Measuring the Thinness of Real Estate Markets*

(Preliminary Version)

David Genesove
Hebrew University of Jerusalem
and CEPR
genesove@pluto.mscc.huji.ac.il

Lu Han
Rotman School of Management
University of Toronto
lu.han@rotman.utoronto.ca

June 19, 2011

*We thank Jim Albrecht and participants from the UW-Fed Housing-Urban-Labor-Macro spring conference for helpful conversations and comments. We also thank Travis Chow, Matthew Derricott, Vincent Guo, Palak Kapasi, Jih-Hong Mao, Pearl Mehra, Areeb Urfani and Shirley Wang for their valuable assistance in conducting surveys. Financial support for this research from the Social Sciences and Humanities Research Council of Canada is also gratefully acknowledged.

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?

This project answers this question through estimation of parameters that capture 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 the number of buyers competed against in purchasing the home. We also collect information about the final and list price and time to sale, as well as buyer demographic information. 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 estimate two measures of market thinness, both based on how the final price increases with the number of bidders. The first uses a log-linear approximation, while the second relies on functional forms implied by the expected second order statistic from a sample size of the number of bidders. The second measure, which is implied by basic auction models, allows us to uncover the dispersion in bidders' valuation of a given property.

The intuition behind inferring the degree of market thinness from how price reacts to the number of bidders is straightforward. If homes are not very different one from another, 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

---

[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.

sellers' search valuations will differ very little, so that the bargaining outcome when only one buyer shows up will differ little from the price when two buyers compete.

However, obtaining an estimate of the causal effect of bidder numbers on the final price is not straightforward. Although under the assumptions employed in this paper, regressing the final price on the number of bidders will yield a consistent estimate of this effect, unobserved house quality typically varies so much as to render it extremely imprecise. An alternative often used in the analysis of the determination of the final price in a housing search environment is to normalize sale price by list price, on the argument that the latter will incorporate unobserved quality. That unfortunately introduces an upward bias, if, conditional on housing attributes observable to buyers, a lower list price induces more buyers to consider the house.

In this paper, we propose a method to address these issues. Our solution relies on a simultaneous estimation of list price, number of bidders, and sale price, by maximum likelihood. We will show how the OLS bias of the effect of bidders on the sale 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 sale price say as when buyers bid higher in response to a higher list price, can also be identified.[4]

We find that, on average, doubling the number of bidders increases the sale price by 2.4%. This effect is statistically significant. Hence we can reject the hypothesis of a thick market in which buyers' valuations are homogenous and so that the final price is invariant to the number of bidders. In term of its magnitude, this effect is nearly equal to half the total compensation that intermediaries earn on the transaction. Nevertheless, the usefulness of this estimate is limited in that it does not give us a measure of the degree of heterogeneity in buyers' valuation, making the economic significance of the deviation from an ideal market difficult to gauge.

. In addition to generating reduced-form estimates of market thinness, the maximum likelihood model also allows us to empirically investigate how the list strategy affects the

[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.

transaction price. Note that how buyers respond to the list price 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, 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.

The auction model based measure is more direct in that it does allow us to uncover the variance in the underlying distribution of match qualities. Conceptually, the thinness of the housing market is an outcome of the heterogeneity of housing attributes on one side of the market, combined with the heterogeneity of buyer tastes on the other. Together, these heterogeneities make finding a suitable match difficult and make the number of houses that are suitable for a specific buyer small. Under certain distribution assumptions, the more homes that a buyer considers before visiting or bidding, the more compressed is the distribution of valuations of those buyers that will show up to bid on a given house, and hence the thicker is the market. Thus, a larger variance in match qualities, reflected by a greater dispersion in the distribution of bidders' valuation of a given house, indicates a thinner market.

This second measure builds on the basic insight from standard auction models that in equilibrium, the house is awarded to the bidder who values it most,[5] and the expected winning bid equals the expected second order statistic from the bidders' valuation. Estimating this relationship, and assuming that the distribution of 'serious' bidders' values is uniform, we find that the spread of bidders' valuation, measured by the standard deviation of the distribution, is

[5] This is so for English auctions. This condition may not hold for sealed bid auctions, when valuations are affiliated (see the recent paper by Menicucci (2009)). The auctions here can be considered as a mix of English and sealed bid auctions (see the description later in the paper).

4.2 percent; assuming the extreme value distribution, we obtain a standard deviation of of 4.9%. In levels,these estimates correspond to $17,760 and $20,for an average house in the sample.

Having a measure of market thinness is tantalizing, because it could provide insights into a series of important questions. For example, despite much theoretical work that emphasizes increasing returns to scale in housing markets, there is little evidence of its empirical relevance. Nor do we know much about how much is the money left on the table in a real estate bidding war and how the total surplus is split between buyers and sellers. Moreover, given that the home matching process entails many real estate brokers who operate independently from each other, there are obviously gains from a coordinated matching scheme (although perhaps at the price of diminished incentives for effort). Our structural estimates of market thinness allow us to bound these gains from above, as well as shed some light on the other questions.

We first provide an argument for the dependence of our measure of market thinness on market size. In markets with more sellers and so more homes for the buyer to prescreen before going out to visit one, and under reasonable distributional assumptions, “serious” buyers who end up coming to visit and perhaps bid on a specific house, will have valuations that are clustered together. Thus thicker markets should have less variance in the underlying matching qualities and higher matching rates. Previous work has often built increasing returns into the home matching function or modeled increasing returns by a first order stochastic dominance shift of the match quality distribution (Diamond, 19xx, and Ngai and Tenreyro, 2010). Consistent with our hypothesis, we find that the estimated dispersion in bidders’ valuation is substantially smaller in summers (versus winters) and in districts or periods with a larger number of listings.

Second, the magnitude of market thinness is also crucial for understanding the potential gains from policies that aim to enhance efficiency in the home search and matching process. An upper bound for the value of improvements in search and matching is the difference between the surplus under some ideal matching process and a measure of the actual surplus. One approach is to assume that the match quality distribution is bounded from above. With a continuum of buyers and sellers, the matched pairs will obtain the maximum match quality under the ideal process. Then following Wolinsky (1998), we compare the market under our investigation with a frictionless market where there are only perfect matches, and compute the mismatching costs as a function of the estimated buyer valuation dispersion, buyer-seller ratio, and the threshold value beyond which match occurs --- all these variables can be recovered from our structural model.

The magnitude of mismatching costs thus serves as an upper bound for the gains from any efficiency enhancing policies in housing markets.

In addition to their theoretical and policy implications, the estimates of market thinness also have practical relevance for market participants. For example, interesting questions in any bidding situation are how much the buyer gains from winning the bidding war, and how that gain changes with the number of bidders he/she competes with. Our estimates for the uniform (?) distribution are that the buyer's surplus, measured by the difference between his valuation of the house and the price he pays, is 5.3% of the sale 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 econometric model that generates an intuitive measure of market thinness. We then describe a simple auction framework that guides the subsequent structural estimation exercise and aids in interpreting the results.

2.1 Log-Linear Specification

As discussed in the Introduction, intuition suggests that a market is thin if an additional bidder increases the sale price substantially. In this section, we lay out an econometric strategy to estimate the effect of the number of bidders on the final sale price with reasonable precision, in the presence of substantial unobserved housing attributes, 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.

