Regional Science and Urban Economics 42 (2012) 189–197

Contents lists available at SciVerse ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

Are homes hot or cold potatoes? The distribution of marketing time in the housing market☆

Paul E. Carrillo [a,*], Jaren C. Pope [b]

[a] Department of Economics, George Washington University, United States
[b] Department of Economics, Brigham Young University, United States

ARTICLE INFO

Article history:
Received 15 April 2011
Received in revised form 19 August 2011
Accepted 20 August 2011
Available online 27 August 2011

JEL classification:
R21
R30
C21
C41

Keywords:
Time on market
Liquidity risk
Quantile decomposition
Duration model
Censoring

ABSTRACT

This paper analyzes how the distribution of marketing time of residential real estate evolves across time. Using real estate data from a large suburb in the Washington D.C. area we first show that the whole distribution of marketing time shifts to the right when a "hot" housing market in 2003 is compared with a "cold" one in 2007. The shift, however, is not homogenous across the distribution: it is negligible at lower percentiles, very large at the median and much smaller at higher percentiles. Moreover, the shift in the distribution cannot be explained by changes in the characteristics of the units. We then compute (quality adjusted) time on the market distributions and hazard functions for each year during the period 1997 to 2007. We find that while there are no changes at the bottom of the (conditional) distribution over time, higher percentiles, such as the first quartile and the median, are notably more volatile. We also find that the distribution of marketing time is heterogeneous across property types and property location. The focus on the distribution of marketing time rather than solely on the mean or on the median provides a comprehensive description of the evolution of this asset's liquidity and may help homeowners and financial institutions to better grapple with liquidity risk.

1. Introduction

The recent housing market boom and bust was staggering. The Federal Reserve's Flow of Funds Report documents that the asset value of owner-occupied housing units for the entire U.S. was approximately 13 trillion dollars in 2000, rose to approximately 23 trillion dollars in 2006, and then fell dramatically to approximately 16 trillion dollars by 2010.[1] The media and popular housing price indices such as the Case–Shiller index have documented the substantial heterogeneity in price fluctuations across metro areas. Recent work has begun to document that there was also substantial heterogeneity in price fluctuations *within* a metro area.[2] It is clear that these price fluctuations have adversely affected many households and financial institutions. However, housing price fluctuations are only part of the story and part of the costs of the recent boom and bust. As the housing market began entering the bust phase, housing assets became much more illiquid as the time it took to sale a house increased. Many homeowners and financial institutions were left with an illiquid financial asset when they desperately needed liquidity. Although this liquidity risk in the housing market has obvious importance and is likely to be heterogeneous within a metro area, there has not been a systematic analysis in the literature of how the *distribution* of marketing time evolves during "hot" and "cold" housing markets.

This paper attempts to fill this gap by providing a simple framework to analyze how the liquidity of residential real estate changes over time, where liquidity is measured by the time that it takes to sell (time on the market). The focus on the distribution of marketing time rather than solely on the mean or on the median provides a more comprehensive picture of how the liquidity of this asset evolves. For instance, using residential real estate data from a large suburb in the Washington D.C. area (Fairfax County, VA), we show that the whole distribution of marketing time shifts to the right when we compare a tight market in 2003 with a slow one in 2007.[3] The shift, however, is not homogenous across the distribution: it is negligible at lower percentiles (in both periods, at least 2% of transactions occurred in less than 1 day), very large

☆ We thank the Editor, anonymous referees, David Jacho-Chávez, Anthony Yezer and colleagues at Brigham Young University and George Washington University for comments and support. We are also grateful to Kyle Harbacek for excellent research assistance and MRIS for sharing the data with us. All errors are our own.

* Corresponding author.
E-mail address: pcarrill@gwu.edu (P.E. Carrillo).

[1] This information was taken from Table B.100 entitled "Balance Sheet of Households and Nonprofit Organizations" in the Federal Reserve's Flow of Funds Report which can be found at http://www.federalreserve.gov/releases/z1/current/ and was last accessed on 3/29/2011.

[2] See for example Kuminoff and Pope (2011) and Landvoigt et al. (2011). McMillen (2003) also documents similar spatial heterogeneity in housing price fluctuations in the early 1990s in the Chicago metro area.

[3] Focusing on data for 2003 and 2007 is convenient for our analysis since the number of listings in each year was basically the same, but the time on the market was substantially higher in 2007 than in 2003.

0166-0462/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.regsciurbeco.2011.08.010

at the median (median time on the market is about 11 times larger in the slow market) and much smaller at higher percentiles. The levels of the hazard function in 2003 contrast with those in 2007, particularly during the first 2 months of the listing. For instance, the likelihood of selling a home given that it has been on the market for 1 week is about 6 times higher in 2003 compared to that in 2007. These differences sharply decline with time on the market and virtually disappear after 6 months. We also show that the distribution of time on the market can vary by home type, and by the spatial location of the home within the county.

For our analysis of the distribution of time on the market it is important to ask if the shift in the *unconditional* distribution of marketing time and in the *unconditional* hazard function between 2003 and 2007 can be explained by differences in the characteristics of the housing units that were traded in these periods. This is a valid concern since homes for sale in 2003 are indeed statistically different than those for sale in the latter period. To address this question, we extend a method developed by DiNardo, Fortin and Lemieux (1996) to decompose the shift in the distribution into a component that can be attributed to changes in observed housing characteristics and an unobserved component. The extension proposed in this paper combines the DiNardo, Fortin and Lemieux approach with the Kaplan–Meier estimator (Kaplan and Meier, 1958) to allow the decomposition to work in cases where the dependent variable is subject to random censoring. This method allows us to simulate the distribution of time on the market in 2007 as if the housing units in this period had the same characteristics as those units for sale in the 2003 sample. The difference between the estimated counterfactual distribution and the actual distribution of time on the market in 2007 is minimal. Thus, we conclude that the observed shift in the distribution and in the hazard function is not a consequence of changes in the characteristics of the houses in our sample. Rather it can be almost fully attributed to other unobserved factors that affect the liquidity of residential real estate.

Finally, we use this approach to compute "quality adjusted" time on the market distributions and hazard functions for each year during the period 1997–2007. In particular, we simulate the duration distribution and hazard function during each year in our sample assuming that housing units had the same characteristics as homes in 2003. The results show that the *conditional* median time on the market is quite volatile: it sharply decreased from 133 days in 1997 to 12 days in 2000, remained somewhat constant between 2001 and 2004, and then increased close to 1997 levels by the end of 2007. However, there is substantially less volatility at lower percentiles. For instance, in almost all periods, at least 2% of transactions occurred in less than 1 day. Similarly, the hazard rate for those properties that have been on the market for 1 week exhibit substantial volatility over time: during our ten year sample it ranges from 0.007 to 0.077. As time on the market increases, this volatility sharply diminishes. For example, the likelihood of selling a home given that it has been on the market for 6 months lies between 0.003 and 0.005. In summary, to measure the liquidity of the housing market and to help homeowners and financial institutions to better analyze liquidity risk, it is important to look beyond the conditional mean and median and focus on the full distribution of time on the market.

Our paper was motivated by the work of McMillen (2008) who analyzes changes in the *distribution* of home prices in Chicago over time.[4] He finds that the shift in home prices between 1995 and 2005 is significantly larger at the right tail of the distribution and that these shifts cannot be fully attributed to changes in the structural characteristics nor the location of the housing stock.[5] Our work is different than McMillen's in two important ways. First, we focus on time on the market instead of housing prices, and second, we use a different decomposition method that lends itself to studying time on the market. The decomposition used by McMillen which is based on quantile regressions (Machado and Mata, 2005) is not suitable for our application, because marketing time is subject to random censoring. The DiNardo, Fortin and Lemieux decomposition has been used to explain changes in the distribution of home prices (Cobb-Clark and Sinning, 2011). The extension to the DiNardo, Fortin and Lemieux method we propose a methodological contribution to the literature that may be easily implemented in other circumstances when the dependent variable is censored.

