Measuring the Thinness of Real Estate Markets

(Preliminary Version)

David Genesove
Hebrew University of Jerusalem
and CEPR

Lu Han
University of Toronto

March 1, 2011

1. Introduction

A number of studies have shown the inadequacy of the perfect capital market paradigm as a model of housing markets.[1] Other research has shown that measures of liquidity such as time to sale help explain construction activity and pricing dynamics.[2] Together, these studies point to the need to incorporate search and matching into any compelling analysis of the housing market. Even a casual acquaintance with the market, and certainly most people's experience of buying or selling a home, would attest to the essential thinness of the market – the variability in the match between buyer and seller, and the inability to assess that without a costly visit to the home, which underlie all search and matching models. Micro studies of time on the market show it to be consistent with simple search theoretic intuition. But how thin is the housing market? Is the extent such as to be able to explain the anomalies that we see at the macro level?

This project answers this question through estimation of parameters that capture the market thinness, based on a survey of recent buyers in a large North American urban area. The novel feature of this survey is that respondents report on their buying experience, including time spent searching and number of buyers competed against in purchasing the home, as well as information about the seller, specifically list price and time to sale. Buyer demographic information is collected as well. We complement the survey information with publicly available information on list price, seller time on the market (both highly correlated with the buyer's report) and home attributes.

Exploiting this unique dataset, we propose two measures of market thinness: a reduced-form measure that reflects how much price increases with the number of bidders, and a structural measure that reflects how disperse buyers valuation of a given house is under certain distribution assumptions.

The reduced-form measure, while less direct, is intuitively appealing. If homes are not very different one from another than one buyer will evaluate it pretty much the same as another,[3] and a third bidder will not result in an appreciably higher price than two will; and with little to gain from further search, buyers and sellers' search valuations will differ very little, so that the

[1] The leading examples include Case and Shiller (1989) and Krainer (2001).
[2] E.g.,Mayer and Somerville (2000) and Berkovec and Goldman (1996).
[3] Buyers of different income levels or quality sensitivity will search among different market segments.

bargaining outcome when only one buyer shows up will differ little from the price when two buyers compete.

Straightforward ordinary least squares estimation is feasible: regressing the final price on the number of bidders will yield a consistent estimate of this effect. Unfortunately, unobserved quality varies so much as to render it extremely imprecise. An alternative often used is to normalize final price by list price, on the argument that the latter will incorporate unobserved quality. That unfortunately introduces an upward bias, as, conditional on housing attributes observable to buyers, a lower list price will have induced more buyers to consider the house.

To address these issues, we propose a simultaneous estimation of list price, number of bidders, and final price, by the exact maximum likelihood method. We will show how the OLS bias of the effect of bidders on the final price can be corrected, via the estimated effect of the list price on the number of bidders. That coefficient, too, is inconsistently estimated by OLS, again because the list price includes unobserved (by the econometrician) quality. The consequent classical errors-in-variable problem is correctable by a reliable estimate of the variance of the unobserved quality. For that we use the covariance of the list price with previous list price, estimated by a method analogous to that of repeat sales indices. Since the number of bidders is stochastic, an independent effect of the list price on the final final price, say as when buyers bid higher in response to a higher list price, can also be identified.[4]

Our simple OLS estimates show that on average, doubling the number of bidders increases the final price by 2.4%. This effect is statistically significant and economically substantial, providing evidence for the thinness of real estate markets. When we use the maximum likelihood method to estimate the simultaneous estimation model, we find that, the effect of the number of bidders on the final price remains close to the OLS estimates. This suggests that our reduced-form estimate of market thinness is robust to the concerns on unobserved house quality and endogenous bidder response.

The simultaneous estimation model also allows us to empirically investigate how the list strategy affects the final house price. Note that how buyers respond to the list price is an important issue in its own right. This is a central issue in the literature on directional search and a crucial factor in determining market efficiency (Merlo and Ortalo-Magne, 2004; Albrecth, et al,

[4] Albrecht, Gautier and Vroman (2010) show how patient sellers can signal their type through high list prices. The effect is not identified under OLS.

2010). If sellers are able to draw additional buyers to them by reducing the list price, then price can play its rationing role (Moen, 1997). But if list prices only indicate the market segment, essentially indicating the unobservable quality or sellers' type, or buyers have difficulty separating the two, that function will be severely impeded. In addition to affecting the final price thorough its effect on buyer behavior, the list price is also correlated with the final price through a selection effect, that is, sellers with higher reservation prices may both set higher list prices and only accept higher bids. Our simultaneous estimation model allows us to distinguish between these different roles of the list price. We find that a reduction in the listing price, held observed housing attributes and unobserved quality constant, indeed increases the final price by intensifying bidding competition. However, this effect is statistically insignificant and quantitatively dominated by the loss of price premium through the signaling and selection effects of the list price.

Having found the reduced-form evidence in support of the market thinness, we then turn to a structural measure of the market thinness by estimating how disperse buyer valuation is. The structural approach is appealing in that its micro foundation allows us to incorporate features of housing markets, such as search and bidding, directly into standard hedonic house price estimation. To do so, we set out a straightforward private independent value auction model and derive the final price as a function of housing attributes and the expected second order statistics from the bidders' valuation. Estimating this relationship, we find that the spread of bidders' valuation, measured by the standard deviation of the distribution of offers, is 4.2 percentage points under the uniform distribution, and 4.9 percentage points under the standard extreme value distribution.

To understand the implications of these estimates, we further compute money left on the table in a real estate bidding war. As expected, the winning bidder (buyer)'s surplus decreases with the number of bidders that he is competing with. Take the standard extreme value distribution as an example. The buyer's surplus, measured by the difference between his valuation of the house and the price he pays, is 5.3% of the final price when he is completing with one bidder; and is reduced to 4.6% when a third bidder participates in the bidding war. At the sample mean, these estimates are equivalent to $22,472 and $19,504 of surplus, respectively. Thus increasing the number of bidders from two to three reduces the winning bidder's surplus by 13.21%.

2. Model: Measuring Market Thinness

We are interested in measuring the thinness of real estate markets. We begin by describing a reduced-from the model. We then describe a simple auction framework that guides the subsequent structural estimation exercise and aids in interpreting the results.

2.1 Reduced-Form Measure of Market Thinness

In this section, we lay out an econometric strategy to precisely estimate the effect of the number of bidders on the final price, in the presence of substantial unobserved quality, an imperfect proxy for it in the form of the list price, and a possible endogeneity bias resulting from the responsiveness of the number of bidders to the list price.

It will prove useful in the interpretation of the statistical model to have a sketch of our understanding of how prices are determined in the market. Potential sellers are characterized by a reservation price, which is a threshold level above which they are willing to part with the housing unit, below which they are not. In a fully specified model this would equal the value of search, which would incorporate their beliefs about the arrival rates of buyers, the bids buyers would make and so on. Sellers set a list price that is advertised to potential buyers. Sellers are not committed to this price, but it may convey information to the buyers about the seller's reservation price. Buyers show up, possibly in response to the advertised list price. They bid, and the resulting price is thus determined.

This model avoids any concept of time, and so is best suited to a market in which an auction is held at some set period of time after the property has listed. The market under consideration in large part fits this description.

With this sketch of a theoretically model in mind, we present a three equation model of the list price, the number of bidders and the final price. Each is modeled as a log linear function of latent variables that represent unobserved quality, the list price strategy, the realized deviation of the number of bidders around the expected mean and the realized deviation of the price around its mean. Exclusion restrictions ensure that the parameters are identified.

The model is as follows:

(1) $$l = v + \eta$$

(2) $$b = \psi\eta + w$$

(3) $$p = \phi b + m\eta + v + u$$

where l is the list price, b is the number of bidders and p is the final price. All three variables are in logs, and all should be understood as the residuals after a regression on observed attributes (and controlling for time and location). ψ, ϕ and m are parameters. The remainder is latent variables.

Equation (1) states that the list price is composed of a quality component (v) that is unobserved by the econometrician, and a list price strategy (η). The latter will be determined by seller's relative patience, beliefs about the best strategy to take, and so on.