2.1.1 Simultaneous Estimation Model

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. As we shall see, 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 sale price. The model starts with the following:

(1) $$L_{it} = Z_{it}\,\gamma^L + e_t^L + \, l_{it}$$

(2) $$B_{it} = Z_{it}\,\gamma^B + e_t^B + \, b_{it}$$

(3) $$P_{it} = Z_{it}\,\gamma^P + e_t^P + \, p_{it}$$

where L_{it} is the list price, B_{it} is the number of bidders and P_{it} is the sale price. All three variables are in logs. We use subscript *it* to indicate house *i* sold at time t. The vector Z_{it} represents observed housing attributes and location fixed effects; while e_t represents period dummies.

Our analysis focuses on the residuals l_{it}, b_{it}, and p_{it} obtained from theregressions in (1)-(3). We further model each of them as a linear function of latent variables that represent unobserved housing attributes (v), the list price strategy (η), the realized deviation of the number of bidders around the expected mean (w), and the realized deviation of the price around its mean (u).[6] Exclusion restrictions ensure that the parameters are identified.

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

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

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

[6] For the ease of exposition, we drop the subscript *it* hereafter.

Equation (4) states that the residual list price is composed of a house 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 (5) states that the residual expected number of bidders is proportional to the list price strategy (η). The factor of proportionality is ψ; presumably, it is negative, as intuition would suggest, and as is predicted on and off the equilibrium path in Peters (1991) and elsewhere. Deviations around the expected number of bidders 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.

The equation that is central to measuring market thinness is Equation (6). It states that the residual sale price is comprised of four distinct components. First is the number of bidders (b), with the accompanying coefficient ϕ. As discussed earlier, we interpret ϕ as a reduced-form measure of market thinness. 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 predict a negative ϕ, but our sense is that the common value aspect is rather limited in this market.

Second is the list price strategy (η), with the accompanying coefficient m. The parameter m reflects both the causal effect of the list price on bids and a non-causal selection effect. First, a higher list price can signal that the seller is more patient (à la Albrecht et al). Such a signal can cause a single bidder to offer more in negotiation with the seller when there is no bidding war, or a winning bidder to submit a higher bid when there are multiple offers. 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. A higher list price can also decrease the number of bidders and thereby reduce the bidding competition effect on the sale price. Thus the list price strategy can operate both indirectly through its effect on the number of bidders and directly though its effect on buyers' offers. The former effect is captured by $\psi\phi$, while the latter would be captured by m in the absence of selection.

That last qualification is needed since m also includes 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 sale 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 third component in Equation (6) is the unobserved house quality (v). Although the hedonic regression literature is generally concerned with omitted housing attributes, with the assumption that the unobserved house attributes are mean independent of observed attribute frequently criticized in the literature (Small, 1975; Bajari, et. al. 2010), that issue does not arise here, as we are not interested in the true causal effect of any given attribute on outcome, but rather on netting out the projection of the outcome on the attribute

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 seller heterogeneity effects not captured by the list price strategy, or via w, should the seller's list price 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 housing attributes, and so will have a high u.

2.1.2 Identification Strategy

Our identification strategy is based on a number of restrictions. First, we assume that the unobserved house quality has an equal effect on the list price and sale price. Equivalently, we are assuming that the price premium, the excess of the log sale price over the log list price, is independent of unobserved quality. Second, we assume that unobserved house quality does not affect the number of bidders. A third, less straightforward, 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. Under these three assumptions, the model is identified, and we need place no restrictions on how observed housing attributes affect the prices and the number of bidders.

One might wonder whether the first two assumptions are too restrictive. One reason that house quality might affect the number of bidders, the list price and the sale price differently is that the ratio of buyers to sellers differs across market segments. Another reason is a differing degree of heterogeneity of properties 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

In the Robustness Section below [NOT IN THIS DRAFT], we will test the first two assumptions explicitly. We will further show that these two assumptions can be replaced by a less restrictive assumption. This alternative assumption permits the unobserved house quality to affect the list price, sale price and the number of bidders freely, but requires that the relative effects of unobserved house quality on any two of these dependent variables be identical to those of observed quality. Estimating the model under this alternative assumption requires a nonlinear approach, which will also be presented later in the paper. we find that although we can 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.

The above equations can be written so that only the latent variables appear on the right 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 under-identified 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

$$l_0 = v + \eta_0 \tag{7}$$

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.) This establishes the 7th moment condition: $Cov(l, l_0) = \sigma_v^2$, 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 sale price to depend on the previous list price for all the reasons raised in our discussion of the model itself – leading to a correlation between the current previous list price. 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 (7) 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; we do not do so as we have very few observations for which we have information on *both* the previous list price (l_0) *and* the number of bidders (b). The issue of data availability will be discussed further in the next section.

2.1.3 OLS Estimates vs. Maximum Likelihood Estimates

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

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

so that the OLS estimator would be biased downwards in magnitude in the manner of an error-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 of the list price around the (not fully observed) mean for that housing type.

The key parameter of interest is ϕ, which estimates how the sale price changes with the number of bidders. Regressing the sale price on the number of bidders, we would obtain an OLS estimator as following:

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

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

So the OLS estimator here is also biased, if the list price strategy has an effect on the number of bidders ($\psi \neq 0$) and either a direct signaling effect on the sale price or a selection effect ($m \neq 0$,). However, as we will show later, in practice this bias is small, mostly because the vast majority of the variance in the number of bidders is random, i.e., due to w. With a large and dominating σ_w^2, $\frac{\psi^2\sigma_\eta^2}{\psi^2\sigma_\eta^2+\sigma_w^2}$ is close to zero, making the magnitude of the bias quite small.

Thus, the problem with the OLS estimate of market thinness in our sample is not bias, but rather lack of precision. In particular, the regression error from regressing the sale price on the number of bidders equals $\sigma_v^2 + \sigma_u^2$. As shown by our maximum likelihood estimates later, σ_v^2 is large, indicating that unobserved house quality varies so much that renders the OLS estimate of ϕ extremely imprecise.

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

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

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

So the OLS estimator here is also biased, if the list price strategy has a non-unitary effect on the premium ($m \neq 1$) and has a non-negligible effect on the number of bidders ($\psi \neq 0$). Whether the bias is exacerbated or mitigated, obviously depends on how far m is from unity. When m is not far from unity, the OLS estimator of ϕ from regressing the premium on the number of bidders would be nearly consistent. Moreover, the estimator would be much more precise because the variance of the regression error is now reduced to σ_u^2.[7]

Yet another OLS solution is to regress the sale price (p) on both the number of bidders (b) and the list price (l). Then we are in the case of a bivariate regression with correlated regressors, one of which, l, suffers from an error-in-variable problem. This regression is less restrictive than the previous one, as we allow the coefficient on the list price to be freely estimated. However, as long as this coefficient is not far from one, the property of the OLS estimator of ϕ from this regression is similar to that in the previous regression.

In contrast, maximum likelihood estimation of equations (4)-(7) will yield consistent estimates of the parameters. By the invariance principle, 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 $\hat{\kappa} = (\psi, \phi, m, \sigma_v^2, \sigma_\eta^2, \sigma_w^2, \sigma_u^2)$.

Fortunately, the solution can be stated explicitly. First, we regress the list price on the previous list price, in the sample of housing units that are sold more than once. We take the estimated coefficient, $\hat{r}$, and multiply it by the variance of the residual of the *previous* list price.[8] This serves as an estimate of the variance of the unobserved house 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 the list price and the ratio of the list

[7] Note that σ_v^2 gets cancelled out under our first identification assumption.

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

price variance to the list price strategy variance:

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

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

The remaining coefficients are then estimated as follows.