This paper is also related to an extensive literature that analyzes the determinants of time on the market. For example, Genesove and Mayer (1997, 2001) have shown that the seller's equity position and loss aversion are key determinants of marketing time: the smaller the equity, the longer the time to sell. Haurin (1998) shows that the time a home stays on the market, is related to the atypicality of the housing unit. Other authors have analyzed the relationship between list prices and marketing time from both a theoretical and empirical point of view (Allen et al., 2009; Anglin et al., 2003; Carrillo, forthcoming; Haurin et al., 2010; Horowitz, 1992; Kang and Gardner, 1989; Knight, 2002; Springer, 1996 and Yavas and Yang, 1995, among others). The focus of the previous literature, however, is on the conditional mean. To the best of our knowledge, ours is the first attempt to evaluate changes in the full distribution of time on the market.

The paper is organized as follows. The next section presents and discusses the data. The third section presents the decomposition method and results from a comparison between the 2003 and 2007 time periods. Section 4 computes the distribution of time on the market for each year during the 1997 to 2007 time period. Finally, the last section concludes the study.

2. Data

Our analysis uses residential real estate data from Fairfax County, Virginia. Fairfax County is located in northern Virginia and it is part of the Washington, D.C. metropolitan statistical area. This county hosts more than one million residents and more than 350,000 housing units (Fairfax County website 2010) and it is one of the richest and best-educated counties in the United States.[6]

We gathered data from the local Multiple Listing Service (MLS) and collected information from all housing listings that were posted on the MLS between January 1, 1997 and December 31, 2007. The data include all listings that ended up in a transaction as well as those that expired or were withdrawn from the market. The data contain detailed property characteristics, such as the number of bedrooms, bathrooms, age and location, as well as list prices, transaction prices and the time that the listing stayed on the market (time on the market).

Time on the market is measured by the number of days that the MLS listing stays "active" on the market. For units that are sold, we compute marketing time as the difference between the date when an offer was accepted and the date when the listing was posted.[7] When a listing is withdrawn from the market or it expires without a sale, we compute the time between the initial listing and withdrawal, and treat it as a censored observation. Notice that we analyze the time that a *listing* stays on the market, which can be different from the total time that the *property* has been on the market. This occurs because sellers can withdraw the listing for a few days, weeks or

[4] The paper by Deng et al. (forthcoming) also looks at distributional issues in real estate prices in Singapore.

[5] A few recent studies in urban economics also analyze changes in the distribution of the dependent variable. For example, Cobb-Clark and Sinning (2011) compare the distribution of home prices between natives and immigrants in Australia, and Carrillo and Yezer (2009) evaluate the differences in homeownership rates between segregated neighborhoods.

[6] It ranked second in median household income in 2008, with 58.5% of adults over the age of 25 holding at least a bachelor's degree.

[7] Notice that the date when the offer is accepted is generally different than the transaction date recorded in court. It usually takes between 3 and 8 weeks from the date when the offer is accepted to complete the sale process (this includes the inspection process, securing financing, etc.).

even months and then put the property back on the market as a "new listing." We exclude from our sample listings with unusually high or unusually low listing prices (top and bottom 1% during each year), about 240 observations that stayed on the market for more than 2 years, and about 800 observations with missing data. After this cleaning process, we are left with 284,678 listings.

In the rest of this section, we focus on two periods that are particularly interesting: 2003 and 2007. Although other years could have been chosen, we focus on 2003 and 2007 since it was clear that time on the market varied dramatically between these two years. During the first period, the housing market could be thought of as a "hot" housing market. Annual home appreciation rates were close to 20% and most properties sold in less than 2 weeks. During the second period the housing market could be thought of as a "cold" housing market where properties took longer to sell and home values were in decline. To analyze these periods we use two datasets. The first (second) dataset includes all listings that were posted between January 1, 2003 (2007) and December 31, 2003 (2007). In Section 4, we discuss the data from all other years.

Table 1 shows descriptive statistics for Fairfax County, VA in 2003 and 2007. It is clear that market conditions in both periods are notably different. First, notice that list prices and home sale values are significantly higher in 2007. Both median list price and median sale price are about $120,000 higher in 2007 when compared to 2003. A nice feature of focusing on these two years is that the number of listings in 2003 and 2007 were approximately the same (around 27.5 thousand listings). The descriptive statistics provide some interesting initial insights about the liquidity of residential real estate. Out of 27,487 residential real estate listings that were posted on the MLS in 2003, about 82% ended up in a sale; on the other hand, of the 27,663 real estate listings that were posted in 2007, only 46% were sold. Among those units that were sold, the mean and median marketing time in 2007 is more than 3 and 4 times larger than their counterparts in 2003. Clearly, the housing market in 2003 was considerably "hotter" than in 2007.

The mean and median time on the market, such as those computed in Table 1, are part of a battery of indicators generally used by real estate practitioners to assess changes in housing market conditions. Homeowners also use information about the time on the market (typically average TOM given to them by a real estate agent) to better understand the potential costs involved with trying to sale their house. However, these indicators can be misleading for many reasons. First of all, these statistics are subject to censoring bias because they refer only to those listings that ended up in a sale. One can expect this bias to be particularly important for the 2007 sample where the share of censored observations is large. Moreover, a focus on the median or on the mean can hide important shifts at other points of the distribution of marketing time. For instance, even if the average and median home takes longer to sell in 2007, it may be the case that a share of units sells fairly quickly in both periods.

Table 1
List prices, transaction prices and marketing time.

	2003				2007			
	Mean	Median	St. Dev.	N	Mean	Median	St. dev.	N
All listings								
List price ($ 000)	392.9	339.9	220.5	27,487	535.2	459.9	289.3	27,663
Sold listings								
Sale price ($ 000)	367.7	325.0	195.3	22,606	509.9	445.0	258.9	12,745
Time on market (days)	18.6	9.0	26.6	22,606	60.8	38.0	65.4	12,745

Notes: Table shows descriptive statistics of Fairfax County, VA, residential real estate listings posted on the MLS during each of the years specified above. The first row shows statistics for all listings posted during each time period. The second and third rows refer only to the subset of listings that ended up in a sale. Notice that the sale or listing withdrawal does not necessarily occur within the specified periods.

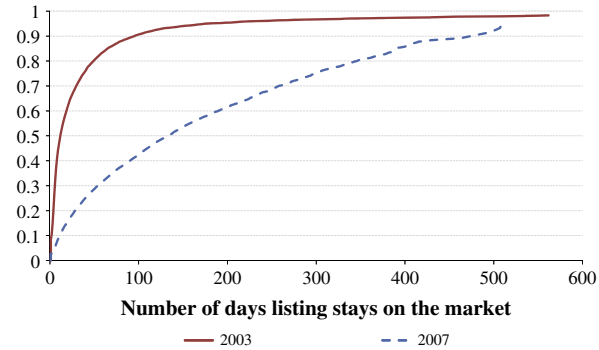

Fig. 1. Distribution of marketing time.