Equation (2) states that the expected number of buyers who bid on the home is proportional to the list price strategy. The factor of proportionality is ψ; presumably, it is negative, as intuition would suggest, and as is predicted in equilibrium and off the equilibrium path in Peters (2001) and elsewhere. In practice, there will be deviations from the expected number of bidders. Those deviations are indicated by w. They arise out of the uncoordinated nature of buyer visits to sellers homes which implies that the number of visitors will be random, the heterogeneity in buyers' valuation of a home, coupled with fixed costs of entering into negotiation or bidding, as well as other inducements of buyers to visit, such as advertising.

Equation (3) states that the final price is comprised of a number of effects. First is the number of bidders. If buyers have private valuations, then the more bidders there are, the higher the price, so that the coefficient ϕ is expected to be positive. Of course, common value auctions could switch the sign, but our sense is that the common value aspect is rather limited in this market.

Second is the list price strategy, with the accompanying coefficient m. The parameter m reflects both the effect of the list price on bids and a selection effect. A higher list price can signal that the seller is more patient (a la Albrecht et al), and so cause a single buyer to offer more in negotiation with the seller, when there are no other bidders present, or a winning bidder to up his bid after the auction. In a sealed bid auction setting, bidders who think the seller will either accept or reject the winning bid, and not negotiate with the winning bidder, will also be induced to bid higher. Thus the list price strategy can operates both indirectly through its effect on the number of bidders and directly though its effect on buyers' offers.

Unfortunately, there is also a selection effect: even if buyers' bids are independent of the list price, if the list price is positively correlated with the seller's reservation price, as one would expect it to be, then there will be a positive correlation between the list price and the final price solely because only sufficiently high winning bids will be accepted by the seller. That is, sellers with high reservation prices will both set high listing prices and obtain high prices, the latter because they will only accept high offers. We hope to be able to differentiate between the signaling effect and the selection effect by incorporating seller time on the market into our statistical analysis, but we have yet to do so.

The bids will also be increasing in the unobserved quality v. This explains the third element in Equation (3).

Finally, there is u. This is most easily understood as the realization of the winning bid around its expected value. However, u will also pick up any additional selection effects not captured by the list price strategy, should the seller's list price does not fully reflect his reservation price. For example, one might imagine that some sellers allow the agent to determine the list price, but express their preferences only at the point at which they must decide whether to accept the offer or not. Those among them who are especially patient will only accept offers that are high with respect to both the list price strategy and unobserved quality, and so will have a high u.

There are a number of identification restrictions implicit in our model. First, we assume that the unobserved quality has an equal effect on the list price and final price. Equivalently, we are assuming that the price premium, the excess of the final price over the list price, is independent of unobserved quality. Second, we assume that unobserved quality does not affect the number of bidders. Given these assumptions and an additional equation that we will introduce forthwith, the model is identified, and we need place no restrictions on how observed attributes affect the prices and the number of bidders. These two assumptions can be dropped while still maintaining identification if they are replaced by alternative assumptions that while permitting a different effect across list and final prices, and an effect on bidder numbers, requires that the relative effects of unobserved quality be identical to those of observed quality. That requires a nonlinear estimation approach, which we will present in our Robustness Section below.

The alternative assumption is based on the reasoning that there is nothing special about attributes that are unobservable to the econometrician – that it is overall quality that matters. One reason that quality might affect the number of bidders and affect the list and final prices differentially is if the ratio of buyers to sellers differs across market segments. Another reason is that the degree of heterogeneity of properties differs across market segment. Of course a linear specification might be inadequate, especially if the middle quality markets are thicker than those at the extremes, but we leave that to a future draft. As it is, we will see that although we can not reject the assumption that quality affects list and final price equally, the estimated effect is so small as to be economically irrelevant. In contrast, the point estimate of the effect of quality on the number of bidders is moderate, but not statistically significant.

A third, less obvious, assumption that we make is a zero correlation between the seller's reservation price and any mechanism other than the list price with which sellers might attract bidders (such as advertising). Thus there is assumed to be no correlation between how much a seller (or a seller's agent) advertises and the seller's reservation price, beyond what is predicted by η. In other words, p depends on w only through b.

The above equations can be written so that all the observed variables are on the left hand side:

$$l = v + \eta$$
$$b = \psi\eta + w$$
$$p = (\phi\psi + m)\eta + \phi w + v + u$$

Thus the variance-covariance matrix of $(l, b, p)'$ is

$$A = \begin{pmatrix} \sigma_\eta^2 + \sigma_v^2 & \psi\sigma_\eta^2 & (\phi\psi + m)\sigma_\eta^2 + \sigma_v^2 \\ & \psi^2\sigma_\eta^2 + \sigma_w^2 & (\phi\psi + m)\sigma_\eta^2 + \phi\sigma_w^2 \\ & & (\phi\psi + m)^2\sigma_\eta^2 + \phi^2\sigma_w^2 \quad + \sigma_v^2 + \sigma_u^2 \end{pmatrix}$$

A is estimable and has six distinct elements. However, the model has seven parameters: the three coefficients (ψ, ϕ, m) and the four variances $(\sigma_v^2, \sigma_\eta^2, \sigma_w^2, \sigma_u^2)$. It is straightforward to see that the system is underidentified by one parameter. However, if an estimate for σ_v^2 can be obtained from outside the system, the system is identified. We do so by appending the equation

$$(4) \qquad l_0 = v + \eta_0$$

where l_0 is the list price from the previous sale of the unit. We assume that v, η_0 and η are mutually independently distributed. (η_0 and η need not be identically distributed.) Thus $Cov(l, l_0) = \sigma_v^2$ and the model is now identified.

The assumption that the covariance between the current and the previous list price arises solely out of unobserved quality requires some justification. In particular, either equity lock-in or loss aversion would lead to a correlation between the previous sale price and the current list price (see Stein 1995 and Genesove and Mayer 1997, 2001) beyond that arising from unobserved quality, while we would expect the previous final price to depend on the previous list price for all the reasons raised in our discussion of the model itself. However those mechanisms are phenomena of markets with declining prices, while the market under consideration was characterized by rising or stable prices over the sample period and many years before that, which together cover the period from which we draw the previous list price.

Were we to consider the four variables together, we would have a four by four variance co-variance matrix, with ten potentially distinct elements. Since the only additional parameter introduced by equation (4) is $\sigma_{\eta_0}^2$, the variance of η_0, the system would be over-identified by two moments. In particular, the model predicts that $Cov(b, l_0) = 0$ and $Cov(p, l_0) = \sigma_v^2$. In principle, we could use those additional moments to improve our estimation of the parameters; however we have very few observations for which we have information on *both* a previous list price *and* the number of bidders. The issue of data availability will be discussed further in the next section.

It is instructive to see what the corresponding ordinary least squares estimation would produce. A naïve approach to estimating the effect of the list price on the number of bidders would be to regress l on b. This would yield the coefficient

$$(5) \qquad \hat{\psi}_{OLS} = \frac{\hat{C}(b,l)}{\hat{V}(l)} = \frac{\hat{C}(\psi\eta+w,\eta+v)}{\hat{V}(v+\eta)} \rightarrow \psi\frac{\sigma_\eta^2}{\sigma_v^2+\sigma_\eta^2}$$

so that the OLS estimator would be biased downwards in magnitude in the manner of an errors-in-variable bias. Simply put, we do not expect the number of bidders to be responsive to variations in the list price per se, but rather to variations around the (not fully observed) mean for that housing type. Thus regressing number of bidders on list price is subject to an errors-in-variable problem.