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

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

and

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

2.2 Structural Measure of Market Thinness

So far our estimator of market thinness has been based on a reduced-form definition of how much the sale price increases with the number of bidders. In this section, we explore a more structural approach to measure market thinness.

2.2.1 Conceptual Framework

Conceptually, what makes a market thin is not necessarily the size of the market (say, a small total number of buyers and sellers), but rather the extent of heterogeneities, created by the dimensionality of housing attributes combined with the idiosyncrasy of buyer tastes. These heterogeneities result in a large variance in the underlying matching quality and make it difficult to form a potential match. We therefore define a structural measure of market thinness as the variance in the matching quality, represented by the dispersion among buyers' valuations. A thicker market has more compressed valuations among buyers who show up to bid on a specific house; while a thinner market has more dispersed valuations.

To empirically uncover the dispersion among buyers' valuations, we set out 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. Obviously, the more housing units a buyer considers, the higher the valuation he is likely to have for the home that he chooses to visit. The value of the house to a bidder is a random variable X distributed according to a continuously differentiable distribution function $G(X; \mu, \sigma)$, with $(X - \mu)/\sigma$ following some baseline distribution G_0 that we will specify below. μ is a location parameter that shifts the distribution an equal amount for each value, and so can be seen as reflecting variations in the hedonic value, or observed quality of the home σ is the scale parameter that reflects how disperse the distribution G is. If σ is large, then the distribution will be more spread out; If σ is small then it will be more concentrated.

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 and 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

(15) $$EP = \mu + \sigma h(N) ,$$

where $h(N)$ is the expected second order statistic, $E\{X(N-1:N)\}$ from the baseline distribution G_0. As discussed above, the structural measure of the market thinness is the dispersion among N buyers' valuations. This is represented by the parameter, σ, scaled by the standard deviation of the standard distribution G_0.

To complete the model, we 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.

Case (a): 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. We assume $(Y_1, Y_2, \ldots, Y_M)$ are drawn from the same distribution F. The buyer 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. That is, $X = Max(Y_1, Y_2, \ldots, Y_M)$. 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.

M is thus the number of homes that the buyer prescreens on the MLS Internet website or the broker's office before visiting one of them, and perhaps bidding on it. Note that M is not necessarily the number of homes that are ex-ante suitable for a given buyer, but maybe a smaller number if it is too onerous for the buyer to look at all of the homes in that set.[9] In general, we expect that in markets with more sellers, there will also be more homes for the buyer to prescreen before he goes out to visit one, and hence a larger M. Note that this is different from the usual sense of the market size that fixes the ratio of buyers/sellers and views increasing the size of the market as an increase in both the number of buyers and the number of sellers. In this scenario, M captures the market size only from the sellers' side.

It is useful at this point to relate the market thinness (σ) to the market size (M). In particular, how does σ behave as M increases? As will be discussed in Section 5.1, this is a question that concerns the microfoundations for increasing returns to scale. The scenario that we described above allows us to provide some insight into this question.

[9] If the buyer ends up not buying that home, he will go back to the website and draw another subset of homes. That is, he won't go to the second best home of those he looks at before. This is analogous to a simplification of drawing with replacements --- a similar simplification is made in all random matching models.

To do so, we consider three families of the extreme value distributions for the baseline distribution G_0: Gumbel (standard extreme value distribution), Frechet, and Weibull.[10] Theses are the only possible limiting distributions of linear transformations of the maximum value X. An important property of the linear transformation is that the scale parameter, σ, needs to be chosen as a function of M. In particular, σ increases with M under the Gumbel and Weibull distributions, and decreases with M under the Frechet distributions. Thus, if the underlying distribution of a buyer's value of a house belongs to the "domain of attraction" of the Gumbel or the Weibull distributions, then we should expect the dispersion parameter, σ, to fall with the size of the market, M. In contrast, if the buyers' valuation for a home is drawn from a Frechet distribution, then we should expect the opposite result. This first case seems the most reasonable to us. However, absent any external validity for which extreme value distributions holds in our data, whether larger markets are thicker is ultimately an empirical point, which will be tested within our structural model framework.

Case (b): In the second scenario, each buyer considers, again via visual and/or textual descriptions, a single property each period, and formulates a willingness to pay, 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 one of the Generalized Pareto Distributions. A special case of this family of distributions is the uniform distribution.

The bidding process described in this section 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 (1990).

2.2.2. Econometric Specification

Equation (15) suggests the following regression

[10] Note that maxima of different distributions (F, as noted earlier) converge to different limiting distributions. For example, normal, exponential, and logistic distributions of Y converge to the Gumbel, the Frechet distribution of Y converges to Frechet, whereas bounded distributions of $-Y$ converge to the Weibull.

(16) $P_{it} = Z_{it}\,\gamma + e_t + \sigma h(N_{it}) + \varepsilon_{it}$

Equation (16) says that the sale price of a house consists of three components. The first component is $Z_{it}\,\gamma + e_t$, where Z_{it} represents observed housing attributes and location fixed effects, e_t represent period dummies. Together, these two terms capture the hedonic value of the house and proxy the location parameter μ in Equation (15).

The second component corresponds to $\sigma h(N_{it})$ in equation (15). We, of course, need to specify h. To do so, we consider two types of distributions for G_0: standard extreme value distribution (Gumbel) and uniform distribution – each corresponding to a special case in one of the two scenarios described in the conceptual framework. Under these distributions, the second order statistics from the distribution G_0 is the following:

$$h(N) = \begin{cases} 0.5772 + N ln(N-1) - (N-1) ln N, & \text{Standard extreme value} \\ 0.5\big((N-2) - (N-1)^2/(N+1)\big), & \text{Uniform distribution} \end{cases}$$

Finally, ε_{it} in Equation (16) denotes the deviations of actual house prices around their mean values.

Notice that in Equation (16), 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 γ and so improving the precision of our estimate of σ. In order to incorporate those observations, we define

(17) $$\tilde{h}(N) = \begin{cases} h(N), & N > 1 \\ 0, & N = 1 \end{cases} \qquad \mathrm{I}(N) = \begin{cases} 1, & N = 1 \\ 0, & N > 1 \end{cases}$$

and estimate the following relationship:

(18) $P_{it} = Z_{it}\,\gamma + \sigma \tilde{h}(N_{it}) + \alpha I(N_{it}) + e_t + \varepsilon_{it}$

Thus the predicted price when there is a single bidder is $Z_{it}\,\gamma + \alpha I(N_{it}) \; + e_t$ and $Z_{it}\,\gamma + \sigma\tilde{h}(N_{it}) \; + e_t$, when there are two or more.

3. Data

3.1. Bargaining and Bidding Environment

Our primary data are based on a survey we are presently conducting among recent buyers in a large North American metropolitan area. Before we present our survey, it is useful to briefly describe the home selling process in this metropolitan area. As in most other markets, house transactions in the sample market take place through a process of search and negotiation. In such a process, the seller puts a house on the market and advertizes some of the house's features and a listing price. For sellers, the listing price serves as a partial commitment. The most common restriction appears to be that the listing contract between seller and agent contains a clause obliging the seller to pay the commission in the event of a rejected offer without any restrictions but equal to or above the list price. This penalty implies that the probability of failing to complete a transaction when at least one offer equals or exceeds the listing price is quite low.