To illustrate the severity of these concerns, we use the Kaplan–Meier estimator to compute a non-parametric estimate of the distribution of time on the market for both the 2003 and 2007 sample. Any listing that is either expired or withdrawn is treated as a censored observation. Results are shown in Fig. 1 and in Table 2. First note that when censoring is accounted for, the estimate of the median marketing time changes by a significant amount. In the 2003 sample where the share of censored observations is relatively small, the estimate of the median duration increases from 9 to 12 days (see estimate in Table 1 compared to Table 2). The difference is much larger for the 2007 sample: the Kaplan–Meier estimate of the median time on the market is 138 days, about 3.5 times larger than the median time from the uncensored observations (38 days as shown in Table 1). Thus it appears that the inclusion of censored observations is critical to accurately measure the liquidity of residential real estate.

Results from Fig. 1 and Table 2 also show that the whole distribution of marketing time shifts to the right when we compare a "hot" market in 2003 with a "cold" one in 2007. However, the shift is *not* homogenous across the distribution. There is virtually no shift at very low percentiles. For example, tabulations in Table 2 show that in *both* periods, more than 2% of transactions occurred in less than 1 day. As we move up to higher quantiles, the shift increases both in magnitude and in relative terms (see Fig. 2). For instance, the median time on the market is about 11 times larger in the slow market (133 days) compared to the tight market (12 days).[8] But the relative shift is much smaller at the top of the distribution. For example, the 75th and 90th percentile in the 2007 sample are only about 8 and 5 times larger than their counterparts in 2003. We also compute hazard rates for each of our samples and show results in Fig. 3.[9] During 2003, there is strong negative duration dependence: the hazard rate is quite large during the first week of the listing and slowly decreases thereafter. The hazard rate in 2007, on the other hand, shows a markedly different pattern: even though there is some evidence of negative duration dependence during the first 45 days, the hazard is more or less constant after that point. The levels of the hazard function in 2003 contrast with those in 2007, particularly during the first 2 months of the listing. For example, the likelihood of selling a home given that it has been on the market for 1 week is about 6.5 times higher in 2003 compared to that in 2007. These differences sharply decline with time on the market and virtually disappear after 6 months. In

[8] The spike in the lowest quantiles is due to the fact that a larger share of properties sale in 1 day in 2003 than in 2007.

[9] We computed cumulative hazard functions using the Nelson–Aalen estimator (Nelson, 1972 and Aalen, 1978). A Gaussian kernel function was used to smooth the cumulative hazard function. The hazard function was then estimated by computing the curvature of the smoothed cumulative hazard. The same bandwidth (3 days) was used for both samples.

Table 2
Distribution of marketing time (Kaplan–Meier).

Percentile	Quantiles (days)		Q2007/ Q2003
	2003	2007	
01	1.0	1.0	1.0
02	1.0	1.0	1.0
05	1.0	5.4	5.4
10	1.7	11.2	6.6
25	4.6	39.8	8.6
50	11.7	133.4	11.4
75	37.8	299.5	7.9
90	95.1	474.6	5.0

Notes: Table shows Kaplan–Meier estimates for the distribution of residential real estate listings' marketing time in Fairfax County, VA.

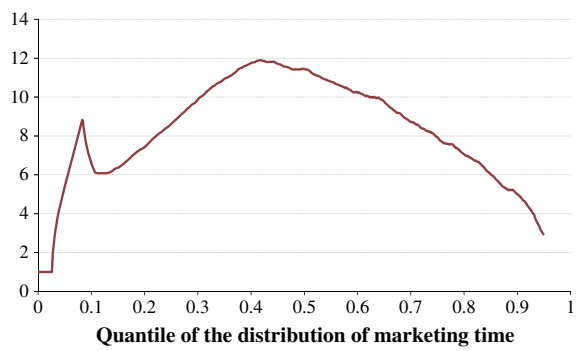

Fig. 2. Ratio between the distribution of marketing time in 2007 and the distribution of marketing time in 2003.

summary, to assess shifts in the liquidity of housing over time, it is important to pay attention to changes in the whole distribution of marketing time, rather than focusing solely on the mean or on the median.

Can the shift in the *unconditional* distribution of marketing time and the *unconditional* hazard function between 2003 and 2007 be explained by differences in the characteristics of the housing units that were traded in these periods? This is a valid concern because homes for sale in 2003 may be systematically different than those for sale in the latter period. To test if this is the case, we have collected data on several characteristics of the units including their number of bedrooms, number of bathrooms, age and type, among others.[10] Descriptions of the variables as well as their descriptive statistics (for the 2003 and 2007 samples) are shown in Table 3. The average home in the pooled sample is a 25 year old unit with 3.4 bedrooms, 3 bathrooms, and 1 fireplace. Homes for sale in 2007 are statistically different than homes for sale in 2003: they are slightly larger (have more bathrooms and bedrooms) and are about 3 years older. Thus, part of the shift of the distribution of time on the market could be potentially explained by the changes in the characteristics of the housing stock.

3. Decomposition

To assess how much of the shift of the marketing time distribution and hazard function can be explained by changes in the characteristics of the housing units, we extend a decomposition method proposed by DiNardo, Fortin and Lemieux (1996) (DFL). The extension we propose simply combines the Kaplan–Meier estimator with the DFL decomposition to allow it to work in cases where the dependent variable is subject to random censoring. To keep our exposition self-contained we first review the decomposition method before we present results.

[10] We use the information on the field "type" on the MLS listing to identify if the housing unit is a detached single family home, a townhome or an apartment.

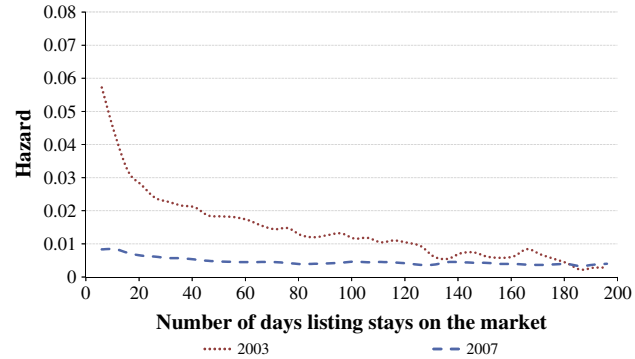

Fig. 3. Hazard function.

3.1. Method

Here we provide a careful description of the DFL approach using the same notation as in Leibbrandt et al. (2010). Let Y be our variable of interest (time a listing stays on the market) and t_0 and t_1 refer to the two mutually exclusive periods (years) we analyze. The cumulative probability function of Y in period t_0 is defined as

$$F(y|T=t_0)=P\{Y\leq y|T=t_0\}=\int F(y|x,T=t_0)h(x|T=t_0)dx, \tag{1}$$

where T is a random variable describing the period from which an observation is drawn and x is a particular draw of observed attributes of individual characteristics from a random vector X. $F(y|x,T=t_0)$ is the (conditional) cumulative distribution of Y given that a particular set of attributes x have been picked, and $h(x|T=t_0)$ is the probability density of individual attributes evaluated at x. The cumulative probability function of Y in period t_1 is defined similarly.