The OLS estimator for the effect of the number of bidders on the final price is

$$\hat{\phi}_{OLS} = \frac{\hat{C}(p,b)}{\hat{V}(b)} = \frac{\hat{C}(\phi b + m\eta + v + u, b)}{\hat{V}(b)} = \phi + \frac{\hat{C}(m\eta + v + u, \psi\eta + w)}{\hat{V}(b)} \tag{6}$$

$$\rightarrow \phi + m\psi \frac{\sigma_\eta^2}{\psi^2\sigma_\eta^2 + \sigma_w^2} = \phi + (m/\psi) \frac{\psi^2\sigma_\eta^2}{\psi^2\sigma_\eta^2 + \sigma_w^2}$$

So the OLS estimator here is also biased, if the list price strategy has an effect on the number and bidders and a direct effect on the final price (whether through signalling or the selection effect). However, in practice, we will see that this bias is small, mostly because the vast majority of the variance in the number of bidders is random, i.e., due to w. In our sample, the problem with OLS estimation is not the bias, but the lack of precision. The regression error equals $\sigma_v^2 + \sigma_u^2$ and we expect σ_v^2 to be large (and our maximum likelihood estimates will confirm that).

A natural solution to consider in dealing with this lack of precision is to 'correct' the final price for the presence of unobserved quality by subtracting off the list price. This strategy has been taken in any number of studies that have investigated the partial correlation between seller time on the market and price. The regression is

$$\hat{\phi}_{OLS} = \frac{\hat{C}(p-l,b)}{\hat{V}(b)} = \frac{\hat{C}(\phi b + (m-1)\eta + u, b)}{\hat{V}(b)} = \phi + \frac{\hat{C}\big((m-1)\eta + u, \psi\eta + w\big)}{\hat{V}(b)} \tag{7}$$

$$\rightarrow \phi + (m-1)\psi \frac{\sigma_\eta^2}{\psi^2\sigma_\eta^2 + \sigma_w^2} = \phi + ([m-1]/\psi) \frac{\psi^2\sigma_\eta^2}{\psi^2\sigma_\eta^2 + \sigma_w^2}$$

So the OLS estimator here is also biased, if the list price strategy has an effect on the premium on the number of bidders is. Whether the bias is exacerbated or mitigated, obviously depends on how far m is from unity. However, the variance of the regression error is now only σ_u^2.

Yet another solution is to regress p on both b and l. Then we are in the case of a bivariate regression with correlated regressors, one of which, l, suffers from an errors-in-variable problem.

In contrast, maximum likelihood estimation of equations (1)-(3) and (4) will yield consistent estimates of the parameters. This involves finding values of the parameters to match the variance co-variance matrix of (l, b, p), that is $A(\hat{\kappa}) = \hat{C}(l, b, p)$ and $\hat{C}(l, l_0) = \hat{\sigma}_v^2$, where is , $\hat{\kappa} = (\psi, \phi, m, \sigma_v^2, \sigma_\eta^2, \sigma_w^2, \sigma_u^2)$.

Fortunately, the solution can be stated explicitly. First, we regress the list price on the previous list price, in the sample of units that are sold more than once. We take the estimated coefficient, $\hat{r}$, and multiply it by the variance of the residual *previous* list price.[5] This serves as an estimate of the variance of the unobserved quality, $\hat{\sigma}_{v,ML}^2$. The variance of the list price strategy, σ_η^2, is then estimated as the difference between the list price variance and σ_v^2: $\hat{\sigma}_{\eta,ML}^2 = \hat{V}(l) - \hat{\sigma}_{v,ML}^2$. The estimated effect of the list price strategy on the number of bidders, $\hat{\psi}_{ML}$, is the product of the regression of bidders on list price and the ratio of the list price variance to the list price strategy variance:

$$\hat{\psi}_{ML} = \frac{\hat{C}(b,l)}{\hat{V}(l)} \frac{\hat{\sigma}_{v,ML}^2 + \hat{\sigma}_{\eta,ML}^2}{\hat{\sigma}_{\eta,ML}^2} = \frac{\hat{C}(b,l)}{\hat{\sigma}_{\eta,ML}^2}$$

This then allows us to estimate σ_w^2 as the excess of the variance of the number of bidders over that part contributed by the list price strategy: $\hat{\sigma}_{w,ML}^2 = \hat{V}(b) - \hat{\psi}_{ML}^{\;2} \hat{\sigma}_{\eta,ML}^2$.

The remaining coefficients are then estimated as follows.

(8) $$\hat{\phi}_{ML} = \frac{\hat{C}(p,b) - \hat{C}(b,l)(\hat{C}(p,l) - \hat{V}(l)\hat{r})}{\hat{V}(l)(1-\hat{r})} \frac{1}{\hat{V}(b) - \hat{C}(b,l)\hat{C}(b,l)/\hat{\sigma}_{\eta,ML}^2}$$

(9) $$\hat{m} = \frac{\hat{C}(p,l) - \hat{V}(l)\hat{r}}{\hat{V}(l)(1-\hat{r})} - \hat{\phi}_{ML}\hat{\psi}_{ML} = \frac{[\hat{C}(p,l)/\hat{V}(l)] - \hat{r}}{1-\hat{r}} - \hat{\phi}_{ML}\hat{\psi}_{ML}$$

and

(10) $$\hat{\sigma}_{u,ML}^2 = \hat{V}(l) - \hat{\phi}_{ML}^{\;2} \hat{V}(b) - \hat{m}_{ML}^2 \hat{\sigma}_{\eta,ML}^2 - \hat{\sigma}_{v,ML}^2$$

2.2 Structural Measure of Market Thinness

[5] We multiply it by the variance of the previous list price, and not the current list price, to allow for the possibility that $\sigma_\eta^2 \neq \sigma_{\eta 0}^2$.

So far our estimator of market thinness has been based on a reduced-form definition of how much the final price increases with the number of bidders. In this section, we explore a more structural approach to measuring market thinness that attempts to uncover the dispersion in serious bidders' valuations of properties they bid on. To do so, we present a simple independent private value model in the context of housing markets. A seller with one unit of a house for sale sets a list price, l, to alert potential buyers that the house is available for sale. This list price serves, imperfectly, as both a signal that indicates the seller's patience and as an indicator of the quality segment. Potential buyers arrive to inspect the house, and N of them choose to bid. Note these are not randomly chosen bidders, but bidders who have self-selected based on their observation of the list price and the home. The extent one can do is greater in thicker markets. The value x of the house to a bidder is a random variable X distributed according to a continuously differentiable distribution function $G(X;\mu,\sigma)$. μ is a location parameter that summarizes the expected value of the house given its attributes and location, and σ is the scale parameter that reflects how disperse the value distribution is among bidders, so that $G(X;\mu,\sigma) = G_0((X-\mu)/\sigma)$, for some baseline distribution G_0 that we will specify below.

The setting described above can encompass both the English and the sealed-bid auction formats. In light of the standard practice in the market under consideration, we assume that the bidding takes the format of sealed-bid auction (although this assumption is not necessary). This requires us to assume that G is known to all buyers, an assumption that can be at least partially justified by the fact that real estate agents, who advise the buyers, have observed past sales of similar houses, have accumulated knowledge of the distribution of offers likely to be received (Haurin, 1988).

In equilibrium, the house is awarded to the bidder who values it most, and each bidder submits a bid that is equal to the expected second highest valuation conditional on his/her valuation being the highest (McAfee and McMillan, 1988). It is well known that within this standard auction framework, the expected winning bid, Ep, is equal to the expected second highest value of a sample of size N drawn from G, and so equal to

(11) $$Ep = \mu + \sigma h(N)\ ,$$

where $h(N)$ is the expected second order statistic, $E\{X(N-1:N)\}$ for distribution G_0, and N is the number of bidders. This suggests the regression

(12) $p = X\beta + \sigma h(N) + v + u$

where we model the location parameter as $\mu = X\beta + v$. X represents observed attributes and time fixed effects.

We, of course, need to specify G_0. Two classes of distributions suggest themselves, with each corresponding to a different decision environment for the buyer *before* he visits the house. In the first scenario, each buyer looks at the descriptions, visual and/or textual, of a large number (M) of homes, idiosyncratically differentiated, by means of an Internet site, or with the help of an agent. He formulates a willingness to pay Y for each. He is assumed able to visit only one, so he will of course choose that property for which his Y is the greatest. Call that maximum Y, X. For M large, the distribution of X will be well approximated by a location-scale distribution, with G_0 one of the extreme value distributions. For this preliminary version, we will consider *the* extreme value distribution (Gumbel).