What makes this market interesting is that sellers usually indicate on the Realtor version of the MLS listing that "offers are being accepted on 'this date'", or something similar, to instruct real estate agents to bring their clients' offers at a certain date and time.[11] Offer dates are typically set for within 4-7 days of the home being listed for sale. Serious buyers would register to bid before the specified time. Insisting that all offers be made at the same time certainly encourages multiple offers or bidding wars and makes this market particularly suitable for studying a real estate auction, which is the setup that we proposed for our structural estimation.

Of course, not all home sales that are structured to receive multiple offers actually ignite bidding wars. Sometimes, there is no single offer on the specified offer date. In this case, as long as the seller's contract with the sales agent has not expired, the seller will have to leave the listing open and wait for next offers. Sometimes, only one offer arrives on the specified date. In this case, the house's sales price is typically set by negotiation between one buyer and the seller. In cases when there are multiple offers, a bidding war takes place at the specified time on the

[11] An example of such indication would be "Offers kindly reviewed on Thursday, May 13th at 7pm, please register by 5pm."

offer date. During this process, each registered bidder, accompanied by his/her agent, is arranged to sit in a private room at the office of the sales agents' brokerage firm.[12] Bidders cannot meet with each other, nor can they meet with the seller directly. Each bidder communicates his/her offer with the seller through agents. Before bidders "present" legal offers to purchase, they learn about how many bidders they are competing with from their agents, but they do not know how much competitors are offering. When all offers are presented to the seller, the seller may accept one of these in which case the bidding war ends. But it is also common for the process to have multiple rounds. Sometimes, the seller will invite all bidders to bid again. Sometimes, the seller will invite a subset of bidders. And sometimes, the bidding will go on to third or even fourth rounds. In the auction framework we proposed in Section 2.2, we model the sealed-bid aspect and the information on the number of bidders, but neglect the possibility of multiple rounds.

3.2. Survey

To conduct surveys in this large metropolitan area, we take the addresses of buyers 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 merging 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 in which they were asked the same questions as in our original questionnaire. In the third round, as yet not done, we plan to

[12] Sellers generally expect that a firm purchase agreement will be established on the specified offer date. This means that if there are multiple offers, frequently it is only the clean offers (offers without any conditions) that have a solid chance of winning. Thus, to be competitive in a multiple offer situation, bidders are advised to have a pre-purchase home inspection and mortgage pre-approval completed in advance of placing the offer. This mitigates the need for two of the most common conditions.

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 1,725 by mail and 1,169 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.

3.3. MLS Data

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 are further divided into 904 squares.

The structure variables include lot front, lot depth, the length and width of the primary room, dummy variables for basement, garage space and occupancy. We also have tax assessments for the year of, or the year prior to listing. Since taxes are a constant percentage of assessed value, taxes is a perfect proxy for assessed value. We add that variable, along with tax year dummies. Some observations lack tax information, and we drop them from the 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 exhibits a 0.5% higher mean list price, a 0.9% lower sale price, and a 0.28% higher price premium measured by the log difference between the transaction and list price. In term of overall attributes, houses in the survey response sample seem of slightly lower quality 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, bargaining 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). 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. It is easy to see that the distribution is hugely overdispersed relative to the Poisson distribution.

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 sale price of the home purchased? The second and fourth questions provide an independent source of buyer-reported price information that can be used to verify the MLS-reported price information. Table 4 shows that on average, buyers report about 0.7% higher sale 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. We plan to use that information in future work.

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 sale price and the initial offer price is smaller when there are competitors. When there are no competitors, 19 percent of the time the sale 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 sale 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 presents the histogram ofthe log difference between transaction and offer prices by presence of competing bidders. Clearly, the distribution is much tighter when there are more than one bidder. Second, Figure 3 shows that the number of days between when the offer was made and when it was accepted is tpically lower when there are more competing bidders, although the relationship is not monotonic. The overall relationship suggests to us that when the buyer reports that there were competiting bidders, there was likely to have been an auction where the seller was committed to accept the winning bid, and so there was no possibility of delay.

Table 6 presents the relationship between the number of bidders and the mean and median of price premium, defined by the difference between log of the sale price and log of the list price. Normalizing by the list price should control for quality, albeit at the cost of confounding the causal effect of bidder numbers on sales price with the causal effect of the list price on bidder numbers – which the maximum likelihood estimation is meant to solve. In any case, 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.

Table 7A presents the variance-co-variance matrix of the three observed variables (L_{it}, B_{it}, P_{it}) 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 sale 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 sale price are nearly exactly the same. Third, the covariance of the number of bidders with the sale price is an order of magnitude greater than that with the list price. That will prove critical. Table 7B shows the variance-co-variance matrix for $(L_{it}, P_{it}, L_{oit})$ in the much larger MLS data. (Recall that L_o is the previous list price, and that bidder information is only available in the interview sample.) Tables 7C and 7D present the parallel matrices for the residuals (l_{it}, b_{it}, p_{it}) and $(l_{it}, b_{it}, l_{oit})$ obtained after conditioning on attributes and taxes, periods and neighbourhoods. All three patterns continue to hold for the residuals.

We also note that the list price residual variance in the survey is 0.010, while the previous list price residual variance is 0.012. In contrast, the residual variance in the MLS population is .013, that is, thirty percent more.

4. Results

This section is divided into three subsections. The first two subsections report the OLS and ML 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.

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 more difficult to calculate ML estimates. Second, we will see that the OLS results are similar to the ML results, notwithstanding the potential endogeneity and errors in variable biases. This will justify our use of OLS estimates later in the paper.

Table 8 considers the log-linear regression of the number of bidders on the list price. Column (1) presents the bivariate regression. To recall, equation (8) shows that the estimated coefficient ought to be downwardly biased in magnitude given the noise in the list price contributed by unobserved quality. At .085, the coefficient is positive and significant, which is hard to interpret. Adding taxes (essentially assessed value) in Column (2) only increases the coefficient, and substantially so, 0.38. Adding period dummies instead (Column (3) has little effect on the estimate. Adding home attributes to taxes (Column (4)) brings the coefficient down somewhat, but it is still unexpectedly large and positive, at 0.25, and this is little affected by adding period dummies (Column (5)). Only when square dummies are included in the last column, i.e., we include fine location information, does the coefficient turn negative. The sign is as expected, but the coefficient, -.055, is imprecise with a standard error of 0.060. We would have thought that taxes would have controlled for the location, but evidently it does not wholly do so.

Table 9 presents the log-linear regressions of the sale price on the number of bidders. The first column shows the bivariate regression. As shown in equation (9), this should yield consistent estimates only if m, the causal effect of bidder numbers on price (combined with the selection effect), 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 (not shown). 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 sale 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 R-squared, but changes the coefficient on the number of bidders only slightly from 0.037 to 0.034, and leaves the precision unchanged.

Finally, in Columns (5) and (6), we move the list price to the right hand side. Since the estimated coefficient on list prices is very close to one, it is not surprising that coefficients on the number of bidders remain the same as those in Columns (3) and (4).). These results also support simply analyzing the premium directly.

If we compare across the specifications in Table 9 that condition on attributes and time, we find that although controlling for the list price, in either restricted or unrestricted form, increases the estimated coefficient by about a half, the size of the standard error in Column (2) is such that differences could pass the Hausman-Wu test. This strengthens our general approach that the main advantage of controlling for the list price lies in the increased precision of the estimates.

4.2. Maximum Likelihood Estimates

The OLS regressions indicate that by conditioning on the list price, we can improve the precision in estimating the effect of the number of bidders on the sale price. The earlier discussion of the model, however, shows that by doing so we will be aggrevating the bias if m is closer to zero than to one. To address this concern, we estimate the simultaneous equation system defined in (4) – (7) with the maximum likelihood method.

