The Effect of Agent Inventory Holdings on Residential Real Estate Transactions

Jason Beck[1], David Hoover[2], Michael Toma[1]

[1]Department of Economics, Armstrong State Unviersity, Savannah, Georgia, United States
[2]Georgia Ports Authority, Savannah, Georgia, United States
Email: michael.toma@armstrong.edu

Abstract. This study uses an OLS hedonic pricing model to compare 94 months of Multiple Listing Service (MLS) listings in Chatham County, Georgia with heterogeneous housing characteristics. Controlling for other factors, this study estimates the effect of listing real estate agents' inventory holdings (number of open listings at time of sale) on the sale price of a house sold via the MLS. Results indicate that agents holding greater levels of inventory result in up to a nearly 6% lower sale price relative to agents with only a few listings. This finding could be considered consistent with the hypothesis of a potential principal-agent problem in the real estate brokerage market whereby agents encourage their clients to accept sub-optimal prices in order to increase total commission income.

Keywords: Real estate, realtors, housing prices, hedonic real estate model.

1 Introduction

The market for real estate brokerage is an environment conducive to the principal-agent problem. In short, the principal-agent problem is a market failure that occurs when one individual, the agent, uses his or her information advantage against the principal. Given that the two individuals have different incentive structures, this can lead to sub-optimal results for the "principal."

In the case of real estate brokerage, sellers of homes face an informational disadvantage in that they have imperfect information on the quality of their agent and the fair market price for their home. It has been suggested that agents are able to influence the decisions of their clients to accept a low offer on a home because the agent has an asymmetrical knowledge advantage about the real estate market as compared to the client (Levitt and Syverson [1], Anglin [2]) and uses that advantage in an attempt to increase volume.

When a house is sold through a Multiple Listing Service (MLS), one agent acts as the "listing agent" while another (sometimes the same agent) acts as the "selling agent." The listing agent acts on behalf of the owner of the property and reports various characteristics to the MLS database. Other agents, acting as selling agents, can then search this database for properties meeting their clients' needs.

This paper will focus on the incentive structure surrounding listing agents. Listing agents work on a commission based system. Conventional wisdom is that agent commissions are a fixed percentage (typically five or six percent) of the sale price of the house when it sells, which is split between the listing and selling agent. By increasing the volume of sales in a given time period, the increase in commissions could outweigh losses from lower sale prices in each transaction. In an agent's attempt to sell more homes within any given time period, it is possible that his or her clients are receiving lower sale prices by being convinced that no better deal is forthcoming or through encouragement to reduce the price of the home when it is not optimal for the homeowner to do so. The number of clients taken on by a listing agent at any point in time can be thought of as an inventory of listed properties which the agent is holding. Effects of this inventory level on the sale price of heterogeneous housing can be estimated using a hedonic pricing model.

2   Hedonic Real Estate Models

The sale price of a house is often modeled using hedonic regression models. Hedonic models assume a house's transaction price is a function of the value placed on its component parts, such as square footage and locational characteristics. Hedonic models are beneficial in that by modeling a house as the sum of its components, it is possible to compare individual heterogeneous houses. Additionally, hedonic models provide an estimated marginal contribution of the housing components themselves, such as the value of each additional bathroom (Beck, Fralick and Toma [3], Dubin [4], Sirmans et al. [5]).

Though useful, hedonic models have limitations. Locational issues may arise in which buyers place a different marginal value on housing components in different areas. For instance, newer houses may be more valuable because they have desirable modern features more closely reflecting current taste or more modern construction methods. While controlling for such issues may be possible, considering large geographic areas can exacerbate such biases. Heterogeneous preferences among homebuyers may also result in biases to estimated means when examining large areas (Sirmans et al. [5]). For instance, consumer behavior on the west coast of the United States may generate different marginal valuation of housing components than consumer behavior on the east coast of the United States.

There is a wide variety of model specifications in the hedonic pricing literature, but the models do share some common characteristics. Sirmans et al. [5] provide an excellent overview of the history of hedonic real estate models, results, and the most common control variables found in the literature.

Hedonic real estate models are commonly specified as log-linear models. The natural log of sales price is used as the dependent variable which helps to account for differences between valuations of a house's component parts when large differences in sale price occur (Sirmans et al. [5]).