We note that the Gumbel distribution has the characteristic that σ does not vary with M. Thus if we would compare estimates from σ from different sized markets, then we would expect to get similar estimates for σ. For the other extreme value distributions, σ will either increase with M (the Frechet distribution) or decrease with M (the Weibull distribution). This last case seems the most reasonable to us, and corresponds to the case in which Y has a bounded distribution.

In the second scenario, each buyer considers, again via visual and/or textual descriptions, a single property each period, and formulates a Y for it. If Y exceeds some threshold value, he will visit the property. As the threshold value increases, the distribution of values of homes that are visited will tend to the one of the Generalized Pareto Distributions. For this preliminary version, we will consider the uniform distribution, which lies in that category, only.

Uniform distribution: $h(N) = 0.5\big((N-2) - (N-1)^2/(N+1)\big)$

Standard extreme value distribution: $h(N) = 0.5772 + Nln(N-1) - (N-1)lnN$

Notice that h is defined for $N > 1$ only: there is no second highest bidder where there is only one buyer. In principle, we could run the regression on the set of observations with competing bidders. However, two-thirds of our observations have no competition, and those observations are useful for estimating β and so improving the precision of our estimate of σ. In order to incorporate those observations, we define

(13) $$\tilde{h}(N) = \begin{cases} h(N), & N > 1 \\ 0, & N = 1 \end{cases}$$

and estimate the following relationship:

(14) $$p_{it} = X_{it}\beta + \sigma\tilde{h}(N_{it}) + \gamma I(N_{it} = 1) + \varepsilon_{it}$$

The goal is to estimate the dispersion of the distribution of valuation of potential buyers who actually bid on a given house. In equation (13), there are two sources of deviations around the mean house value based on housing attributes. The first source is the variation in the number of bids. This is represented by the sum of two terms: $\tilde{h}(N_{it}) + \gamma I(N_{it} = 1)$. The dummy $I(N_{it} = 1)$ = 1 defines the benchmark case where there is only one bidder. When there is an auction (i.e, two or more bidders show up), the dispersion of potential buyers' valuation, conditional on they actually bid, is represented by the product of σ and the standard deviation of the underlying distribution for $h(N)$. Obviously, the estimate of σ is allowed to vary depending on the distribution assumption for $h(N)$. The parameter of interest is σ. A larger σ indicates greater variance in the distribution of bidder valuation and hence a thinner markets.

The second source of deviations of actual house price around the mean value is unobserved quality. This is represented by ε_{it}. In order to obtain a consistent estimate of σ, we need to assume that unobserved quality affects the list price and final price in a proportional way. As we will shown in Section 4.2, this is indeed the case, and therefore a simple and straightforward estimation of equation (13) will yield a consistent estimate of dispersion of buyer valuation.

The bidding process described above is efficient in the sense that the house is awarded to the bidder with the highest valuation. This is efficiency conditional on the bidders who show up.

However, since the winning bidder will not pay the full amount of his valuation, he receives a winner's surplus equal to the difference between his valuation and his expectation of the second highest order statistic. The existence of this surplus allows for the possibility of inefficient entry into the search process by buyers, akin to that of a model of bilateral bargaining in a search framework, as in Hosios. We thus calculate the expected surplus, or the 'money left on the table'. For both of the distributions we consider, it is a function of the number of bidders, and indeed a decreasing function.

3. Data

Our primary data are based on a survey we are presently conducting among recent buyers in a large North American metropolitan area. The addresses of buyers are taken from transaction records of single-family homes available at the local Multiple Listing Service (MLS), covering one-third of the area. Names of these buyers are purchased from the deeds office. Difficulties in lining across the various data sources for condominiums led us to not cover that segment of the market. The universe, the sample and the response rates on the survey are described in Table 1.

From the universe of transaction records, mail samples of 3,523 were drawn at random for 2006, 4,032 for 2007, 6,707 for 2008, and 4,340 for the first three quarters of 2009. For each household, there are at most three rounds of interviews. In the first round, each household in the sample was sent a 4-page questionnaire with a personalized cover letter hand-signed by both authors. In the second round, for those who have not returned questionnaires and whose numbers can be found on the yellow pages, we conduct phone interviews by asking them the same questions in our original questionnaire. In the third round, as yet not done, we plan to mail the duplicate questionnaire with a new personalized cover letter to those who have not responded yet either by phone or by mail.

The overall mail list contains 18,602 addresses, out of which 1,816 addresses are invalid for survey purpose. Among these invalid addresses are some who bought land only, some as institutional buyers, etc. With these excluded, the total number of questionnaires we sent out in the first round is 16,977. A total of 351 surveys were returned "households-moved" or "address unknown" by the Post Office.

In total, 2,894 interviews have been conducted, among which 1725 by mail and 1169 by phone interviews conducted by our research assistants. The overall response rate so far is 17.4%. Given that the second round phone interview is still not completed and the third round follow-up mailed survey has not started yet, this rate should be considered as a lower bound for the final response rate. Although low, this response rate is considerably higher than average response rates of other homebuyer surveys, such as those conducted by the National Association of Realtors, which have been the basis of almost all other surveys of buyers.

Our survey data are complemented with publicly available information from the local MLS, which covers 212,063 transactions that occurred between 2001 and 2009. (Recall that we cover only one-third of the area.) Our survey covers transactions that occurred between January 2006 and September 2009. This is a period that experienced a boom market, followed by a slow and uncertain market trigged by the global financial crisis started in September 2008. However, the market did not experience out of the ordinary rates of foreclosures. During this period the MLS records transactions of 57,431 properties, among which 10,117 properties have been transacted more than once. Table 2 lists the number of transactions for each of these properties during the sample period. Each transaction is characterized by a set of variables, including location, price, time of the sale, and structure.

Properties are identified in the MLS data by district, MLS number, address, unit number (if applicable). For each property, the MLS also defines its geographical coverage in terms of its rows and columns on the map. Using this information, we create a square dummy that captures squares on the map. The overall MLS sample covers 27 districts, which is further divided into 904 squares.

The structure variables include lot front, lot depth, the length and width of the first room, dummy variables for basement, garage space and occupancy. These variables, along with the tax information, provide sufficient information for conducting the hedonic analysis.

Table 3 presents summary statistics of the variables of interest for the overall MLS sample and the Survey sample. Compared with the overall sample, the sample for which we have collected survey responses tend to exhibit 0.5% higher list price, 0.9% lower final price, and 15% higher price premium measured by the difference between the transaction and list price. In term of overall attributes, houses in the survey response sample seem slightly worse than those in the overall MLS sample. However, sampling rates are not uniform over time and our full analysis of the difference between the survey and the MLS population awaits a more careful analysis on our part.

In our surveys, we sought information on home search and bidding behavior. Figure 1 shows a histogram of the number of competing bidders. This variable, the response to the question "Were there other people actively bidding on the home when you submitted your first offer?" and "IF YES, about how many other bidders were there?", has never been explored before in any analysis of residential housing market search (and we suspect in labour economics as well).[6] The figure shows that in two-thirds of the cases, there is no competing bidder. In one-sixth of the cases, there is a single competitor, and in somewhat less than half of that, there are two competitors. There are more than five bidders in three percent of the observations.

In addition, we obtain information about prices through the following questions: (1) How much were you thinking about spending for the home? (2) What was the seller's asking price at the time that you made your first offer on the home? (3) How much was your first offer on the home? (4) What was the final final price of the home purchased? The second and fourth questions provide an independent source of buyer-reported price information for which we can use the MLS-reported price information to verify. Table 4 shows that on average, buyers report about 0.7% higher final prices and 1.7% higher list prices than are recorded in the MLS data. The first and third questions provide information on buyers' expected budget and initial offer price. Together, these questions provide a unique opportunity for us to examine the bidding and bargaining behavior in real estate transactions.