Suppose we would like to assess how the distribution of Y (marketing time) in period t_1 (2007) would look if the individual attributes x (number of bathrooms, bedrooms and age, for example) were the same as in period t_0 (2003). We denote this counterfactual as $F_{t_1\to t_0}$ and express it symbolically as[11]

$$F_{t_1\to t_0}(y)=\int F(y|x,T=t_1)h(x|T=t_0)dx. \tag{2}$$

In order to compute such a counterfactual, Bayes' rule is used to obtain $h(x|T=t)=P(X=x)P(T=t|X=x)/P(T=t)$. DFL recognized that

$$\frac{h(x|T=t_0)}{h(x|T=t_1)}=\frac{\frac{P(T=t_0|X=x)}{P(T=t_0)}}{\frac{P(T=t_1|X=x)}{P(T=t_1)}}=\frac{\frac{P(T=t_0|X=x)}{1-P(T=t_0|X=x)}}{\frac{P(T=t_0)}{1-P(T=t_0)}}=\tau_{t_1\to t_0}(x). \tag{3}$$

One may use expression (3) to substitute $h(x|T=t_0)$ in Eq. (2) and thereby obtain expression (4).

$$F_{t_1\to t_0}(y)=\int F(y|x,T=t_1)h(x|T=t_1)\tau_{t_1\to t_0}(x)dx \tag{4}$$

Notice that this expression differs from Eq. (1) only by $\tau_{t_1\to t_0}(x)$. DFL refer to $\tau_{t_1\to t_0}(x)$ as "weights" that should be applied when computing the counterfactual distribution of our variable of interest. However, given that the weights are unknown, they need to be estimated.

The DFL method described above cannot be directly used in our application because marketing time is subject to random censoring; that is, some properties are not sold and withdrawn from the market. The extension to the DFL method we propose here is straightforward. Because the random variable Y (marketing time) is subject to random

[11] The subscript "$t_0\to t_1$" indicates that the attributes data from period t_0 will be "replaced" by data from period t_1.

Table 3
Characteristics of housing units.

Variable	Description	Pooled sample		2003 sample		2007 sample		Mean diff. (2007–2003)	T-test
		Mean	St. dev.	Mean	St. dev.	Mean	St. dev.		
Bedrooms	Total number of bedrooms	3.380	1.096	3.354	1.045	3.406	1.144	0.053	5.63
Bathrooms	Total number of full and half bathrooms	3.071	1.132	3.059	1.099	3.084	1.163	0.025	2.58
Age	Age of unit in years (zero if new)	24.55	16.15	22.86	15.56	26.22	16.54	3.357	24.55
Fireplaces	Total number of fireplaces	0.909	0.729	0.929	0.730	0.888	0.728	−0.041	−6.65
Central heat	Indicator if unit has central heat	0.936	0.245	0.933	0.251	0.939	0.239	0.007	3.12
Detached	Indicator if single family unit	0.488	0.500	0.490	0.500	0.486	0.500	−0.005	−1.11
Townhome	Indicator if unit is townhome	0.349	0.477	0.355	0.479	0.342	0.475	−0.012	−3.08
Apartment	Indicator if unit is in an apartment complex	0.163	0.370	0.155	0.362	0.172	0.377	0.017	5.47

censoring, the counterfactual distribution can be computed using the Kaplan–Meier estimator, with sampling weights given by $\tau_{t_1 \to t_0}(x)$. Similarly, one could compute the counterfactual hazard function using the Nelson–Aalen estimator and the same sampling weights.

To be specific, as in Leibbrandt et al. (2010), we summarize the estimation algorithm for the counterfactual given that a random sample of N_0 and N_1 observations (that include both censored and uncensored observations) for periods t_0 and t_1 is available:

Step 1 Estimate $P(T=t_0)$ using the share of observations where $T_i=t_0$; that is, compute: $\hat{P}(T_i=t_0)=N_0/(N_0+N_1)$.

Step 2 Estimate $P(T=t_0|X=x)$, by estimating a logit model using the pooled data. The dependent variable equals one if $T_i=t_0$, and explanatory variables include the vector of individual attributes x_i.

Step 3 For the subsample of observations where $T_i=t_1$, estimate the predicted values from the logit $\hat{P}(T_i=t_0|X=x_i)=\frac{\exp\{x_i\hat{\beta}\}}{1+\exp\{x_i\hat{\beta}\}}$, where $\hat{\beta}$ is the parameter vector from the logit regression. Then, compute the estimated weights as

$$\hat{\tau}_{t_1\to t_0}(x)=\frac{\frac{\hat{P}(T=t_0|X=x)}{1-\hat{P}(T=t_0|X=x)}}{\frac{\hat{P}(T=t_0)}{1-\hat{P}(T=t_0)}}.$$

Step 4 For the subsample of observations where $T_i=t_1$, compute a weighted empirical cumulative distribution function using the Kaplan–Meier estimator. Weights are given by $\hat{\tau}_{t_1\to t_0}(x_i)$.

Step 5 For the subsample of observations where $T_i=t_1$, compute a weighted hazard function using the Nelson–Aalen estimator. Weights are given by $\hat{\tau}_{t_1\to t_0}(x_i)$.

It is useful to analyze the differences between the distributions of interest at each quantile. Specifically, define $Q_\theta[Y|T=t]$ as the θth quantile of the distribution of Y in period t. Similarly, let $Q^*_\theta[Y|T=t_1]$ be the θth quantile of the counterfactual distribution of Y in period 1 (year 2007) if individual attributes where identical to those in period 0 (year 2003). Notice that Q^*_θ is implicitly defined by $\hat{F}_{t_1\to t_0}(y)$. In the Results section, we essentially compare $Q_\theta[Y|T=t_1]$ with $Q^*_\theta[Y|T=t_1]$.

It is useful to note at this point that there are alternative approaches to compute the counterfactual distribution of marketing time. For example, one could use proportional hazard models to compute the 2007 counterfactual distribution and hazard function. Proportional hazard models, however, generally rely on strong functional form assumptions. The DFL method, on the other hand, relies on minimal parametric assumptions. Another option to compute the counterfactual is the semi-parametric decomposition based on quantile regressions proposed by Machado and Mata (2005). This method requires fewer parametric assumptions and it has been used in several recent applications in labor, development and urban economics.[12] It is not clear, however, how to apply this methodology (or its properties) when the dependent variable is subject to random censoring.[13] We prefer the DFL decomposition to other alternatives because it can be easily implemented when censored observations are present, it makes mild parametric assumptions and it is computationally straightforward.

3.2. Results

Using the methods and data described above we have estimated the counterfactual distribution of marketing time and its corresponding counterfactual hazard function. The counterfactual simulates marketing time in 2007 treating housing characteristics in this period as identical to housing characteristics in 2003. Covariates include all variables displayed in Table 3.[14]

To assess how much of the relative shift in the distribution of Y between period t_0 and t_1 can be "explained" by changes in individual attributes x, we first compare the relative shift of the unconditional distribution $Q_\theta[Y|t_1]/Q_\theta[Y|t_0]$ with the relative shift of the counterfactual distribution $Q^*_\theta[Y|t_1]/Q_\theta[Y|t_0]$. If the actual and counterfactual shifts are similar, this would suggest that changes in the distribution of Y over time are not explained by changes in individual covariates. Results shown in Fig. 4 suggest that, indeed, the differences between the actual and the counterfactual shift are minimal throughout the distribution both in absolute and relative terms. For example, the unconditional and counterfactual median time that a listing stays on the market in 2007 is 11.1 and 10.8 times greater than the median time in 2003, respectively. Thus the shift in the unconditional median is only 0.3 days or 3% higher than the shift in the counterfactual median. These small differences virtually disappear as higher percentiles are reached.[15] We also compare the unconditional hazard ratio $h[y|t_1]/h[y|t_0]$ with the counterfactual hazard ratio $h^*[y|t_1]/h[y|t_0]$ and show results in Fig. 5. It is clear that the differences between the actual and counterfactual hazard ratios are negligible. In summary, virtually none of the shift of the marketing time distribution and hazard function between 2003 and 2007 can be explained by changes in the characteristics of the housing units. Hence, the shift can be almost fully attributed to other unobserved factors that affect the liquidity of

[12] See for example, Albrecht et al. (2003), Nguyen et al. (2007), McMillen (2008) and Carrillo and Yezer (2009).

