.	Zip Code	Census Tr.	Zip Code	Census Tr.
Fixed Effects, House and Buyer Controls	✓	✓	✓	✓	✓	✓
R-squared	0.637	0.638	0.638	0.640	0.659	0.660
N	391,279	391,248	391,279	391,248	299,662	299,642

Note: This table shows results from regression (16). The dependent variable is the annualized return 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 (5), and at the quarter × 4-digit census tract level in the other columns. Columns (1), (3) and (5) include sales quarter pair fixed effects and zip code fixed effects and have standard errors clustered at the quarter × zip code level, while columns (2), (4) and (6) include sales quarter pair × census tract fixed effect and have standard errors clustered at the quarter × census tract 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 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. Significance Levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

Table 10 shows the results from this regression. In column (1), which tests Proposition 5, 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 80 basis points annually relative to otherwise observationally similar houses purchased by agents that are not real estate agents. The model-based explanation for this is that these buyers are better able to pick out good deals from the set of homes on offer. In columns (2) - (4) we show that, consistent with Proposition 6, the difference in the return of houses purchased by real estate agents and other individuals is particularly big in neighborhoods that are predicted to underperform. As discussed above, the intuition for this derives from the fact that in good neighborhoods, informed buyers find most houses to be a good deal, and thus behave similar to uninformed buyers, who cannot tell good and bad houses apart. In bad neighborhoods, however, informed buyers use their expertise 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 10: Effect of Buyer Characteristics - Real Estate Professionals

	(1) Return	(2) Return	(3) Return	(4) Return
Real Estate Prof.	0.801***	0.645***	0.0904	2.743***
	(0.050)	(0.075)	(0.299)	(0.703)
Share Informed Sellers		-2.683***		
		(0.877)		
Real Estate Prof. × Share Informed Sellers		5.477**		
		(2.219)		
Average Land Share			-10.49***	
			(0.616)	
Real Estate Prof. × Average Land Share			1.200**	
			(0.510)	
Share in Zip of Tenure > 3				4.973***
				(0.321)
Real Estate Prof. × Share in Zip of Tenure > 3				-2.387***
				(0.886)
R-squared	0.704	0.704	0.704	0.722
N	391,278	391,278	391,278	299,662

Note: This table shows results from regression (17). The dependent variable is the annualized return 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 relative informedness of buyers, which also use the fact that we can observe the names of all buyers and sellers in the deeds records. For those houses bought by people that we observe selling a house anywhere in Los Angeles 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 (17).[13] We conjecture that the further these buyers have 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 Tabel 11.

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

	(1) Return	(2) Return	(3) Return	(4) Return
Log(Distance)	-0.241***	-0.261***	0.011	-0.476**
	(0.016)	(0.025)	(0.080)	(0.218)
Share Informed Sellers		-5.651***		
		(2.044)		
Log(Distance) × Share Informed Sellers		0.667		
		(0.711)		
Average Land Share			-11.28***	
			(1.087)	
Log(Distance) × Average Land Share			-0.427***	
			(0.136)	
Share in Zip of Tenure > 3				3.925***
				(0.834)
Log(Distance) × Share in Zip of Tenure > 3				0.305
				(0.281)
R-squared	0.697	0.697	0.697	0.735
N	99,317	99,317	99,317	68,847

Note: This table shows results from regression (17). The dependent variable is the annualized return 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 buyers who previously lived further away buy houses that under-perform otherwise similar houses bought by people that lived closer by. This is consistent with agents that lived closer by having superior information about characteristics of the neighborhood that allow them to pick better deals. Column (3) shows that the return differ-

[13] 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.

ence between those houses bought by neighborhood insiders and outsiders is particularly big in bad neighborhoods, i.e. those where the average land share sold is particularly high. The interactions with the other two measures of seller composition is not statistically significant.

As our final measure of "infomed buyer," we check whether we observe someone with the particular 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. The results are presented in Table 12, and look very similar to those presented in Table 11.