To understand the nature of the competitive environment in the presence of competing bidders, we consider how some additional variables from our interview vary with the number of bidders. First, from Table 5, we see that the gap between the final final price and the initial offer price is smaller when there are competitors. When there are no competitors, 19 percent of the time the final final price equals the respondent's first offer. That fraction nearly doubles to 36 percent when there are competitors. We understand the presence of a gap between final price and first offer as indicating that bargaining has taken place, so that the correlation suggests that bargaining is more likely to have occurred the fewer the number of bidders – with the alternative being some form of auction. This interpretation is further supported by Figure 2, which shows that with one bidder, the spread of the difference between transaction and offer prices is

[6] One might imagine a situation in which a competing bid was made after the respondent made his/her bid.

relatively large, whereas with more than one bidders, the distribution of the price gap is much tighter. Second, Figure 3 shows that the number of days between when the offer was made and when it was accepted falls with the number of competing bidders. That suggests to us the likely presence of an auction where the seller was committed to accept the winning bid, and so there was no possibility of delay, when the buyer reports competing bidders.

Table 6 presents the variance co-variance matrix of the three observed variables in the interview sample. A number of things stand out, which will be reflected later in the estimated parameters of the model. First, the list price variance and the final price variance are of similar size, although the latter is larger. Second, the list price variance and the covariance of the list price with the final price are nearly exactly the same. Third, the covariance of the number of bidders with the final price is an order of magnitude greater than that with the list price.

We also note that the list price residual variance in the survey is .010, while the previous list price residual variance is .012. In contrast, the residual variance in the MLS population is .013, that is, thirty percent more.

Table 7 presents the relationship between the number of bidders and the mean and median of price premium, defined by the difference between log of the final price and log of the list price. Overall, the MLS data reports slightly higher mean premium than the survey responses. As the number of bidders increases, both the mean and the median of price premium increases, lending support to the notion that market are thin and hence each additional bidder increase the premium that the winning bidder has to pay.

4. Results

This section is divided into three subsections. The first two subsections report the estimates of reduced-form measures of market thinness from the OLS regressions and the simultaneous estimation, respectively. The third subsection reports the structural estimates of market thinness and examines the association between buyer surplus and the number of bidders.

4.1 OLS Estimates

Tables 8 and 9 show OLS regressions. They serve two purposes. First, they suggest what we expect to find in the ML estimates. Second, we will see that the OLS results are similar to the ML results, not withstanding the potential endogeneity and errors in variable biases. This will justify our use of OLS estimates later in the paper.

Table 1 considers the regression of the log of the number of bidders on the log list price. Column (1) presents the bivariate regression. To recall, equation (5) shows that the estimated coefficient ought to be downwardly biased in magnitude given the noise in the list price contributed by unobserved quality. The bivariate regression is positive and highly significant, which is hard to interpret. In the remaining columns, we control variously for home attributes, taxes (essentially assessed value), period dummies and neighborhood dummies, the coefficient varies. When all three sets of regressors are included, in the last column, the coefficient has the expected negative sign, but it is imprecise, with a coefficient of -.055 and a standard error of 0.060.

Table 9 presents the regression of the final price on the number of bidders. As shown in equation (6), this should yield consistent estimates only if m equals zero. In all cases, we expect the estimates to be noisy. The bivariate regression is .072 and significant. When we control for attributes alone, the coefficient falls to .038. When we control for attributes, taxes, month dummies, and square dummies, the coefficient falls to 0.024, with a standard error of 0.012, and the corresponding R-squared increases from 0.01 to 0.65.

In Columns (3) and (4) of Table 9, we regress the price premium – the difference between the log of the final price and that of the list price – on the number of bidders. This is equivalent to controlling for list price and imposing its coefficient to 1. As shown in equation (7), this should reduce the bias if m is near one, and in general should substantially reduce the regression error and increase the precision of the estimates. Consistent with our expectation, regressing the price premium on the number of bidders improves the precision tremendously – the standard errors in Columns (3) and (4) are one-sixth of that in Column (2). Comparing Column (4) with Column (3), we find that adding housing attributes, taxes, month dummies and location fixed effects doubles the prediction power, but changes the coefficient on the number of bidders only slightly from 0.037 to 0.034.

Finally, in Columns (5) and (6), we move the list price to the right hand side. Consistent with the imposed restriction in Columns (3) and (4), the estimated coefficient on list price is not significantly different from 1. Thus not surprisingly, the coefficients on the number of bidders remain the same as those in Columns (3) and (4).

Comparing across specifications in Table 9, we also find that the coefficient on the number of bidders in Column (2) is not significantly different from those in Columns (4) and (6), suggesting that including the list price, either as a regressor, or with an imposed coefficient of 1, could pass the Hausman-Wu test. Thus, regardless whether we control for the list price, the OLS regression yields a reliable estimate of the number of bidders on final price. Intuitively, this is probably because the list price reflects the seller's or his agent's estimate of home value, most of which is already captured by assessed tax value, housing attributes, location and transaction period.

4.2 Maximum Likelihood Estimates

The OLS regressions indicate that we can improve the precision in the estimation of the effect of the number of bidders on the final price, although possibly at the cost of a bias. To address this concern, we estimate the simultaneous equation system defined in (1) – (4) with the maximum likelihood method. The results are reported in Table 10, along with bootstrap standard errors. The bootstrap includes the initial regression that generates the residuals for use in the estimation procedure outlined above. We start by presenting our baseline results, that is, assuming that unobserved quality has the no effect on the number of bidders, and the same effect on list as final prices. The autocorrelation parameter r is estimated at 0.566. When that is multiplied by the variance of the (residual) previous list price, .012, we get an estimate for the variance of the unobserved quality of list price of .007. Thus, some 68 percent of the unexplained (current) list price variance is accounted for by unobserved quality. The remainder is the list price strategy.

Our estimate for ψ is -.067, indicating that increasing the list price decreases the number of bidders, but the estimate is immensely imprecise. Why is ψ measured imprecisely? Note that we can write

$$\hat{\psi}_{ML} = [\hat{C}(b,l)/(\hat{V}(l) - \hat{C}(l,l_0))] = (\hat{C}(b,l)/\hat{V}(l))(1-\hat{r})^{-1}.$$

Closer investigation of the bootstrap results shows that $\hat{V}(l)(1-\hat{r})$ is fairly constant around y. The individual components are also fairly constant. $\hat{C}(b,l)$, however, is quite noisy, leading to the large standard error of ψ.

Subtracting ${\hat{\psi}_{ML}}^2 \hat{\sigma}^2_{\eta,ML}$ from the variance of the number of bidders, we obtain the variance of w, the noise in the number of bidders. In particular, the estimated $\hat{\sigma}^2_w$ is 0.213 with a standard error of 0.010. This is substantially larger than the estimated ${\hat{\psi}_{ML}}^2 \hat{\sigma}^2_{\eta,ML}$. Clearly, the variation in the number of bidders is largely dominated by that part that derives from w, i.e. is unrelated to the list strategy (η). This provides an additional reason for why ψ is not precisely estimated.

In contrast to our estimate for ψ, our estimate for ϕ is extremely precise, and at .034 is very close to the OLS estimates. This number implies that doubling the number of bidders will increase the final price by 2.4 percent, on average.

We estimate m at 1.004, with a standard error of .025. Our estimate is not so dissimilar from implicit estimates of it that can be derived from previous papers in which both the list and final prices are regressed on some attribute of the seller that is assumed uncorrelated with buyers' bids. For example, Genesove and Mayer (1997) regress both the list price and the final price on the excess of loan-to-value over 80 percent, and obtain coefficients of .19 and .16 respectively, implying an estimate of m of .19/.16= 1.19. Genesove and Mayer (2001), who also consider loss aversion, find that loan-to-value affects list price and final price equally, consistent with an m of one.

A few details about the identification of m bear highlighting. First, it is possible that the number of bidders is mis-reported. In this case, $\hat{\psi}$ will be biased downwards in magnitude, by the classical errors in variable bias, and thus $\hat{\phi}$ will be biased downwards as well. Consequently, $\hat{m}$ will overestimate the true value of m. From (9) we see that $\hat{m}$ is essentially the regression of the final price on the list price, adjusted for the effects of unobserved quality and the number of bidders.