[13] Powell et al. (2002) develop an estimator for a quantile regression when the dependent variable is subject to random censoring. One could potentially use this estimator to estimate the quantile regressions and perform the Machado–Mata algorithm. This estimator, however, is computationally expensive and more difficult to implement.

[14] Covariates include the number of bedrooms, number of bathrooms, seven discrete categories for the unit's age, an indicator if the unit has one fireplace, an indicator if the unit has more than one fireplace, an indicator if the unit has central heat, an indicator if the unit is detached, and an indicator if the unit is a townhome.

[15] Confidence intervals (and standard errors) can be computed using a bootstrap procedure. For example, one could randomly re-sample data (with replacement) from the 2003 and 2007 periods and use the methods described in this paper to estimate the relative shift of the counterfactual distribution $Q^*_\theta[Y^j|t_1]/Q_\theta[Y^j|t_0]$, where $j=1..K$ denotes a re-sample iteration. After this procedure is performed K times, one can compute standard errors and confidence intervals for each chosen quantile θ. Similar procedures have been performed by Carrillo and Yezer (2009) and Nguyen et al. (2007).

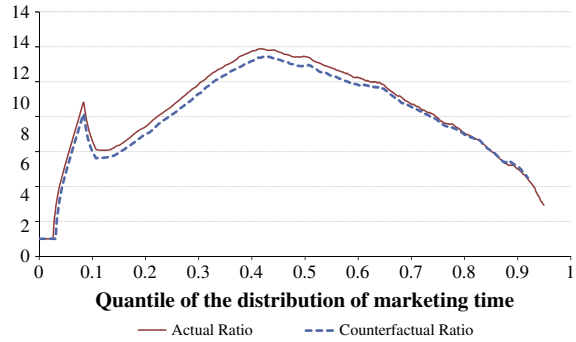

Fig. 4. Ratio between the counterfactual distribution of marketing time in 2007 and the actual distribution of marketing time in 2003.

residential real estate. Nonetheless, as will be shown in Section 4 changes in the characteristics of housing may be more important as the gap between years being compared increases, making it more likely that housing characteristics changed substantially.

4. Measuring changes in the liquidity of housing

In this section we use the methods described above to compute "quality adjusted" time on the market distributions for each year during the period 1997–2007 in Fairfax County, VA. This allows us to compute several indicators that describe the liquidity of this market over time.

The data used in this section come from the same source (MLS) and contain the same variables as the data described in Section 2. That is, for all listings that were posted on the Fairfax County, VA, MLS between January 1, 1997 and December 31, 2007, we observe the date when it was listed, the number of days that the listing stayed active on the market, whether the unit was sold or whether the listing was withdrawn from the market, as well as all housing characteristics listed in Table 3. On average, our sample contains about 26,000 observations per year.

The first panel of Table 4 shows Kaplan–Meier estimates for the unconditional distribution of time on the market. Listings that do not end up in a sale are treated as censored observations. Each row refers to listings posted in a particular year. For example, the first row shows seven percentiles of the marketing time distribution as well as the sample size and share of censored observations of all housing listings that were posted on the MLS between January 1, 1997 and December 31, 1997.

The upper panel of Table 4 provides interesting insights about how the unconditional distribution of marketing time has changed over time. For example, notice that the first and second percentiles of the distribution remain constant in almost all periods: it appears that in about every year at least 2% of listings were sold in less than

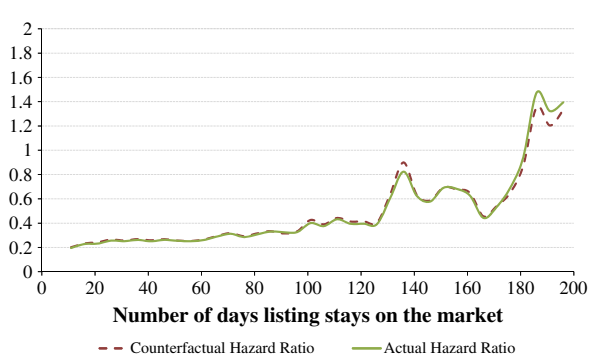

Fig. 5. Ratio between the counterfactual hazard function in 2007 and the actual hazard function in 2003.

Table 4
Distribution of marketing time (days on the market) in Fairfax County, VA.

a) Unconditional distributions (Kaplan–Meier)

Year	Percentile							N	Fraction censored
	01	02	05	10	25	50	75		
1997	1.0	1.8	6.1	14.2	45.9	126.2	316.9	23,100	0.438
1998	1.0	1.0	2.2	5.7	22.1	74.5	189.3	23,439	0.325
1999	1.0	1.0	1.0	1.7	6.9	31.0	92.8	21,875	0.212
2000	1.0	1.0	1.0	1.0	3.2	11.9	48.3	21,921	0.148
2001	1.0	1.0	1.0	1.1	3.5	14.1	58.2	25,349	0.190
2002	1.0	1.0	1.0	1.3	4.0	16.3	62.1	27,241	0.214
2003	1.0	1.0	1.0	1.7	4.6	11.7	37.8	27,487	0.178
2004	1.0	1.0	1.0	1.3	4.5	8.3	25.0	29,055	0.157
2005	1.0	1.0	1.0	3.0	6.3	20.2	88.6	32,578	0.309
2006	1.0	1.0	4.6	10.4	35.7	105.9	238.3	33,353	0.514
2007	1.0	1.0	5.4	11.2	39.8	133.4	299.5	27,663	0.539

b) Conditional distributions: marketing time CDF if housing units had the same characteristics as homes in 2003 (DFL and Kaplan–Meier)

Year	Percentile						
	01	02	05	10	25	50	75
1997	1.0	1.7	6.3	15.0	47.6	130.4	326.3
1998	1.0	1.0	2.3	6.0	23.9	79.2	199.4
1999	1.0	1.0	1.0	1.8	7.2	31.6	94.2
2000	1.0	1.0	1.0	1.0	3.3	12.1	48.5
2001	1.0	1.0	1.0	1.1	3.5	13.9	57.3
2002	1.0	1.0	1.0	1.4	4.0	16.1	61.7
2003	1.0	1.0	1.0	1.7	4.6	11.7	37.8
2004	1.0	1.0	1.0	1.5	4.5	8.4	25.2
2005	1.0	1.0	1.0	3.1	6.3	20.2	87.2
2006	1.0	1.0	4.7	10.5	35.7	105.3	235.3
2007	1.0	1.0	4.8	10.2	37.6	127.1	294.1

Notes: First panel of table shows Kaplan–Meier estimates for the distribution of residential real estate listings' marketing time in Fairfax County, VA. Second panel uses the DiNardo, Fortin and Lemieux decomposition to simulate the distribution of marketing time in each corresponding year if housing units had the characteristics of units for sale in 2003. Covariates include all variables shown in Table 3.

1 day. Higher percentiles of the distribution, on the other hand, are notably more volatile. For instance, the 25th and 50th percentile of the distribution of time on the market in 1997 are about 10 and 11 times higher than their counterparts in 2003 suggesting that 1997 was a very "cold" housing market much like 2007.