Table 12: Effect of Buyer Characteristics - Same Zip Code

	(1) Return	(2) Return	(3) Return	(4) Return
Same zip	1.010***	1.042***	0.371	1.645**
	(0.056)	(0.091)	(0.282)	(0.722)
Share Informed Sellers		-2.386***		
		(0.876)		
Same zip $\times$ Share Informed Sellers		-1.086		
		(2.620)		
Average Land Share			-10.49***	
			(0.617)	
Same zip $\times$ Average Land Share			1.072**	
			(0.478)	
Share in Zip of Tenure > 3				4.932***
				(0.321)
Same zip $\times$ Share in Zip of Tenure > 3				-0.804
				(0.919)
R-squared	0.704	0.704	0.705	0.722
N	391,278	391,278	391,278	299,662

Note: This table shows results from regression (17). The dependent variable is the annualized return 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 $\times$ zip code level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

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. 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 the predictions of Proposition 5. In addition, and consistent with Proposition 6, this advantage is lower when the predicted neighborhood quality is higher, in particular based on the average land share seller composition measure.

6 Conclusion

In many markets, sellers of an asset are better informed about a payoff-relevant aspect of that 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* the neighborhood component and the structure component of the value of a house, but both among buyers and among sellers some are better informed than others. We find that the correlations between buyer, seller and house characteristics and house prices are consistent with the patters than one would predict under these conditions. In particular, this indicates that asymmetric information about neighborhood characteristics is an important aspect of real estate markets, and that heterogeneity in information across different agents can have significant distributional consequences.

References

Akerlof, G.A., “The market for” lemons”: Quality uncertainty and the market mechanism,” *The 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.*

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

Bayer, P., C. Geissler, and J.W. Roberts, “Speculators and middlemen: The role of flippers in the housing market,” Technical Report, National Bureau of Economic Research 2011.

Bond, E.W., “A direct test of the” lemons” model: the market for used pickup trucks,” *The American Economic Review*, 1982, pp. 836–840.

Chiappori, P.A. and B. Salanie, “Testing for asymmetric information in insurance markets,” *Journal of Political Economy*, 2000, *108* (1), 56–78.

Davis, M.A. and J. 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,” *The Quarterly Journal of Economics*, 2002, *117* (4), 1529–1570.

Finkelstein, A. and J.M. Poterba, “Adverse selection and the choice of risk factors in insurance pricing: Evidence from the uk annuity market,” *Journal of Political Economy*, 2004, *112* (1), 183–208.

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, M.J. and T.J. Moskowitz, “Confronting information asymmetries: Evidence from real estate markets,” *Review of Financial Studies*, 2004, *17* (2), 405–437.

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

Guerrieri, V., D. Hartley, and E. Hurst, "Endogenous gentrification and housing price dynamics," *NBER Working Paper*, 2010, *16237.*

Guerrieri, Veronica, 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.

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.

Stroebel, J., "The impact of asymmetric information about collateral values in mortgage lending," *Working Paper*, 2012.

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

Wong, Siu, C. Yiu, and K. Chau, "Liquidity and information asymmetry in the real estate market," *Journal of Real Estate Finance and Economics*, 2011, pp. 1–14.

A Theoretical Appendix

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 neighbourhood* j *is increasing in* θ_j*.*

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

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

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

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

Taking the derivative with respect to v:

$$\begin{aligned}
\frac{d\pi_{A|Sell}\left(v\right)}{dv} &= \frac{-\pi_A g_A\left(\frac{p^*}{v}\right)\frac{p^*}{v^2}\left[\pi_A G_A\left(\frac{p^*}{v}\right) + \pi_B G_B\left(\frac{p^*}{v}\right)\right] + \left[\pi_A g_A\left(\frac{p^*}{v}\right)\frac{p^*}{v^2} + \pi_B g_B\left(\frac{p^*}{v}\right)\frac{p^*}{v^2}\right]\pi_A G_A\left(\frac{p^*}{v}\right)}{\left[\pi_A G_A\left(\frac{p^*}{v}\right) + \pi_B G_B\left(\frac{p^*}{v}\right)\right]^2} \\
&= \frac{-\pi_A g_A\left(\frac{p^*}{v}\right)\frac{p^*}{v^2}\left[\pi_B G_B\left(\frac{p^*}{v}\right)\right] + \left[\pi_B g_B\left(\frac{p^*}{v}\right)\frac{p^*}{v^2}\right]\pi_A G_A\left(\frac{p^*}{v}\right)}{\left[\pi_A G_A\left(\frac{p^*}{v}\right) + \pi_B G_B\left(\frac{p^*}{v}\right)\right]^2} \\
&= \frac{\pi_A \pi_B \frac{p^*}{v^2} G_A\left(\frac{p^*}{v}\right) G_B\left(\frac{p^*}{v}\right)}{\left[\pi_A G_A\left(\frac{p^*}{v}\right) + \pi_B G_B\left(\frac{p^*}{v}\right)\right]^2}\left[\frac{g_B\left(\frac{p^*}{v}\right)}{G_B\left(\frac{p^*}{v}\right)} - \frac{g_A\left(\frac{p^*}{v}\right)}{G_A\left(\frac{p^*}{v}\right)}\right] > 0 \qquad (18)
\end{aligned}$$