There is another bias that might operate on the estimation of m. The effect of unobserved quality is removed by the autocorrelation of list and previous list price. However, variation in the improvement and depreciation of units will tend to mean that the true variance of unobserved quality is greater than our measurement. From (9) we see that if $Cov(p, L) > Var(L)$ (the OLS estimate greater than one) this will lead to an underestimate of m; for $Cov(p, L) < Var(L)$ this will tend to overestimate. In either case, the extent of mis-estimation would have to be large in order to change the results. [7]

What can we learn from Table 10? First, the estimate for the parameter of interest, $\phi,$ is statistically significant, economically substantial, and robust to endogeneity concerns due to unobserved house quality. Recall that ϕ is a reduced-form measure of market thinness. An estimate of 0.034 implies that doubling the number of bidders will increase the final price by 2.4 percent, on average. Thus the larger number of bidders, the larger is the deviation of the actual house price from its baseline value predicted by the house's location and structure, consistent with the notion that housing market are thin in nature.

Another interesting finding from Table 10 is that the maximum likelihood estimate of the effect of the number of bidders on the final price is very close to the corresponding OLS estimates reported in Table 9. Recall that the OLS estimate of ϕ are labeled "naïve" because the unobserved house quality and endogenous list strategy are not controlled. Interestingly, the ML estimate of ϕ is about the same as the OLS estimates reported in Columns (4) and (6) of Table 8. Clearly, this seemingly surprising equivalence between the OLS and ML estimates arises out of the finding that ML estimate of m is close to 1. As shown in equation (8), as long as m is not far from 1, then one should expect a consistent estimate of $\hat{\phi}_{OLS}$ yielded by OLS regressions where the list price is controlled for. An ML estimate of m of close to 1 is fortuitous, although it is arguably what one would expect from a signaling explanation. It is also consistent with a selection bias if the distribution of winning bids follows an exponential distribution.

Finally, Table 10 also reveals how the seller's list strategy affects the final price. It follows from equations (2) and (3) that

$$\frac{\partial p}{\partial \eta} = \hat{\psi}\,\hat{\phi} + \hat{m}\,. \tag{14}$$

Equation (14) makes clear that the effect of an increase in list strategy (η) on the final price is the sum of two opposite effects. The first term, $\hat{\psi}\,\hat{\phi},$ captures the effect of intensified bidding competition through a lower listing price, holding constant the house quality. Although both $\hat{\psi}$

[7] We can get a reasonable value for $Var(v)$ by considering the variance in expenditures on housekeeping as a function of home value. Since home repairs are likely to be lumpy, a single cross section will greatly overestimate the variance.

and $\hat{\phi}$ are sizeable individually, the product of the two estimates amounts to -0.0014, which is clearly dominated by the second term $\hat{m}$, which is about 1.004. Note that $\hat{m}$ itself represents the sum of two effects: signaling and selection. Controlling for observed housing attributions, location, and market conditions, a higher list price can signal that the seller is more patient, and thereby causing bidders to submit a higher bid. A higher list price also sets a higher minimum for acceptable bids, thus only bidders with sufficiently strong desire for the house would submit bids. Although we cannot distinguish between these two different sources of positive effects, our estimates clearly show that, for an average sellers, held house quality constant, the loss from lowing the list price through signaling and selection strongly dominates the possible gain through induced bidding competition effect

4.3 Structural Estimates

So far we have been relying on a reduced-form measure of the market thinness, namely, how much the final price increases with the number of bidders. Section 2.2 presents a structural model that allows us to directly infer market thinness by recovering the standard deviation of the distribution of potential buyers' valuation conditional on those that actually bid on the house. We rely on the closeness of the OLS and ML estimates established in Section 4.2 to justify the use of OLS approach in obtaining our estimates.

In Table 11, we report estimates from several variants of equation (12). Column (1) reports the benchmark case where the final price is regressed on a set standard controls, including housing attributes, tax assessments, neighborhood and year effects. This is a typical house price hedonic model. In the absence of unobserved quality and search frictions, the model would provide reliable estimates of home value.

In Columns (2) and (3), we account for search frictions by including a dummy that indicates whether the number of bidders is one, and a second order statistics $\tilde{h}(N)$ defined in equation (12). The parameter of interest is σ, the coefficient on $\tilde{h}(N)$, which reflects how spread the value distribution is among potential bidders in the event of a bidding war. Its consistency is assured by the equivalence result between the OLS and ML estimates of the effect of the number of bidders. Bidders' valuation is assumed to be drawn from a standard extreme value distribution in Column (2) and from a uniform distribution in Column (3). In both specifications, the estimates of σ are positive and statistically significant, providing evidence for the presence of the price premium that is dependent on the number of bidders in a bidding war.

To obtain the estimate of the standard deviation of the distribution of valuations of potential buyers who actually bid on the home, one needs to multiply the estimated σ by the standard deviation of the corresponding underlying standard distribution. The latter is $\pi/\sqrt{6}$ in the case of the standard extreme value distribution, and $1/\sqrt{12}$ in the case of the uniform distribution. In Column (2), the estimate of ϕ is 0.038, which implies a standard deviation of 0.049. For the standard extreme value distribution, this implies that difference in valuation between the 25th and the 75th percentile is equivalent to 7.71 percentage points (.049*(ln(-ln(.25))-ln(-ln(.75)))). In Column (3), the estimate of ϕ is 0.145, which implies a standard deviation of 0.042. For the uniform distribution, this implies that the range of the distribution of bidders' valuation is 4.2 percentage points. Both estimates suggest that markets are thin and hence there is consideration amount of variation among buyers valuation of the same house.

In Column (4), we further explore a non-parametric method by approximate the second order statistics with a set of dummy variables for the number of bidders. Consistent with what we expected, an increase in the number of bidders increases the price premium, although such relationship is not strictly monotonic. Compared to the benchmark case where there is only one bidder, having one additional bidder increases the premium significantly by 1.7%, while having two additional bidders only increases the premium by 0.8% in an insignificant way. However, the increase in price premium jumps to 4.2% when the number of bidders is four, and 8.4% when the number of bidders is 5. In the rare case when the number of bidders exceed 7, the premium increased by over 10%. Overall, these estimates lend strong support to the hypotheses that in markets where search is costly, the winning bidder pays a substantial amount of premium that exceeds the hedonic value of the home, and that such premium increases with the number of bidders that he competes.

The estimates in Table 11 can be used to compute how much money is left on the table in the event when there is a bidding war. The auction setting described in Section 2.2 predicts that the buyer (winning bidder) surplus is the difference between his valuation and the expectation of the second highest order statistics conditional on his valuation being the highest. Applying this formula, we compute the buyer surplus in a bidding war under various assumptions on the value distribution and the number of bidders. The results are reported in Table 12. When there are two bidders, we can approximate the buyer surplus by $0.038(E[X(2{:}\,2)] - E[X(1{:}\,2)])$, which is equivalent to 0.053 if bidders' value is drawn from the standard extreme value distribution; and 0.048 if it is drawn from the uniform distribution. These estimates amount to about 5% of home value, which is $21,081.75 at the sample mean. Note that the standard extreme value and uniform distributions predict similar amount of money left on the table when there are two bidders, which accounts for over 45% of the bidding wars. As the number of bidders increases, the predicted amount of money left on the table under two distributions get far apart as there is less data. For example, when the number of bidders increases to four, the buyer surplus decreases to 4.2% under the standard extreme value distribution, and 2.4% under the uniform distribution. In both cases, there is a clear pattern that buyer surplus decreases substantially with the number of bidders, and such decrease is marginally decreasing. The finding is consistent with the notion that the winning bidder's expected surplus is small when the bidding is well attended.

5. Conclusion

This paper makes two contributions. Substantively, it provides credible estimates of the thinness of real estate markets. A large body of housing literature on market efficiency and liquidity builds on the assumption that real estate markets are thin, though there is few empirical work that examines how thin real estate markets are. We tackle this important question by conducting a new survey among recent home buyers and by developing an econometric framework to estimate the market thinness. We find that doubling the number of bidders increases the final price by 2.4 percent, on average. In addition, the range of the distribution of bidders' valuation varies from 4.2% (under the uniform distribution assumption) to over 7.7% of the final price (under the standard extreme value distribution). Clearly, these estimates reflect a substantial amount of dispersion among buyers valuation for the same house, thereby establishing solid evidence for the thinness of real estate markets. Given the increasing research attention on the frictions in housing markets, our estimates should prove useful in future search-based calibration models of housing markets.