The number of days a house spends listed for sale may help explain sale price. By being listed for longer, the market has more opportunities to match it with a buyer who will pay more. However, the longer a house remains on the market unsold, the more likely the seller is to reduce the list price in order to entice a buyer's offer. Thus, there is a simultaneity effect present when including some form of a control variable for time on market. This effect has been addressed in the literature in multiple ways. Most studies simply exclude the variable due to its correlation with sale price (Hamilton and Schwann [6]; Beck, Fralick and Toma [3], Sirmans et al. [5]). Others include listing time, and typically find a negative effect on transaction price (Sirmans et al. [5]). However, Sirmans et al. [5] note several studies with conflicting results. Judd, Seaks, and Winkler [7] find a positive relationship between days on market and sale price. Haag, Rutherford, and Thomson [8] also find a negative relationship. The use of two-stage least squares regression is not uncommon to handle this simultaneity issue, however Sirmans et al. [5] note that such models are difficult to properly specify. Forgey, Rutherford, and Springer [9] and Rutherford, Springer, and Yavas [10] utilize two-stage least squares methodologies and find a positive relationship between sale price and days on market, as well as time on market's dependence upon search effort and various other market factors.

In their review of 125 hedonic real estate pricing models, Sirmans et al. [5] found that the age of the house was the most often controlled-for variable. In order of frequency of appearance, the remainder of the ten most popular control variables was square footage, garage spaces, existence of a fireplace, lot size, number of full bathrooms, air conditioning, existence of a pool, and existence of a basement.

3   Hypothesis and Model

A hedonic pricing model is used to analyze the effects of real estate agent inventory holdings on real sale price of residential properties. It is hypothesized that listing agents may choose to trade some degree of sale price in return for a higher rate of turnover. Given the marginal increase in agent income from an increase in home sale price is small, an agent might more than make up lost commissions via lower sale prices on any given house with an increased number of sales.[1] Listing agents can take advantage of such

---

[1]For example, if through higher effort an agent increases sale price by $10,000, total commission at 6% would increase $600. That $600 is first split between the listing agent and selling agent resulting in an increase of only $300 for the listing agent. This amount may be split again by the brokerage house resulting in only $150 for the listing agent. An effective 1.5% commission rate on a $200,000 sale is $3,000. Therefore, it may be more efficient for

a personally beneficial exchange through their asymmetrical knowledge of current real estate markets to convince sellers to lower their asking price or accept a price which is sub-optimal for the seller, but maximizes commissions for the agent.

Listing agents' inventories are measured by recording the number of other listings currently active for a given agent on the day of the sale of a property, including the sold property itself. These form the "Agent Inventory" variables. The value of this metric is further illustrated in Figure 1.
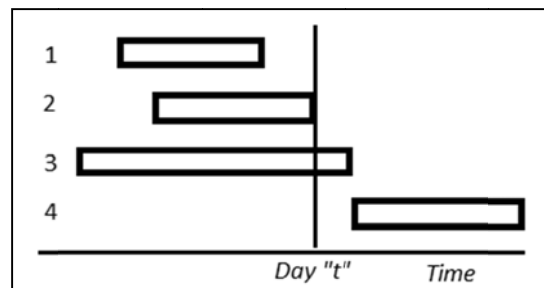
Figure 1. Each rectangle represents a distinct listing with the left edge being the list date and the right edge being the day of removal from the MLS. When a listing sells, a count of all other active listings is made and recorded as that observation's "inventory" level. For instance, listing #2 in this graphic has an inventory of 2 because listing #3 is also held in inventory at the time of its sale on day "t".

It is hypothesized that while holding the number of days each property takes to sell constant, greater inventories reflect greater volume during the pertinent period. Greater volume may reflect agent quality as agents who are able to increase volume without negatively affecting their clients do so. If this is the case, then inventory should have a statistically insignificant effect on sale price. Alternately, agents may be taking on increased volume at the expense of sale price in order to increase turnover. If this is the case, increased inventory should correlate with a reduction in sale price as agents neglect clients' interests in order to maximize their own commissions.

Only observations from agents who listed 25 total properties over the entire 95 months in Chatham County, Georgia are analyzed in order to eliminate those properties handled by agents well outside the mainstream of full-time employment in the industry.