4.2.1. Results

The results are reported in Table 10, along with bootstrap standard errors. The bootstrap includes the initial regression on attributes, location and time effects, i.e. equations (1)-(3), that generates the residuals for use in the estimation procedure outlined above. In this draft we present only our baseline model given by equations (4)-(7), that assume that unobserved quality has the no effect on the number of bidders, and the same effect on the list price as on the sale price.

The autocorrelation parameter r is estimated at 0.542. When that is multiplied by the variance of the (residual) previous list price, 0.011, we get an estimate for the variance of the unobserved quality of list price of 0.007. 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 -0.059, 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 rewrite equation (11) as

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

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. The estimated $\hat{\sigma}^2_w$ is 0.230 with a standard error of 0.012. 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 ψ, the key estimate of our interest, ϕ, is extremely precise, with an estimate of 0.034 and a standard error of 0.004. Recall that ϕ is interpreted as the reduced-form measure of market thinness. We will discuss its implications in the next subsection.

We estimate m at 1.015, with a standard error of 0.403. From (13) 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. Our estimate is not so dissimilar from implicit estimates of it that can be derived from previous papers in which both the list and sale 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 sale price on the excess of loan-to-value over 80 percent, and obtain coefficients of 0.19 and 0.16 respectively, implying an estimate of m of 0.19/0.16= 1.19. Genesove and Mayer (2001), who also consider loss aversion, find that loan-to-value affects list price and sale 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.

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 (13) 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. [13]

4.2.2. Implications

A striking result from Table 10 is that the maximum likelihood estimate of the effect of the number of bidders on the sale price is very close to the corresponding OLS estimates reported in Table 9 that condition on list price, or substitute the price premium for the sales price. Recall

[13] We can get a reasonable value for this bias 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. We have yet to do this.

that conditioning thus is problematic, since it assumes that the list price strategy and the unobserved quality affect the sales price in an identical fashion, i.e., $m = 1$. Since the ML estimate of m is close to 1, evidently in practice this is not a problem. As shown in equation (11), as long as m is not far from 1, then one should expect a near 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 sale price. It follows from equations (5) and (6) that

$$\frac{\partial p}{\partial \eta} = \hat{\psi}\,\hat{\phi} + \hat{m}\,. \tag{15}$$

Equation (15) makes clear that the effect of an increase in list strategy (η) on the sale 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 $\hat{\psi}$ and $\hat{\phi}$ are sizeable individually, the product of the two estimates amounts to -0.002 with a standard error of 0.092, which is clearly dominated by the second term $\hat{m}$, which is about 1.015 with a standard error of 0.403. 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. So if selection is not so important, we can conclude that signaling dominates induced bidding competition: conditional on quality, a higher list price will generate a higher price. Obviously, the higher price will be at the cost of a lower likelihood of sale.

4.3. Structural Estimates

So far we have been relying on a reduced-form measure of the market thinness, namely, how much the sale 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 OLS estimation in what follows.

In Table 11, we report estimates from several variants of equation (18). Columns (1) and (5) report the benchmark case where the sale price is regressed on a set standard controls, including housing attributes, tax assessments, neighborhood and year effects. They differ only in that column (5) includes the list price. This is a typical hedonic house price model. In the absence of unobserved quality and search frictions, the model would provide reliable estimates of home value.

In the remaining columns, we account for search frictions by including a dummy for a single bidder, and $\tilde{h}(N)$ as defined in equation (17). The parameter of interest is σ, the coefficient on $\tilde{h}(N)$, which reflects the dispersion of a given bidder's value distribution. The reasonableness of this approach is assured by the closeness of the OLS (conditional on the list price) and ML estimates of the bidder numbers effect. Nonetheless, for purposes for comparison, we show regressions both conditional and unconditional on the list price. Bidders' valuations are assumed to be drawn from a standard extreme value distribution in Columns (2) and (6) and from a uniform distribution in Columns (3) and (7). In both specifications, the estimates of σ are positive and statistically significant, providing evidence for the dependence of the price premium on the number of bidders. Conditioning on the list price increases σ by about a quarter and more than triples the precision. As in our earlier regressions, we cannot reject a coefficient of one on the list price.

To impute the standard deviation of the distribution of valuations among serious bidders, 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. For the standard extreme value distribution, Column (2) shows that the estimate of ϕ is 0.038, implying a standard deviation of 0.049, and so an interquartile difference of 7.71 percentage points (.049*(ln(-ln(.25))-ln(-ln(.75)))). In term of the magnitude, the estimate of ϕ suggests that the standard deviation in the distribution of bidders' valuation of an average house in the market is about $21,100 (0.049 x $430,617), which is quite substantial. For the uniform distribution, Column (3) shows that the estimate of ϕ is 0.145, which implies a standard deviation of 0.042 and of course an interquartile difference of 2.1 percentage points. In both cases, we find strong evidence that there is consideration amount of variation among buyers' valuation of the same house, providing direct evidence for market thinness.

In Columns (4) and (8), we take a non-parametric method and approximate the second order statistics by a set of dummy variables for the number of bidders. Considering first Column (4), where we do not conditional on the list price, we see an overall pattern of increasing coefficients, but the relationship is not monotonic (although we have yet to conduct a statistical test). In Column (8), however, where we do condition on list price, the estimated coefficient falls in only one case, and one can be very confident that a test of weak monotonicity would not fail. Adding a second bidder increases price by 1.1 percent; adding a third increases price by a further 0.7 percent, and a fourth by a further 0.8 percent. We do not see evidence, however, of a concave relationship between the bidder number and its estimated effect (adding a firth bidder increases price by a further 4.5 percent!), a property shared by the expected second order extreme value from bot then uniform and extreme value distribution, and, indeed, either distribution can be rejected at less than the 1% level when conditioning on the list price. One can reject neither distribution at a reasonable level of significance when not conditioning on the list price, indicating again the importance of conditioning on the list price.

5. Implications From Market Thinness Estimates

Overall, the evidence in the previous section is consistent with the conventional wisdom that real estate markets are thin. But why do we care how thin real estate markets are? The reason is that without understanding the magnitude of the market thinness, the policy recommendations and the testable economic hypotheses are limited. Which markets do we expect to be thinner than others? Why are summer markets always "hotter" than winter markets? Is there a loss from uncoordinated matching between buyers and sellers? How much can we gain from improvements in such process? The answers to these questions would not be very clear if we only have reduced-form evidence for market thinness. In contrast, the structural approach taken in this paper, by providing microfoundations, makes predictions about where we should expects markets to be thinner, how much buyers gain from winning a bidding war, how much policy makers can gain from improving the home matching process. In this section, we explore these implications from our market thinness estimates.

5.1. Increasing Returns to Scale

Despite much literature indicating that increasing returns to scale may be important in housing markets, little work has investigated the underlying economic rationale. Quite often increasing returns to scale in matching markets are mechanically built into a matching function or modeled by a first order stochastic shift of the match quality distribution. However it is not clear why the matching function necessarily exhibits the increasing returns or why the match quality distribution should shift up when the market is bigger. The structural model presented above allows us to link the variance in the match quality to the number of sellers, hence providing a microfoundation for the increasing returns to scale.