In this longer time frame the concern remains that shifts in the *unconditional* distributions of marketing time could be at least partially explained by differences in the characteristics of the housing units that were traded in each of these periods. To alleviate this concern, we again use the DFL method to estimate the distribution of time on the market during each year in our sample assuming that housing units had the same characteristics as homes in 2003, the base year. We construct these counterfactual distributions following the methods described in Section 3 and present results at the bottom panel of Table 4. Results suggest that controlling for home characteristics introduces some changes in the distribution of marketing time. For instance, while the actual and counterfactual bottom percentiles of the distributions are virtually identical, there are some differences between the actual and counterfactual at the median and other upper percentiles. For example, the counterfactual median marketing time in 1997 (2007) is about 5 days higher (lower) than the actual median marketing time. Thus, it can be seen that over a 10 year period (rather than the 4 year period we looked at in Section 3), controlling for home characteristics becomes more important as housing characteristics have a longer time frame to evolve.

We also construct estimates for the hazard function and show results in Table 5. The top panel displays unconditional hazard functions, while the bottom panel computes counterfactual ones: hazard functions assuming that homes listed in each year in our sample

Table 5
Hazard function of marketing time in Fairfax County, VA.

Year	Days on the market							
	7	14	31	45	60	90	120	180
a) Unconditional hazard (%)								
1997	0.69	0.66	0.55	0.54	0.53	0.49	0.47	0.38
1998	1.36	1.08	0.81	0.76	0.72	0.68	0.63	0.53
1999	2.76	1.85	1.32	1.26	1.12	0.98	0.74	0.54
2000	4.76	2.63	1.73	1.64	1.54	1.14	0.76	0.48
2001	4.40	2.34	1.51	1.45	1.31	0.93	0.71	0.33
2002	4.24	2.25	1.47	1.46	1.37	0.95	0.78	0.37
2003	5.72	3.48	2.27	1.90	1.73	1.25	1.06	0.45
2004	7.75	4.53	2.66	2.03	1.66	1.35	1.34	0.53
2005	4.52	2.41	1.27	1.03	0.88	0.90	0.80	0.40
2006	0.87	0.80	0.66	0.60	0.58	0.55	0.56	0.50
2007	0.88	0.77	0.57	0.49	0.45	0.41	0.42	0.40
b) Conditional hazard (%): hazard function if housing units had the same characteristics as homes in 2003 (DFL)								
1997	0.65	0.63	0.53	0.52	0.52	0.49	0.46	0.39
1998	1.28	1.01	0.78	0.73	0.67	0.66	0.64	0.50
1999	2.72	1.84	1.31	1.22	1.12	1.04	0.76	0.54
2000	4.77	2.62	1.77	1.62	1.53	1.10	0.83	0.54
2001	4.43	2.37	1.52	1.46	1.34	0.93	0.74	0.32
2002	4.28	2.27	1.48	1.46	1.38	0.96	0.81	0.39
2003	5.72	3.48	2.27	1.90	1.73	1.25	1.06	0.45
2004	7.72	4.51	2.66	2.03	1.67	1.37	1.35	0.55
2005	4.56	2.44	1.30	1.06	0.90	0.90	0.81	0.41
2006	0.88	0.81	0.66	0.59	0.59	0.56	0.56	0.49
2007	0.91	0.78	0.58	0.51	0.46	0.40	0.44	0.37

Notes: First panel of Table shows estimates for the unconditional hazard function of residential real estate listings marketing time in Fairfax County, VA. Cumulative hazard functions were first computed using the Nelson–Aalen estimator. The hazard functions were estimated by computing the curvature of the smoothed cumulative hazards (a Gaussian kernel function was used to smooth the cumulative hazard function; the same bandwidth –3 days– was set in all periods). Second panel uses the DiNardo, Fortin and Lemieux (DFL) decomposition to simulate the hazard function of marketing time in each corresponding year if housing units had the characteristics of units for sale in 2003. Covariates include all variables shown in Table 3.

have the same characteristics as homes in 2003.[16] Note that as was the case with the marketing time distributions, there are small differences between the actual and the counterfactual hazard functions. Moreover, the hazard rate for those properties that have been on the market for 1 week exhibits substantial volatility over time: during our sample it ranges between 0.007 and 0.077. As time on the market increases, the hazard rate volatility diminishes. For instance, the likelihood of selling a home given that it has been on the market for 6 months lies between 0.003 and 0.005.

The counterfactual distributions and hazard functions can be used to construct "quality adjusted" indicators that measure the liquidity of the housing market over time. For example, one can compute the ratio between the counterfactual distributions (hazard functions) in each year of the sample and the distribution (hazard) in the base year. We have estimated this ratio for the first, fifth and fiftieth percentile of the distribution and show results in Fig. 6. Similarly, in Fig. 7 the hazard ratio has been estimated for properties that have been on the market for 7, 90 and 180 days. The trends illustrated in Figs. 6 and 7 provide several insights about how the liquidity of housing has evolved in Fairfax County. While there are no changes over time at the bottom of the distribution (first percentile), the median time on the market shows significant changes: it sharply decreased from 1997 to 2000, remains relatively constant between 2001 and 2004, and then increased back to 1997 levels by the end of 2007.

[16] Cumulative hazard functions were first computed using the Nelson–Aalen estimator. The hazard functions are estimated by computing the curvature of the smoothed cumulative hazards (a Gaussian kernel function was used to smooth the cumulative hazard function; the same bandwidth –3 days– was set in all periods). The DiNardo, Fortin and Lemieux (DFL) decomposition is used to estimate the counterfactual hazard function. Covariates include all variables shown in Table 3.

Fig. 6. Conditional quantile ratio = [counterfactual quantile (year t) / actual quantile 2003].

Fig. 7 shows that the probability that a home sells after staying 1 week on the market, has sharply changed over time reaching a peak in 2007. Thus once again it is clear that to measure the liquidity of the housing market it is important to look beyond the conditional mean and median and analyze the full distribution of time on the market.

4.1. Additional heterogeneity in the distribution of TOM

Investors and real estate practitioners may care about a particular subset of the housing market. For example, a focus on a specific type of home or on a particular location may be needed to assess liquidity risk. To test if the distribution of time on the market depends on the home's type and/or location, we use the Kaplan–Meier estimator to compute a non-parametric estimate of the *unconditional* distribution of time on the market for several subsets of homes. Figs. 8–11 show that the distribution of marketing time not only differs substantially by the type of home and by the three zip code locations that we

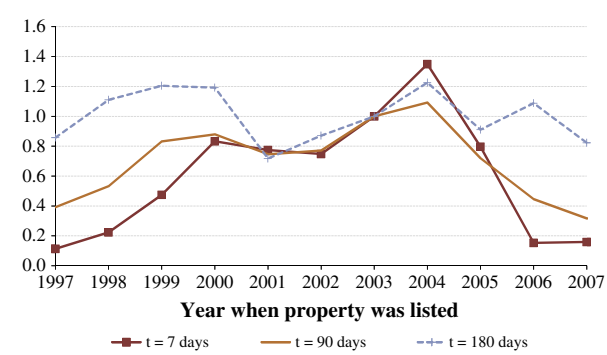

Fig. 7. Conditional hazard ratio = [counterfactual hazard (year t) / actual hazard 2003].

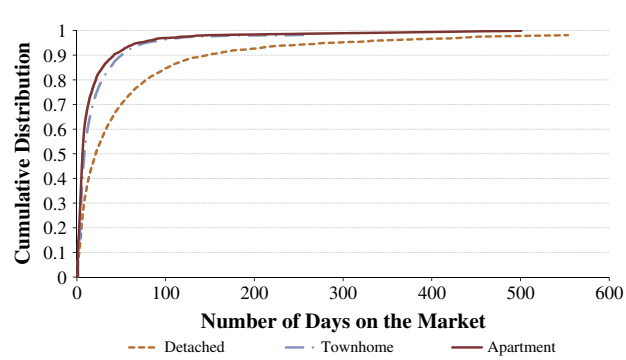

Fig. 8. Distribution of marketing time by type of home year 2003.