Given a correlation between number of bedrooms, bathrooms and square footage as well as the wide variety of property sizes in the dataset, a bedrooms/bathrooms ratio is included, in addition to the number of full bathrooms, as a control for effects from the existence of these rooms (and a desirable balance between the two) that are not observed by simply controlling for square footage.

Binary variables are used in the model to control for other features such as the existence of high/vaulted ceilings, laundry rooms, porches, as well as dummy variables which track the number of fireplaces and garage spaces of each observation property. It is expected the existence of each of these amenities will be positively related with sale price of the home.

Certain construction and exterior improvement characteristics are also controlled for in the model. The number of floors/stories is controlled for as a category of dummy variables relative to single-story homes. Brick, stone, frame, and log houses are controlled for by the model using dummy variables. In addition, cedar and vinyl siding types are controlled for using dummy variables. Type of sewer system and exterior attributes such as a pool are also included as categories of dummy variables.

Finally, locational features of the house are also used as predictor variables. The size of the property on which the house is built (lot size) is controlled for by fixed effects. Since Chatham County is located adjacent to the Savannah River and the Atlantic Ocean, whether or not the house is located on the waterfront is used as a control variable.

Most importantly, the six-digit zip code of each property is controlled for using 128 distinct fixed effects. Six-digit codes were used to balance the need for precision beyond the standard five-digit code

---

the agent to accept a $200,000 offer and move on to the next sale than to expend extra effort for a marginal gain of $150.

and the burden of an exponentially growing number of degrees of freedom necessary to control for additional digits and greater geographic disaggregation of the data.

Many other useful locational details which are difficult to specify explicitly in the model such as tax status, school district, crime, neighborhood desirability, nearby transportation infrastructure (Center for Neighborhood Technology [11]), etc. are controlled for as proxies using the six-digit zip code fixed effects (Beck, Fralick and Toma [3]),

The primary model takes the form:

$$Ln\big(price\big) \; = \; B0 \; + \; B1\big(inventory9\_26\big) \; + \; B2\big(inventory27\_43\big) \; + \; B3\big(inventory44+\big) \; + \; \Omega\big(Y\big) \; + \; \big(Z\big) \; + \; e \qquad (1)$$

where Y is a vector of structural characteristics and Z is a vector of locational and time of sale controls.

Control variables were chosen, as much as possible, in accordance with the extensive literature reviewed by Sirmans et al. [5] in his compilation of trends in hedonic real estate model specifications. One aspect of hedonic OLS modeling strategies of considerable difficulty is controlling for simultaneity between price and time on market. Many studies simply exclude this endogenous variable (Sirmans et al. [5].) Including a control variable for days on market has no notable effect on results, and thus it is excluded from the model.

4   Data Sources

The data analyzed were obtained from the sav.mlxchange.com MLS system covering Chatham County, Georgia. Chatham County contains the city of Savannah as well as the neighboring cities of Pooler, Garden City, Tybee Island, and Port Wentworth. Chatham County is the fifth most-populated county in Georgia with an estimated 278,434 total residents as of 2013 (Census 2013 one-year estimate).

All valid[2] and complete entries sold for at least $50,000 in real 2012 dollars in Chatham County between January 1, 2007 and November 2014 by an agent who listed 25 or more MLS listings within that period were included in the dataset. $50,000 was chosen as a lower bound in accordance with Levitt and Syverson [1] in order to eliminate sales which are clearly unusual. In addition, a relatively small number of realty firms appear to enter their firm's owner/broker in lieu of individual agent names into the MLS system. This type of entry leads to a fringe of inventory outliers. Using a web search, several of these very-high inventory listings' realty firms were discovered to be owned or brokered by the owner of the firm. The names of the other realtors with the firm did not appear in the MLS dataset. Thus observations were restricted to those occurring with no more than three standard deviations above the mean inventory level (an inventory of 68 or less). Figure 2 shows the distribution of agent inventory at time of sale and illustrates the abnormal relation of these entries to the rest of the dataset as well as the magnitude of data captured by eliminating observations with inventories further than three standard deviations above the mean.
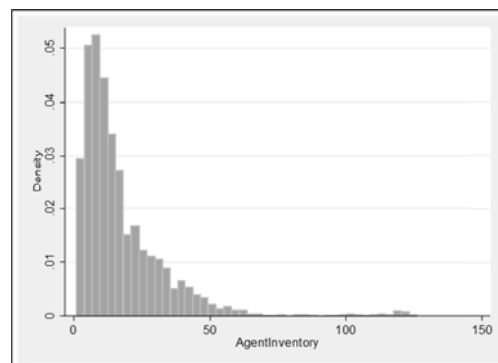

Figure 2. Distribution of agent inventories

---

[2]Listings containing obvious data-entry errors, such as houses listed or built before the 18th century, were removed entirely from the dataset.