The second contribution of the paper is methodological. We demonstrate that our estimates of market thinness are robust both to the presence of unobserved house quality and to the bidders' endogenous response to list strategy. Moreover, we show that market thinness itself is an important determinant of house prices. Conventional hedonic estimation techniques model house price as a function of housing attributes. However, in thin markets housing attributes alone may not be enough to explain house prices. In the event of a bidding war, the bidder who values the home most is willing to pay a price that exceeds the hedonic value of the house, and the premium he pays depends on the number of bidders he competes with. By explicitly incorporating search frictions into standard hedonic price estimation, the structural price estimation method developed in this paper corresponds more closely to the market value of houses in thin markets.

References

Albrecht, James, Axel Anderson, Eric Smith and Susan Vroman, "Opportunistic Matching in the Housing Market." *International Economic Review*, 2007, 48 (2), 641-774.

Albrecht, James, Pieter Gautier and Susan Vroman, "Directed Search in the Housing Market." International Economic Review, 2010, working paper.

Berkovec, James and John Goodman, "Turnover as a Measure of Demand for Existing Homes." 1996, *Real Estate Economics*, 24 (4), pp. 421-440.

Case, Karl and Robert Shiller, "The Efficiency of the Market for Single-Family Homes." *New England Economic Review*, 1989, 79, pp. 125-137.

Delacroix Alain and Shouyong Shi (2008), "Pricing and Signaling with Frictions." mimeo.

Genesove, David and Christopher Mayer, "Equity and Time to Sale in the Real Estate Market." *American Economic Review*, 1997, Vol. 87(3), pp. 255-269.

Genesove, David and Christopher Mayer, "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *Quarterly Journal of Economics*, 2001, Vol. 116(4), pp.1233-1260.

Krainer, John, "A Theory of Liquidity in Residential Real Estate Markets", *Journal of Urban Economics*, 2001, Vol. 49, pp.32-53.

Mayer, Christopher and Tsuriel Somerville, "Land use regulation and new construction," *Regional Science and Urban Economics*, 30(6), pp. 639-662

Merlo, Antonio and Francois Ortalo-Magné (2004), "Bargaining over Residential Real Estate: Evidence from England." *Journal of Urban Economics*, 56, 192-216.

Moen, Espen, Competitive Search Equilibrium, Journal of Political Economy, 1997, 105, pp. 385-411.

Peters, Michael, "Common Agency and the Revelation Principle." *Econometrica*, 2001, 69 (5) pp 1349-1372.

Stein, Jeremy, "Prices and trading volume in the housing market: A model with down-payment effects." *Quarterly Journal of Economics*, 1995, Vol. 110(2), pp. 379-406.

Table 1: Survey Universe, Samples and Response Rates

Year	Month	MLS Sample (overall)	MLS Sample (full info.)	Mailed Sample	Bad Addr.	Total Sent	Returned Unusable	Mailed Response	Phone Response	Total Response
2006	Jan-Dec	23,205	19515	3523	238	3285		341	341	682
2007	Jan-Dec	25,753	20026	4032	238	3794		378	361	739
2008	Jan-Dec	19,561	14861	6707	696	6202		674	333	1007
2009	Jan-Sep	23,368	12973	4340	644	3696		332	134	466
Total		91,187	67375	18602	1816	16977	351	1725	1169	2894
Response Rate (so far): 17.41%										

Table 2: Frequency of Transactions

properties transacted once	47,414
properties transacted twice	9,087
properties transacted three times	943
properties transacted more than four times	87
properties in total	57,531

Note: The survey covers transactions that occur between Jan. 2005 and Sep. 2009, but the transaction history tracks back to Jan. 2001.

Table 3: Descriptive Statistics Across Samples

Variables	MLS Sample	Survey Response Sample
Original Price	$441,222.9 (191.940.3)	$430,617.8 (235,144.7)
List Price	$425,811.7 (255797.7)	$427,867.8 (234186)
Final price	$417,905.8 (248719.7)	$421,634.9 (230955.8)
Premium (lnTRAN-lnLIST)	-.1.91% (4.15%)	-.1.63% (4.44%)
# of bedrooms	3.36 (.74)	3.33 (.71)
# of washrooms	2.87 (.97)	2.82 (.96)
Lot front (feet)	42.56 (80.81)	41.34 (59.90)
Lot depth (feet)	119.95 (113.98)	119.90 (102.39)
Room1 length (feet)	16.12 (19.91)	15.45 (6.09)
Room1 width (feet)	12.01 (22.13)	11.66 (19.00)
Garage space	1.29 (.76)	1.26 (.77)
Taxes	3585.93 (6945.57)	3156.97 (1554.56)
Days on market	29.47 (30.55)	26.72 (27.15)
# of districts	27	27
# of squares (neighborhoods)	835	515
Period covered	Jan, 2006-Sep, 2009	Jan, 2006-Sep, 2009
# of observations	67375	2894

Table 4A: Comparing Survey Responses and MLS Reports of Price Information

	Survey Response	MLS Reported Value	Observations
Final price	$422,852.3 (223,138.8)	$419,784.2 (228,012.8)	2569
List Price	$433,123.4 (235,011.2)	$425,775.1 (231,607.3)	2572
Premium	-2.19% (11.57%)	-1.56% (4.56%)	2531

Table 4B: Descriptive Statistics of Responses to Some Survey Questions

Selected Questions	Mean	S.D.	Responses
"How much was your first offer on the home?"	414,789	218,219.4	2525
"Were there other people actively bidding on the home when you submitted your first offer?"	34.67%	47.60%	2801
"If yes, how many bidders were there?"	1.85	2.32	1085
"How long did you actively search before locating the home you purchased?" (in days)	100.03	157.33	2784
"About how many homes, including the one you bought, did you visit before making your purchase?"	18.61	27.78	2866
"How many homes, including the one you bought, did you make offers on?"	1.56	1.41	2880
"How many days after you made your first formal offer was it accepted?"	3.45	7.35	2817

Table 5: Final Price vs. Offer Price

	Final < Offer	Final = Offer	Final > Offer
N = 1 (No Auction)	71 (4.44%)	297 (18.56%)	1,232 (77.00%)
N > 1 (Auction)	28 (3.56%)	284 (36.09%)	475 (60.36%)

Table 6A: Full Covariance Matrix (Survey Sample)

	lnASK	lnFINAL	lnBIDDER
lnASK	.1389		
lnFINAL	.1387	.1443	
lnBIDDER	.0076	.0209	.3358

Table 6B: Full Covariance Matrix (MLS Sample)

	lnFINAL	lnLIST	lnLIST_PREVIOUS
lnFINAL	.1231		
lnLIST	.1208	.1199	
lnLIST_PREVIOUS	.1051	.1048	.1117

Table 6C: Residual Covariance Matrix (MLS Sample)

	Ln List Price res.	Ln Trans. Price res.	Ln No. of Bidders res.
Ln List Price residual	.0075		
Ln Final price residual	.0077	.0088	
Ln Number of Bidders	.0007	.0059	.1664

Table 6D: Residual Covariance Matrix (Survey Sample)

	Ln List Price res.	Ln Final Price res.	Ln No. of Bidders res.
Ln List Price residual	.0099		
Ln Final price residual	.0100	.0114	
Ln Number of Bidders	- .0002	.0077	.2304

Table 6E: Residual Covariance Matrix (Survey Sample, with MLS-Reported Price)

	Ln List Price res.	Ln Final Price res.	Ln No. of Bidders res.
Ln List Price residual	.0140		
Ln Final price residual	.0141	.0160	
Ln Number of Bidders	- .0002	.0077	.2304