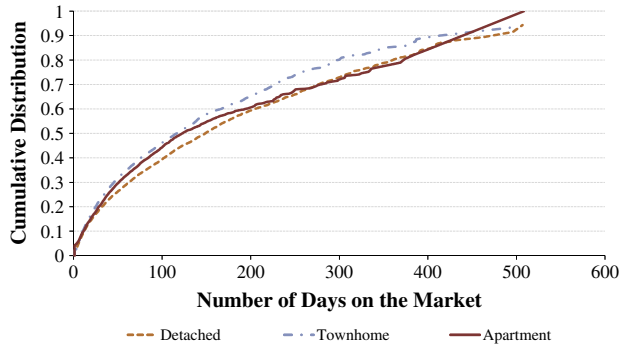

Fig. 9. Distribution of marketing time by type of home year 2007.

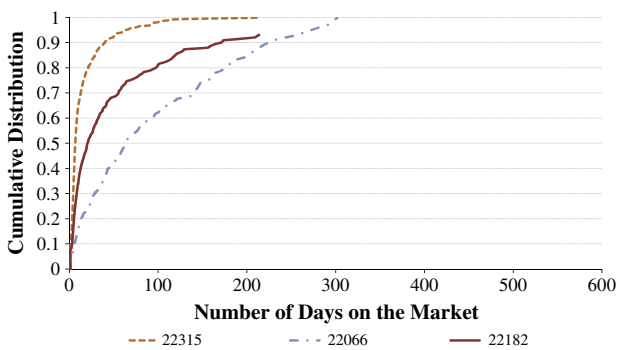

Fig. 10. Distribution of marketing time by zip code year 2003.

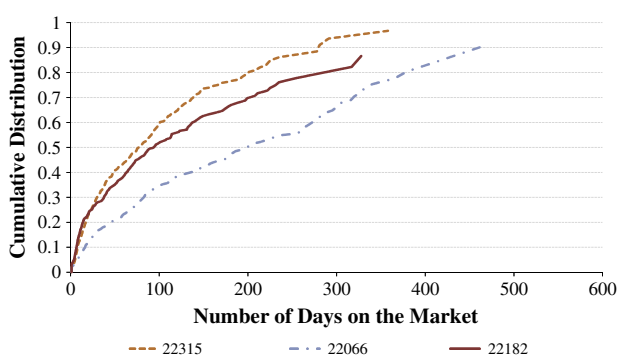

Fig. 11. Distribution of marketing time by zip code year 2007.

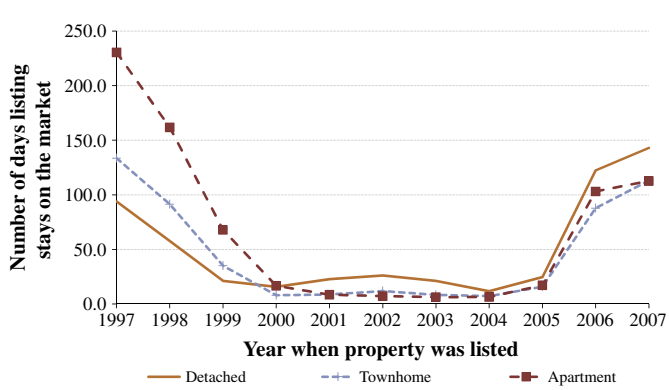

Fig. 12. Median time on market (TOM) by type of home keeping constant characteristics of housing units (as in 2003).

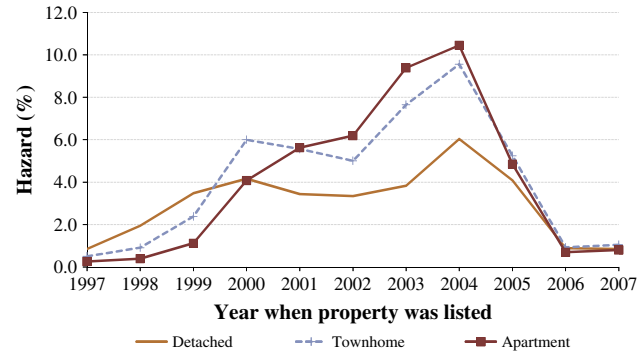

Fig. 13. Hazard Rate ($t = 7$ days) by type of home keeping constant characteristics of housing units (as in 2003).

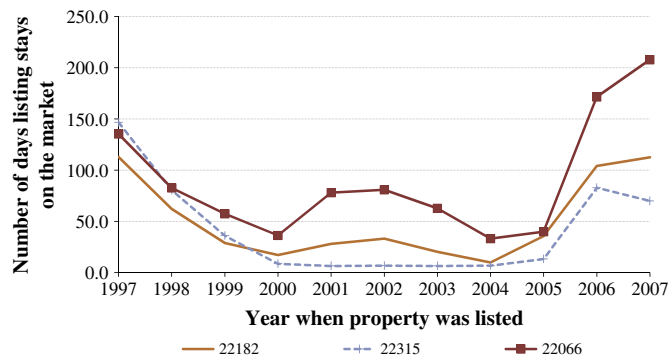

Fig. 14. Median marketing time by zip code keeping constant characteristics of housing units (as in 2003).

have chosen, but it also differs over time within each of these subsets of homes between 2003 and 2007.

The indices developed in this paper can be easily used to measure the liquidity of these subsets of housing types. To illustrate this, we have computed indicators like those shown in the bottom panels of Tables 4 and 5 for three separate types of homes: single family homes, townhomes, and apartments, and for all homes located in three specific zip codes within Fairfax County, VA. To keep our results parsimonious, we only show graphical results for the (quality adjusted) median marketing time and one week hazard rates. Results displayed in Figs. 12 and 13 suggest that the liquidity of housing depends upon the unit's type. For example, until 2000 the median time on the market of apartments and townhomes is much larger than for detached homes. Other interesting differences appear when analyzing hazard rate trends in Fig. 13. While trends are somewhat similar, it is clear that the variance in the hazard rate is much smaller for detached units. Results shown in Figs. 14 and 15 also suggest that marketing time depends upon the location of the units. For example, in most years housing units in zip code 22315 sell at much faster rates than units located in zip codes 22066 and 22182.

The distribution of time on the market has been computed for every year in our sample. To promptly assess housing market conditions, however, higher frequency data is often needed. The methods developed in this paper can be used to compute the (quality adjusted) distribution of time on the market during a period shorter than a year, such as a semester, a quarter or even a month. For example, to compute quality adjusted marketing time distributions and hazard rates for one quarter (month) in our sample, one needs to use all listings that were posted on the MLS between the first and last day of that quarter (month).[17] We have computed quality adjusted median

[17] To construct the DFL counterfactual, one needs to pick a "base-period", and apply the methods described above.

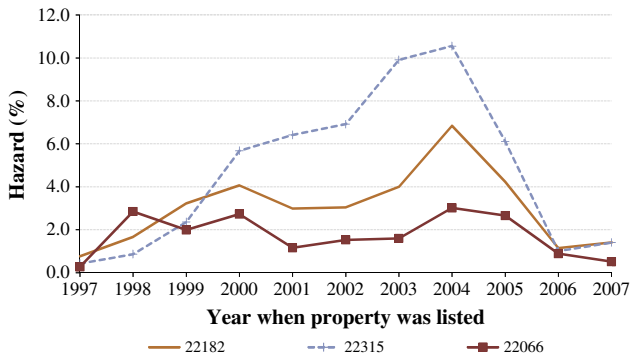

Fig. 15. Hazard rate ($t = 7$ days) by zip code keeping constant characteristics of housing units (as in 2003).

time on the market and one week hazard rates for each quarter in our full dataset (quarter 1, 1997 through quarter III, 2010) and show results in Fig. 16. While the trends are quite consistent with those found using the annual data, quarterly indices highlight important seasonal patterns in the liquidity of the housing market.