Zip code data was found by matching addresses from the MLS system to ten-digit zip code values via SmartyStreets.com online software. Observations with addresses which the software failed to match were excluded. These consisted of 12% of the data.

Within the eight-year time period under study, 20,759 useable observations were listed and sold through the sav.mlxchange.com database. Additional observations not appearing in the final dataset due to incompleteness or failure to sell before being removed from the MLS were used to compute values dealing with agents' real estate inventory levels despite not appearing as observations in the final dataset.

Table 1. Selected summary statistics

Variable	Mean	Std Dev	Min	Max
Sale Price	$239,660	$173,335	$50,000	$3,000,000
Inventory 0-8	0.453	0.498	0	1
Inventory 9-26	0.401	0.490	0	1
Inventory 27-43	0.101	0.301	0	1
Inventory 44 plus	0.046	0.208	0	1
Full Baths	2.291	0.639	1	9
Water Front	0.128	0.334	0	1
Age 0-2	0.315	0.465	0	1
Age 3-5	0.134	0.341	0	1
Age 6-10	0.173	0.378	0	1
Age 11-20	0.168	0.374	0	1
Age 21-30	0.088	0.284	0	1
Age 31-40	0.044	0.205	0	1
Age 41-50	0.023	0.150	0	1
Age 51-100	0.053	0.225	0	1
Age 100+	0.004	0.060	0	1
Sold 2007	0.148	0.355	0	1
Sold 2008	0.118	0.322	0	1
Sold 2009	0.114	0.318	0	1
Sold 2010	0.108	0.310	0	1
Sold 2011	0.123	0.328	0	1
Sold 2012	0.135	0.342	0	1
Sold 2013	0.152	0.359	0	1
Sold 2014	0.102	0.302	0	1
n=20,759				

Note: fireplace, sales quarter, garage, laundry room, and 6-digit zip code variables are included in the model but omitted here for brevity.

5   Results

An OLS model was fit to the data using the natural log of the real sale price of a property in 2012 dollars as the dependent variable (lnprice). The natural log of real sale price is commonly used in the literature when comparing sales of homes with large differences in sale price using OLS analysis (Sirmans et al. [5]).

The lnprice regression model yields an adjusted R^2 of 0.77 and an F-statistic of 284. Results for most of the 268 control variables are not reported for the sake of brevity. Categories of variables which are not shown include controls for month sold, sewer system type, 6-digit zip code, construction characteristics, and exterior characteristics.

Results of the lnprice regression model are listed in Table 2. Variables of interest are listed in bold. "Inventory" denotes the agent's inventory of listings at the time of sale. The width of the "Inventory 9-

26" and "Inventory 27-43" fixed effects are one standard deviation each with the mean value of 17.6 centered inside "Inventory 9 – 26" while "Inventory 44+" simply represents all inventories greater than or equal to 44. The results of each inventory variable are reported relative to sales while the listing agent held inventories of between zero and eight homes. Real estate agents who held higher inventories transacted at a lower price on behalf of their clients, *ceterus paribus*. Agents holding between 9 and 26 listings were found to transact at a discount of 1.7% as compared to agents with 8 or less listings. Agents who held between 27 and 43 listings at the time of sale were found to transact at a 5.9% discount, and agents holding more than 43 listings transacted with a discount of 5.7% on the sale price of the house.