The basic idea is simple. We return to case (a) from before. Buyers prescreen M homes before they know which one they will visit and perhaps bid on. The more homes they have prescreened, the more likely they will put a high value on the home that they decide to visit and bid on. This is true for all buyers who show up and bid on a certain house. Furthermore, when buyers' valuations follow distributions that converge to Gumbel or Weibull, the valuations of those "serious" buyers are going to be not only higher, but also clustered. As a result, there is less variance in the underlying match quality, making the market thicker. This provides a justification for the presence of the increasing returns to scale for the housing market.

Is the microfoundation proposed above plausible? Unfortunately, there is no direct evidence for whether the actual valuations among buyers indeed converge to the Gumbel or Weibull distributions. Nevertheless our structural model allows us to seek empirical support for the prescreen story above. To see this, note that our argument immediately implies that markets with a large number of suitable homes to prescreen have lower dispersion among buyers' valuation. We do not directly observe the number of homes that each buyer prescreens before going out to visit. Therefore, we devise several alternative measures of market size and estimate variants of our structural model that allows the dispersion parameter to vary with measured market size. Overall, these analyses provide evidence that larger markets tend to be thicker, lending support to the increasing returns to scale in housing markets.

(1a): We first investigate whether the summer markets are thicker than winter markets. We denote the period between April 1 and September 30 as the summer time, and the remaining period as the winter time. Summer months are clearly more active than winter months in this market, by a factor of two or three, whether one looks at listings or transactions. In general, summer months are thought to be more active as families with children want to get settled down in their new residence before the start of the new school year. Another possible reason is the weather. Although Ngai and Tenreyro (2010) reject weather-based explanations for seasonal differences in transactions in their work, this market might be different. Summer temperatures normally range from 15C (60F) to 25C (80F), while winter temperature hovers below freezing and a snowfall of less than 10cm is not unusual. Thus, for potential buyers, the costs of going out to visit a house should be much higher during winter months than during summer months. Whatever the source of the differential activity across seasons, if larger markets are indeed thicker markets, then we should expect buyers' valuations be less disperse in the summer.

To test this hypothesis, we estimate a version of equation (18) that controls for the list price and allows both $\tilde{h}\ (N)$ and one-bidder dummy to interact with a summer dummy variable Q_t.

(19) $$P_{it} = Z_{it}\,\gamma + (\sigma+\sigma_S Q_t)\tilde{h}(N_{it}) + (\alpha + \alpha_S Q_t)I(N_{it}) + e_t + \varepsilon_{it}$$

where Q_t is one for a summer month, and zero otherwise. The main effect of this variable is absorbed by month dummies.

Table 12 presents estimates from this equation. The underlying distribution is assumed to be standard extreme value in Column (1) and uniform in Column (2). We exclude observations with more than 10 bidders from our sample so that the results not be influenced by extreme, and suspect, reported outcomes. Consistent with what we expected, the coefficients on the interaction between summer and $\tilde{h}\ (N)$ are negative in both specifications, indicating less dispersed valuations among buyers in the summers, specifically a reduction of 18 percent (.008/.043) in the measure of market thinness in the summer.

Figure 4 plots the thin market component of the predicted price against the number of bidders by seasons, i.e. $summer\ predicted\ price\ (N_{it}) = (\hat{\sigma}+\hat{\sigma}_S)\tilde{h}(N_{it}) + (\hat{\alpha} + \hat{\alpha}_S)I(N_{it})$, and $winter\ predicted\ price\ (N_{it}) = (\hat{\sigma})\tilde{h}(N_{it}) + (\hat{\alpha})I(N_{it})$. The left panel is based on estimates from the standard extreme value distribution, while the right panel is based on estimates from the uniform distribution. We see that in both seasons, as the number of bidders increases, the predicted price increases. The result is consistent with the finding of a positive relationship between the number of bidders and the sale price from the reduced-form model. Second, when there are multiple bidders ($N_{it} \geq 2$), price increases more with an additional bidder in the winter than in the summer, indicating a thinner market in the winter time.

The estimates are statistically insignificant though, so that one cannot reject the absence of seasonality in market thinness.

(1b): A more direct proxy for the market size (*M*) in the conceptual framework is the number of listings that are available for potential homebuyers to choose among. We therefore proceed by looking at the number of listings in the district where the buyer bought the house during the month of transaction and the number of listings in the entire metropolitan during that month. The listing data are assembled by the local real estate board and are publicly available in its monthly market reports. In practice, a buyer may look at homes across several districts and/or over a few months. However, we have no measure of the number of listings that each buyer looks at before visiting home, nor do we know how widely they search . While measurement error may be an issue, having the number of listings from the lowest searchable geographical range (a district) to the largest searchable geographical range (entire metropolitan area) should give us a reasonable range for the estimated relationship between the market size and the market thickness. Using $listing$ to denote the log of number of listings, we estimate the following model:

(20) $P_{it} = Z_{it}\,\gamma + (\sigma+\sigma_S listing)\tilde{h}(N_{it}) + (\alpha + \alpha_S listing)I(N_{it} = 1) + \rho\ listing + e_t + \varepsilon_{it}$

In Columns (3)-(4) of Table 12, we present estimates when listings are measured at the district level. The estimated coefficients on the interaction between listing and $\tilde{h}\ (N)$ are negative and statistically significant, indicating that market thickness increases with the number of listings. Taking the estimates in Column (3) as an example, if the number of listings increases by 3, corresponding to the difference between summer and winter markets, then the underlying dispersion parameter decreases from 0.055 to 0.048 – about 12% change in market thinness. Of course, some buyers may search for houses in several districts. This should lead to an error-in-variable problem and will bias our estimated change downwards in magnitude.

In Columns (5)-(6) of Table 12, we present estimates when the number of listings is measured for the entire metropolitan area. The signs of the estimates are again consistent with what expected, although the statistical significance is much reduced, as we would expect if buyers search more narrowly than the whole market.

5.2. Money Left on the Table

The estimates in Table 11 can also be used to compute how much money is left on the table. The auction setting described in Section 2.2 predicts that, in the event of a bidding war, 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 distributions and the number of bidders. The results are reported in Table 13. When there are two bidders, we approximate the buyer surplus by $0.038(E[X(2:2)] - E[X(1:2)])$, which is equivalent to 5.3% if bidders' value is drawn from the standard extreme value distribution; and 4.8% if it is drawn from the uniform distribution. These estimates roughly amount to 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 observations with bidding wars. As the number of bidders increases, the predicted amount of money left on the table under two distributions diverges. 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. Of course this pattern is entirely driven by the two functional forms, with only the scale of the divergence determined by our estimation.

Is it possible to use our structural estimates to infer the expected buyer's surplus when there is only one bidder? Note that, in the absence of an auction setting, the single buyer's surplus cannot be defined in a way as above. Instead, we define the single buyer's surplus as

$E[X] - P(1)$, where $E[X]$ is the expected value of a random draw from the distribution of serious bidders, and $P(N)$ is the price when there are N bidders. This expression can be written as $E[X] - P(1) = (E[X(1{:}2)] + (E[X] - E[X(1{:}2)]) - P(1) = \left(P(2) - P(1)\right) +$ $(E[X] - P(2))$, using the auction theory result that P(2) = E[X(1:2)]. From Equation (12), it follows that $(E[X] - P(2))$ = $(\mu + \sigma \times mean) - \left(\mu + \sigma \times h(2)\right) = \sigma \times (mean - h(2))$, where $mean$ denotes the mean of the baseline distribution G_0.Thus, $E[X] - P(1) = \left(P(2) - P(1)\right) +$ $\left(\sigma \times (mean - h(2))\right) = (\sigma \times h(2) - d) + \sigma \times \left(mean - h(2)\right) = -d + \sigma \times mean$, where d denotes the coefficient on the single buyer dummy from our structural model. Using the estimates from Table 11, we therefore obtain a single buyer's surplus of 0.037 under the standard extreme value distribution, and 0.034 under the uniform distribution. These estimates are reported in Table 13.