5. Conclusions

Housing is one of the most important markets in the U.S. and is a major source of wealth. Furthermore, it may be the single most important financial decision that the typical household in the U.S. makes. Banks and other financial institutions also invest heavily in residential real estate markets. Thus it is not surprising that significant efforts are made on a regular basis to measure the level and volatility of housing values. Much less attention is given, however, to the systematic measurement of this asset's liquidity. This is surprising given that liquidity is a key determinant of any asset's desirability and that high quality data that document the marketing process of real estate are often available.

This paper is the first attempt at analyzing how the *distribution* of time on the market changes over time. Our methodology takes into account changes in housing quality between time periods, and also adjusts for censored observations where houses are listed but have not been on the market long enough to sell. We use this methodology to describe how the distribution of time on the market evolved from a "hot" housing market in 2003 to a "cold" market in 2007 in a county within the Washington D.C. metropolitan area. Key finding from this analysis include: (i) focusing on only the mean or median masks considerable heterogeneity in the distribution of time on the market, (ii) changes in the distribution of time on the market between a "hot" and a "cold" market are much more complex than what is suggested by a mean or median, (iii) changes in housing characteristics over time explain little of the shift in the distribution, and (iv) there is substantial heterogeneity in the distribution of time on the market across property types and across property locations within the county.

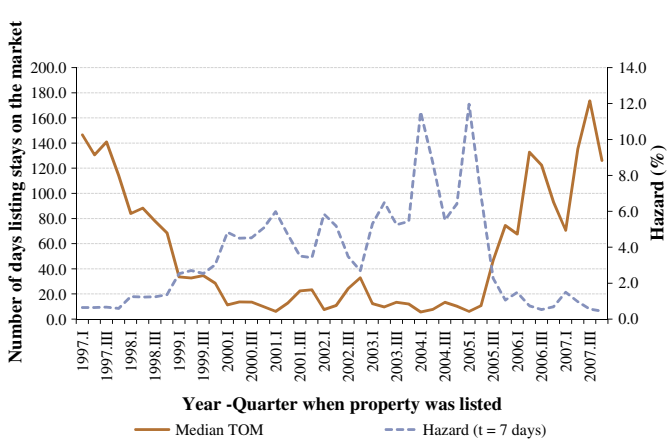

Fig. 16. Median marketing time and hazard rates by quarter keeping constant characteristics of housing units as in 2003.I.

As the recent housing boom and bust has shown there is significant risk involved in including housing assets in a financial portfolio. This risk stems not only from swings in housing prices, but also from changes in how quickly a house can be sold. Households, investors and financial institutions that purchase housing assets, value knowing when homes can be thought of as "hot or cold potatoes" so as to deal with liquidity risk. The methodology outlined in this paper can potentially be used by parties that invest in real estate to better understand how the marketing time of residential properties evolves over time. Furthermore, it might be used by policymakers to better understand the full costs of a boom and bust (beyond just housing price swings) or the impact that policies like the home buyer's tax credit had on marketing times.

References

Aalen, O., 1978. Nonparametric inference for a family of counting processes. The Annals of Statistics 6, 701–772.
Albrecht, J., Bjorklund, A., Vroman, S., 2003. Is there a glass ceiling in Sweden? Journal of Labor Economics 21 (1), 145–177.
Allen, M., Rutherford, R., Thomson, T., 2009. Residential asking rents and time on the market. Journal of Real Estate Finance and Economics 38 (4), 351–365.
Anglin, P., Rutherford, R., Springer, T., 2003. The trade-off between the selling price of residential properties and time-on-the-market: the impact of price setting. Journal of Real Estate Finance and Economics 26 (1), 95–111.
Carrillo, P. forthcoming. "An Empirical Stationary Equilibrium Search Model of the Housing Market." *International Economic Review*.
Carrillo, P., Yezer, A., 2009. Alternative measures of homeownership gaps across segregated neighborhoods. Regional Science and Urban Economics 39 (5), 542–552.
Cobb-Clark, D., Sinning, M., 2011. Neighborhood diversity and the appreciation of native- and immigrant-owned homes. Regional Science and Urban Economics 41 (3), 214–226.
Deng, Y., D. McMillen and T.F. Sing, 2011. "Private Residential price Indices in Singapore: A Matching Approach." forthcoming in *Regional Science and Urban Economics*
DiNardo, J., Fortin, N., Lemieux, T., 1996. Labor market institutions and the distribution of wages, 1973–1992: a semiparametric analysis. Econometrica 64, 1001–1044 (September).
Genesove, D., Mayer, C., 1997. Equity and time to sale in the real estate market. The American Economic Review 87, 255–269.
Genesove, D., Mayer, C., 2001. Loss aversion and seller behavior: evidence from the housing market. The Quarterly Journal of Economics 1233–1260.
Haurin, D., 1998. The duration of marketing time of residential housing. Real Estate Economics 16 (4), 396–410.
Haurin, D., Haurin, J., Nadauld, T., Sanders, A., 2010. List prices, sale prices and marketing time: an application to U.S. housing markets. Real Estate Economics 38 (4), 659–685.
Horowitz, J., 1992. The role of the list price in housing markets: theory and an econometric model. Journal of Applied Econometrics 2, 115–129.
Kang, H., Gardner, M., 1989. Selling price and marketing time in residential real estate market. The Journal of Real Estate Research 4, 21–35.
Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457–481.
Knight, J., 2002. Listing time, time on market, and ultimate selling price: causes and effects of listing price changes. Real Estate Economics 30 (2), 213–237.
Kuminoff, N. and J. Pope, 2011. "The Value of Residential Land and Improvements during the Great Housing Boom and Bust." mimeo.
Landvoigt, T., M. Piazzesi and M. Schneider, 2011."The Housing Market(s) of San Diego." mimeo.
Leibbrandt, M., Levinshon, J., McCrary, J., 2010. Incomes in South Africa after the Fall of Apartheid. Journal of Globalization and Development 1 (1).
Machado, J., Mata, J., 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. Journal of Applied Econometrics 20, 445–465.
McMillen, D., 2003. Neighborhood house price indexes in Chicago: a Fourier repeat sales approach. Journal of Economic Geography 3, 57–73.
McMillen, D., 2008. Changes in the distribution of house prices over time: structural characteristics, neighborhood or coefficients? Journal of Urban Economics 64, 573–589.
Nelson, W., 1972. Theory and applications of hazard plotting for censored failure data. Technometrics 14, 945–965.
Nguyen, B., Albrecht, J., Vroman, S., Westbrook, D., 2007. A quantile regression decomposition of urban–rural inequality in Vietnam. Journal of Development Economics 83 (2), 466–490.
Powell, J., Honoré, B., Khan, S., 2002. Quantile regression under random censoring. Journal of Econometrics 109, 67–105.
Springer, T., 1996. Single-family housing transactions: seller motivations, price, and marketing time. Journal of Real Estate Finance and Economics 13 (3), 237–254.
Yavas, A., Yang, S., 1995. The strategic role of listing price in marketing real estate: theory and evidence. Real Estate Economics 23, 347–368.