Table 2. Truncated OLS regression results with heteroskedasticity insensitive SE's

Dependent variable: ln(real sale price)							
Variable	*Coef*	*Std. Error*	*Sig.*	*Variable*	*Coef*	*Std. Error*	*Si g.*
Inventory 9 − 26	**-0.016**	**0.004**	*	Age 3 - 5	-0.008	0.009	*
Inventory 27 − 43	**-0.059**	**0.006**	*	Age 6 - 10	-0.139	0.008	*
Inventory 44+	**-0.057**	**0.009**	*	Age 11 - 20	-0.220	0.009	*
Square Feet 681 − 1536	0.414	0.027	*	Age 21 - 30	-0.292	0.011	*
Square Feet 1537 − 2391	0.632	0.027	*	Age 31 - 40	-0.303	0.013	*
Square Feet 2392 − 3246	0.848	0.028	*	Age 41 - 50	-0.306	0.014	*
Square Feet 3247 − 4101	1.007	0.030	*	Age 51 - 100	-0.280	0.013	*
Square Feet 4102 − 4957	1.172	0.033	*	Age 100+	-0.278	0.019	*
Square Feet 4958 - 5812	1.327	0.039	*	1.5 Story	0.050	0.007	*
Square Feet 5813+	1.359	0.043	*	2 Story	0.006	0.005	
Full Bathrooms	0.052	0.004	*	Waterfront	0.141	0.006	*
Bedrooms / Bathrooms	-0.019	0.005	*	Laundry Room	0.035	0.004	*
1 Garage Space	0.071	0.006	*	High Ceilings	0.019	0.007	*
2 Garage Spaces	0.188	0.005	*	Hardwood Floors	0.128	0.004	*
3 Garage Spaces	0.248	0.011	*	1 Fireplace	0.062	0.004	*
4 Garage Spaces	0.299	0.025	*	2 Fireplaces	0.100	0.011	*
5 Garage Spaces	0.306	0.061	*	3 Fireplaces	0.167	0.024	*
R^2:	0.77						
F-Statistic:	284		*: Statistical significance at the 1% level				
# of observations	20,759						
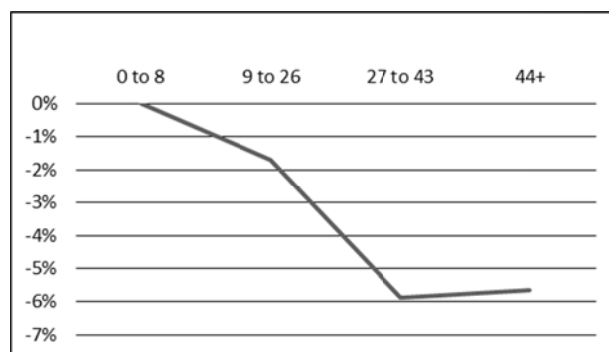
Figure 3. Inventory parameters by category

As figure 3 shows, the effect of agent inventory on the natural log of real sale price appears to grow at inventory levels in the 27 to 43 category, but then plateau afterward. Estimating the model's parameters again with a continuous version of agent inventory for only inventory levels of 43 or fewer produces an estimated -0.21% per each additional listing held in the listing agent's inventory at the time of sale, up to an inventory of 43. Repeating this process using only observations with inventories between 44 and 68 (the cap) did not yield a discount of statistical significance in the highest level of inventory.

These results could be interpreted as evidence for the hypothesis of agents sacrificing some degree of sale price in order to increase turnover volume. However, the marginal effect of inventory on home price does not appear to be constant in this model's results.

"Square Feet" variables are specified in the same way as the Inventory variables. The mean square footage of listings in the dataset is 1,963.8, which is centered in "Square Feet 1537 – 2391" with all variables spanning one standard deviation total. The standard deviation of the square footage of listings in the dataset is 855.2. Each square footage estimate is relative to houses with square footage below 681 square feet. Consistent with the bulk of the literature, increased square footage is found to increase the sale price of a house.

As expected, and consistent with the literature, properties with larger lot sizes, garage spaces, or fireplaces were found to correlate to an increased average sales price of listings while the increased age of a house was found to decrease the sales price. Additionally, being along the waterfront, containing a laundry room, and having hardwood floors increase the average sales value of properties in the dataset.

The listing's number of stories showed a significant positive relationship of 1.5 story homes. No notable significance was found in two, three, or five story homes while four story homes had an effect of 15.7% on sale price. Given that the square footage of a house was accounted for otherwise, this result implies that buyers prefer 1.5 story homes more than one story, but not houses with an abnormally higher number of stories for the region. Of all homes in the dataset, 61% were either single story or 1.5 story homes. The large effect of homes with four stories is likely due to their relatively high correlation with the 31401 zip code which corresponds to historic downtown Savannah.