As in the two-bidder case, the standard extreme value and uniform distributions predict similar amount of money left on the table when there are is only a single bidder, which accounts for about two-thirds of the total observations.

5.3. Mismatching Costs

[To be completed]

6. Conclusion

This paper makes two contributions. Substantively, it provides the first direct empirical estimates for the thinness of real estate markets. Although a large body of housing literature on market efficiency and liquidity builds on the assumption that real estate markets are thin, there is few empirical work that examines how thin real estate markets are. We tackle this important question by conducting a survey among recent home buyers and by developing an econometric framework to estimate the market thinness. The reduced-form estimates show that doubling the number of bidders increases the sale price by 2.4 percent, on average. The structural estimates further reveal that the standard deviations of the distribution of the valuation of serious bidders (i.e., those who show up) range from 4.2% (uniform distribution) to 4.9% of the sale price (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.

These estimates also have important implications. First, the thinness estimates explain quite naturally the presence of increasing returns to scale in housing markets. Estimated valuation dispersion among buyers is bigger in winters and in markets with fewer available listings, lending support to the hypothesis that bigger markets are thicker. Furthermore, our estimates of market thinness allow us to recover the buyer's surplus from a completed transaction. Our results allow us to calculate the reduction in the winning bidder's expected surplus when the bidding is well attended. Finally, we demonstrate (in a future version) that the inefficiency from uncoordinated matching between buyers and sellers can be computed as a function of the market thinness estimates. The imputed loss can serve as an upper bound for the potential gains from policies that aim to improve the efficiency in the matching process.

Another contribution of the paper is methodological. We propose a simultaneous estimation approach that estimates house prices controlling for the presence of unobserved house quality and to the bidders' endogenous response to list strategy. The resulting model allows us to estimate how price increases with the number of bidders and how bidders respond to the seller's list strategy. We are unable to estimate the effect of list prices on the number of bidders with any reasonable degree of precision. We find that conditional on the list price suffices to estimate the effect of the number of bidders on the transaction price with high precision.

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.

Hosios, Arthur. “On the Efficiency of Matching and Related Models of Search and Unemployment”, *Review of Economic Studies*, Vol. 57,pp. 279-298.

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.

Ngai, Rachael and Silvana Tenreyro (2009), “Hot and Cold Seasons in the Housing Market”, LSE wp, 2010.

Peters, Michael, “Ex ante Price Offers in Matching Games Non-Steady States.” *Econometrica*, 1991, 59 (5) pp 1425-1454.

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.

Wolinsky, Asher “Dynamic Markets with Competitive Bidding”, *Review of Economic Studies*, Vol. LV, p. 71-84.

Wheaton, William C., “Vacancy, Search and Prices in a Housing Market Matching Model”, *Journal of Political Economy*, Vol. 2(6), p. 1270-92.

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)
Sale 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
Sale 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: Sale 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 6: 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 7A: Full Covariance Matrix (Survey Sample)

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

Table 7B: Full Covariance Matrix (MLS Sample)

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

(lnBIDDER row is missing here)

Table 7C: Residual Covariance Matrix (MLS Sample)

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

Table 7D: Residual Covariance Matrix (Survey Sample)

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

Table 7E: 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 Sale price residual	.0141	.0160	
Ln Number of Bidders	- .0002	.0077	.2304

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)	0.0904 (0.0336)	0.2471 (0.0696)	0.2313 (0.0694)	-0.0538 (0.1307)
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	0.0029	0.0216	0.0518	0.0723	0.1220	0.3537
# observations	2184	2184	2184	2155	2155	1952

Table 9: OLS Price Regression

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Sale Price		Sale - List		Sale 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

$\hat{\psi}$	-0.059
	(3.290)
$\hat{\phi}$	0.034
	(0.004)
$\hat{m}$	1.015
	(0.403)
$\hat{\sigma}_v^2$	0.007
	(0.0003)
$\hat{\sigma}_\eta^2$	0.003
	(0.0005)
$\hat{\sigma}_w^2$	0.230
	(0.012)
$\hat{\sigma}_u^2$	0.001
	(0.0001)
r	0.542
	(0.011)

Bootstrapped Standard Errors appear in parentheses beneath the estimate coefficients.

Table 11: Structural Model Estimation

Dep. Variable	Sale Price							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Underlying Distribution		Standard Extreme	Uniform			Standard Extreme	Uniform	Non-Parametric
$\tilde{h}\ (N)$		0.038 (0.010)	0.145 (0.037)			0.049 (0.003)	0.175 (0.011)	
I(N= 1)		-0.015 (0.008)	-0.034 (0.007)			-0.008 (0.003)	-0.032 (0.002)	
N= 2				0.017 (0.010)				0.011 (0.003)
N=3				0.008 (0.013)				0.018 (0.004)
N=4				0.042 (0.016)				0.026 (0.005)
N=5				0.084 (0.023)				0.071 (0.007)
N=6				0.054 (0.031)				0.069 (0.009)
N=7				0.093 (0.043)				0.070 (0.013)
N=8				0.130 (0.055)				0.103 (0.017)
$N \geq 9$				0.100 (0.029)				0.145 (0.009)
List Price					1.004 (0.01)	1.008 (0.009)	1.006 (0.009)	1.008 (0.009)
Tax & Tax year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Attributes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Square dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.9119	0.9139	0.9139	0.9144	0.9898	0.9922	0.9920	0.9925
# observations	1929	1929	1929	1929	1929	1929	1929	1929

Table 12: Market Thinness vs. Market Size

Dep. Variable	Sale Price					
	(1)	(2)	(3)	(4)	(5)	(6)
Underlying Distribution	Standard Extreme	Uniform	Standard Extreme	Uniform	Standard extreme	Uniform
$\tilde{h}\ (N)$	0.043 (0.005)	0.152 (0.018)	0.055 (0.007)	0.202 (0.027)	0.044 (0.008)	0.159 (0.028)
I(N= 1)	-0.017 (0.004)	-0.038 (0.003)	-0.011 (0.005)	-0.038 (0.005)	-0.015 (0.006)	-0.036 (0.006)
$\tilde{h}\ (N)\ \times$ Summer	-0.008 (0.006)	-0.028 (0.024)				
I(N= 1) × Summer	0.011 (0.004)	0.015 (0.004)				
$\tilde{h}\ (N)\ \times$ listing			-0.006 (0.002)	-0.021 (0.008)	-0.001 (0.002)	-0.0037 (0.008)
I(N= 1) × listing			0.0012 (0.0014)	0.0039 (0.0016)	0.002 (0.002)	0.0024 (0.0017)
listing			-0.0012 (0.0015)	-0.0040 (0.0014)	-0.00006 (0.0026)	-0.0005 (0.0025)
List Price	1.012 (0.008)	1.012 (0.008)	1.009 (0.009)	1.009 (0.009	1.013 (0.010)	1.012 (0.010)
Tax & Tax year	Yes	Yes	Yes	Yes	Yes	Yes
Attributes	Yes	Yes	Yes	Yes	Yes	Yes
Period dummies	Yes	Yes	Yes	Yes	Yes	Yes
Square dummies	Yes	Yes	Yes	Yes	Yes	Yes
# observations	1911	1911	1672	1672	1604	1604

Table 13: Money Left on the Table: E[X(N:N)]-E[X(N-1)]

Number of Bidders	Standard Extreme Value (σ =0.049)	Standard Uniform (σ =0.175)
1	0.03628	0.032
2	0.06811	0.05828
3	0.05978	0.04375
4	0.05635	0.035
5	0.05488	0.02923
6	0.05341	0.02503
7	0.05292	0.02188
8	0.05243	0.01943
9	0.05194	0.0175
10	0.05165	0.01593

Note: When there is one bidder, buyer surplus is imputed as $-d + \sigma \times mean$, where d is the coefficient on the single buyer dummy from Table 11; σ is the coefficient on $\tilde{h}(N)$ from Table 11; *mean* is 0 under the standard uniform distribution; and 0.5772 under the standard extreme value distribution. When there are multiple bidders, buyer surplus is imputed as $\sigma \times (E[X(N{:}N)] - E[X(N-1){:}N])$.