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

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

Taking the derivative with respect to θ and using (18):

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

□

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

$$\Pr\left[\text{Sell}|\theta, \text{Informed}\right] = \int G\left(\frac{p^*}{\beta\theta + (1-\beta)\,\eta}\right) dF_\eta\left(\eta\right) \tag{19}$$

Similarly, for uninformed sellers,

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

so integrating across realizations of x:

$$\Pr\left[\text{Sell}|\theta, \text{Uninformed}\right] = \int\int G\left(\frac{p^*}{\beta\hat{\theta}\left(x\right) + (1-\beta)\,\eta}\right) dF_\eta\left(\eta\right) dF_{x|\theta}\left(x\right) \tag{20}$$

For any nondegenerate distribution $F_{x|\theta}$, $0 < \hat{\theta}\left(x\right) < \bar{\theta}$ for all x. Equations (19) and (20) 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:

$$\Pr\left[\text{Sell}|\theta, \beta_h\right] = \int G\left(\frac{p^*}{\beta_h\theta + (1-\beta_h)\eta}\right) dF_\eta(\eta) \tag{21}$$

□

1. *Taking the derivative of (21) 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

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

2. *Taking the derivative of (21) with respect to θ:*

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

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 4. *Assume $\bar{\theta} > \frac{b^{\max}}{\beta}$. Then the expected house price appreciation obtained by a buyer of type b is increasing in b conditional on buying in relatively bad neighborhoods but does not depend on b in sufficiently good neighborhoods.*

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

$$\begin{aligned}\bar{v}(b,\theta) &= \frac{\int_{v\geq b} vG\left(\frac{p^*}{v}\right) f_{v|\theta}(v)\,dv}{\int_{v\geq b} G\left(\frac{p^*}{v}\right) f_{v|\theta}(v)\,dv} \\ &= \frac{\int_{v\geq b} vG\left(\frac{p^*}{v}\right) f_\eta\left(\frac{v-\beta\theta}{1-\beta}\right) dv}{\int_{v\geq b} G\left(\frac{p^*}{v}\right) f_\eta\left(\frac{v-\beta\theta}{1-\beta}\right) dv}\end{aligned}$$

so taking the derivative with respect to b:

$$\frac{\partial \bar{v}(b,\theta)}{\partial b} = \frac{G\left(\frac{p^*}{b}\right) f_\eta\left(\frac{b-\beta\theta}{1-\beta}\right) \int_{v\geq b} (v-b)\,G\left(\frac{p^*}{v}\right) f_\eta\left(\frac{v-\beta\theta}{1-\beta}\right) dv}{\left[\int_{v\geq b} G\left(\frac{p^*}{v}\right) f_\eta\left(\frac{v-\beta\theta}{1-\beta}\right) dv\right]^2} \tag{22}$$

If $\theta > \frac{b}{\beta}$, then $\frac{b-\beta\theta}{1-\beta} < 0$ and therefore $f_\eta\left(\frac{b-\beta\theta}{1-\beta}\right) = 0$ and $\frac{\partial\bar{v}(b,\theta)}{\partial b} = 0$, so for $\theta > \frac{b^{\max}}{\beta}$, then $\frac{\partial\bar{v}(b,\theta)}{\partial b} = 0$ for all b. Otherwise, all the terms in (22) are positive, so $\frac{\partial\bar{v}(b,\theta)}{\partial b} > 0$.[14] □

[14]For $\theta < \frac{b-\bar{\eta}(1-\beta)}{\beta}$, then $\frac{\partial\bar{v}(b,\theta)}{\partial b}$ is not defined because b-type buyers would never buy in such a neighborhood.