The number of full bathrooms was found to have a positive relationship with the sale price of the house. This result is in agreement with the bulk of the literature (Sirmans et al. [5]). Each additional bathroom was found to increase the sale price of the house by an estimated 5.8%. Given that such a wide range of houses were compared by this study, a ratio of bedrooms to bathrooms, "Bedrooms/Bathrooms," was included as a control variable in order to make the estimated value of "Full Bathrooms" interpretable. Houses which contain a large number of bedrooms relative to the number of available bathrooms were found to lower sale price, all else equal.

Houses were found to devalue continuously with age up until reaching 100+ years old. It is surmised that properties aging over 100 years in the Savannah area are quite likely to be of some historic significance, which could decrease the negative effect of age on sale price at very high age values.

6 Conclusion

This study finds evidence that could be interpreted as consistent with the previous literature on the existence of asymmetrical information in the market for real estate. Significant negative correlations exist between the inventory holdings of real estate agents and the average sale price of one of their listings. This result offers evidence for the hypothesis that real estate agents neglect their clients to some degree by taking on above-optimal levels of inventory, from the perspective of their individual clients, in order to increase their own commission income. Losses to the homeowner (discounted transaction price) are estimated at 0.21% per each additional property held in "inventory" by the listing agent, up to inventory levels of 43 in a model where agent inventory enters linearly. In the preferred model, where inventory is allowed to have nonlinear effects, agents holding 27 homes or more sell them for almost 6% less than agents holding only a few listings.

As expected, fireplaces, garage spaces, laundry rooms, waterfront adjacency, wood floors, high ceilings, desirable construction materials, additional lot size, public sewers, square footage, and additional bathrooms were all found to have positive relationships with sale price while houses were found to depreciate with age.

A clearer understanding of the tradeoff between higher inventory and lower sales price could have implications for the management of brokerage firms and agents. While larger holdings do seem to imply a discounted sales price and thus a lower commission, the marginal effect on sales price of an additional holding is relatively small (around $300 at the median home price) and thus an agent adding the marginal listing to inventory is likely to generate a net benefit for the agent and firm.

An extension for possible future consideration is to compare inventory levels, and the effects of inventory levels on sales outcomes, across real estate markets with different median home prices. Given the fixed commission rate structure common in the brokerage industry, a higher median transaction price may encourage additional entry, thus lowering the inventory holdings of a typical agent and yielding more favorable (higher) sales price transactions for their clients.

References

1. Levitt, S.D. and C. Syverson, "Market distortions when agents are better informed: The value of information in real estate transactions," Review of Economics and Statistics, vol. 90, no. 4, pp. 599-611, 2008.
2. Anglin, P. M., "Residential real estate brokerage as a principal-agent problem," Journal of Real Estate Finance and Economics, vol. 4, no. 2, pp. 99-125, 1991.
3. Beck, J., J. Fralick, and M. Toma, "The Housing price bubble in a suburban Georgia setting: Using the hedonic pricing model in the new south," The Journal of Applied Business Research, vol. 28, no. 4, pp. 651-659, 2012.
4. Dubin, R., "Predicting house prices using multiple listings data," Journal of Real Estate Finance and Economics, vol. 17, no. 1, pp. 35-59, 1998.
5. Sirmans, S., D. Macpherson, and E. Zietz, "The Composition of hedonic pricing models," Journal of Real Estate Literature, vol. 13, no. 1, pp. 3-43, 2005.
6. Hamilton, S. W. and G. M. Schwann, "Do high voltage electric transmission lines affect property values?" Land Economics, vol. 71, no. 4, pp. 436-44, 1995.
7. Judd, D., T. Seaks, and D. Winkler, "The Impact of residential brokerage," Journal of Real Estate Research, vol. 22, no. 3, pp. 447-458, 1996.
8. Haag, J., R. Rutherford, and T. Thomson, "Real estate agent remarks: help or hype? Journal of Real Estate Research, vol. 20, no. 1, pp. 205-215, 2000.
9. Forgey, F., R. Rutherford, and T. Springer, "Search and liquidity in single-family housing," Real Estate Economics, vol. 24, no. 3, pp. 273-292, 1996.
10. Rutherford, R., T. Springer, and A. Yavas, "The Impacts of contract type on broker performance," Real Estate Economics, vol. 29, no. 3, pp. 389-410, 2001.
11. Center for Real Estate Technology, "The New real estate mantra: Location near public transportation," Chicago: March 21, 2013.