Figure 1: Histogram of Number of Bidders

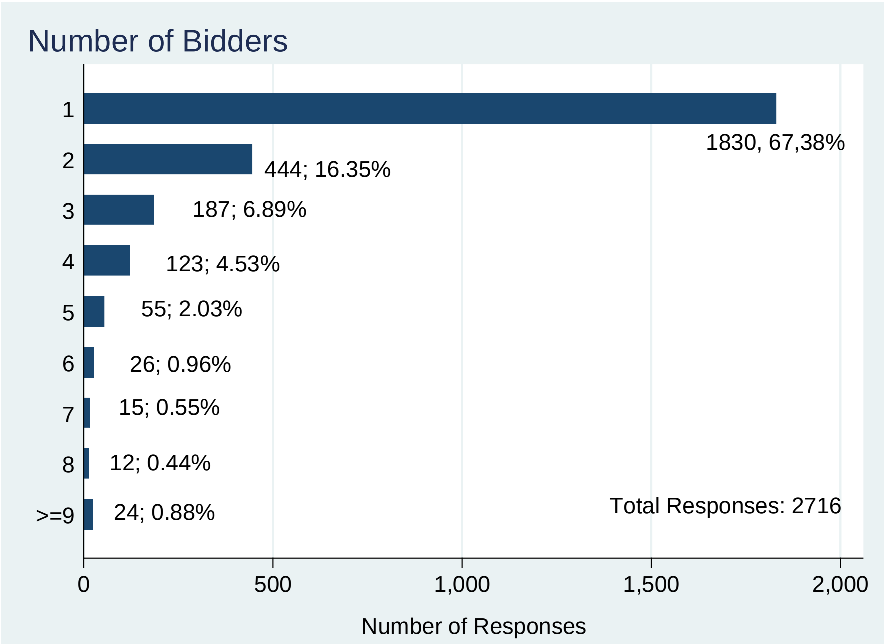

Figure 2: Sale price -List Price, By Bidders

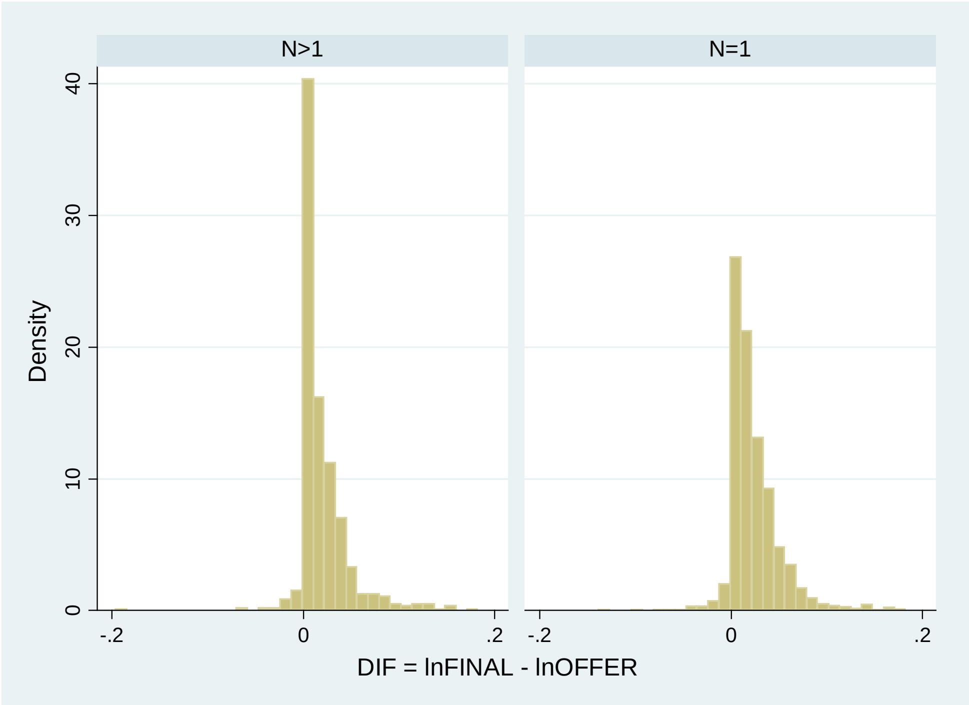

Figure 3: Average Days Until Offer Accepted, By Bidders

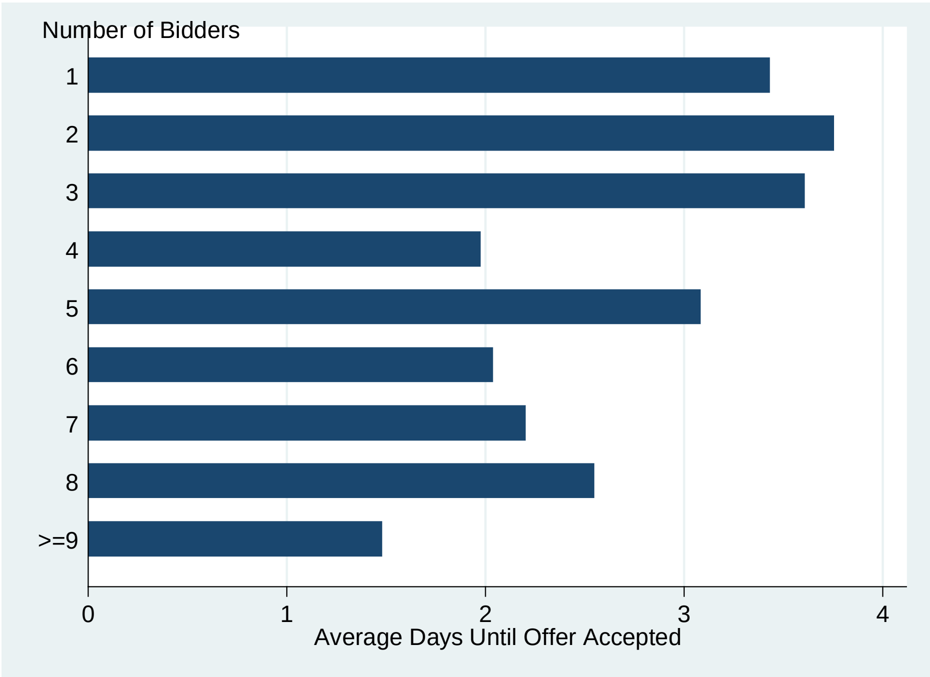

Figure 4: Seasonality Effect in Market Thinness