Table 7: Price Premium, by Number of Bidders

Number of Bidders	Survey Responses		MLS Reported Values		Number of Observations
	Mean	Median	Mean	Median	
N=1	-3.61%	-2.85%	-2.63%	-2.40%	1615
N=2	-1.34%	-0.95%	-0.95%	-1.34%	396
N=3	-0.17%	-0.85%	0.07%	-0.78%	170
N=4	0.28%	0%	0.64%	-0.09%	108
N=5	4.73%	3.37%	4.75%	2.87%	52
N=6	3.84%	4.18%	.4.16%	3.59%	23
N=7	5.78%	6.40%	5.91%	5.09%	15
N=8	10.19%	9.78%	6.99%	3.36%	11
N=9	9.82%	6.87%	12.66%	9.53%	22

Table 8: Bidder Regression (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Number of Bidders					
List Price	0.085 (0.034)	0.3754 (0.0621)	0.090 (0.034)	0.363 (0.061)	0.090 (0.034)	-0.055 (0.060)
Tax & Tax year	No	Yes	No	Yes	Yes	Yes
Attributes	No	No	No	Yes	Yes	Yes
Period dummies	No	No	Yes	No	Yes	Yes
Square dummies	No	No	No	No	No	Yes
R-squared	0.0012	0.0216	0.0518	0.0705	0.0319	0.3126
# observations	2184	2184	2184	2184	2184	1979

Table 9: Final Price Regression (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Transaction Price		Tran. - List		Transaction Price	
Bidders	0.072 (0.014)	0.024 (0.012)	0.037 (0.002)	0.034 (0.002)	0.037 (0.002)	0.034 (0.002)
List Price					1.005 (0.002)	0.996 (0.004)
Tax & Tax year	No	Yes	No	Yes	No	Yes
Attributes	No	Yes	No	Yes	No	Yes
Period dummies	No	Yes	No	Yes	No	Yes
Square dummies	No	Yes	No	Yes	No	Yes
R-squared	0.0123	0.6515	0.2277	0.4760	0.9886	0.9914
# observations	2184	1979	2184	1979	2184	1979

Table 10: Maximum Likelihood Estimates

$\hat{\psi}$	-.067
	(.347)
$\hat{\phi}$	.034
	(.003)
$\hat{m}$	1.004
	(.027)
$\hat{\sigma}_v^2$	.007
$\hat{\sigma}_\eta^2$	.006
	(.002)
$\hat{\sigma}_w^2$	.213
	(.010)
$\hat{\sigma}_u^2$	.002
	(.0001)
r	.566
	(.011)

Bootstrapped Standard Errors appear in parentheses beneath the estimate coefficients.

Table 11: Structural Estimation

Dep. Variable	Final price			
	(1)	(2)	(3)	(4)
$\tilde{h}\,(N)$		0.038 (0.010)	0.145 (0.037)	
I(N= 1)		-0.015 (0.008)	-0.034 (0.007)	
N= 2				0.017 (0.010)
N=3				0.008 (0.013)
N=4				0.042 (0.016)
N=5				0.084 (0.023)
N=6				0.054 (0.031)
N=7				0.093 (0.043)
N=8				0.130 (0.055)
N=9				0.100 (0.029)
Tax & Tax year	Yes	Yes	Yes	Yes
Attributes	Yes	Yes	Yes	Yes
Period dummies	Yes	Yes	Yes	Yes
Square dummies	Yes	Yes	Yes	Yes
R-squared	0.9119	0.9139	0.9139	0.9144
# observations	1929	1929	1929	1929

Table 12: Money Left on the Table

	Money Left on the Table: E[X(N)]-E[X(N-1)]			
N	Standard Extreme Value	Standard Extreme Value With sigma=.038	Standard Uniform	Standard Uniform With sigma=.145
2	1.39	.053	.333	.048
3	1.22	.046	.250	.036
4	1.15	.044	.200	.029
5	1.12	.042	.167	.024
6	1.09	.042	.143	.021
7	1.08	.041	.125	.018
8	1.07	.041	.111	.016
9	1.06	.040	.100	.015
10	1.054	.040	.091	.013

Figure 1: Histogram of Number of Bidders

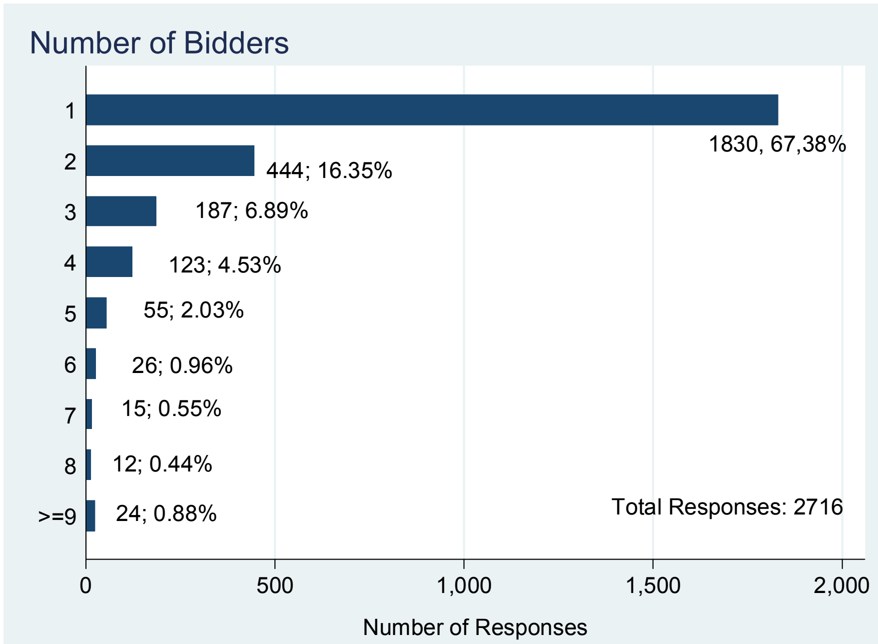

Figure 2: Final price -List Price, By Bidders

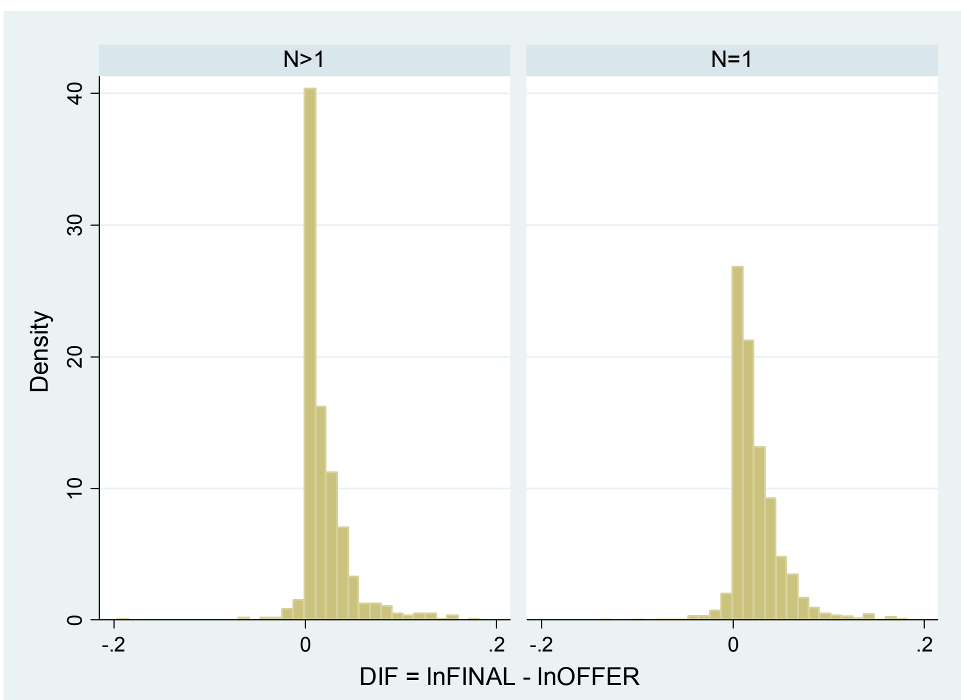

Figure 3: Average Days Until Offer Accepted, By Bidders