The Repeat Time-On-The-Market Index*

Paul E. Carrillo†
George Washington University

Benjamin Williams‡
George Washington University

January 2017

Abstract

We propose two new indices that measure the evolution of the time it takes a home to sell or "time-on-the-market" (TOM). The key features of both indices are a) their ability to control for unobserved heterogeneity exploiting *repeat* listings, b) their use of censored durations (listings that are expired and/or withdrawn from the market), and c) their computational simplicity. The first index computes proportional displacements in the home sale baseline hazard rate. The second estimates the relative change in median marketing time. The indices are computed using about 1.8 million listings in 15 US urban areas. Results suggest that both accounting for censoring and controlling for unobserved heterogeneity are key to measure housing market liquidity. Finally, by combining the TOM indices with repeat housing price indices, we develop a framework to describe the evolution of the *joint* distribution of prices and TOM.

Keywords: Housing liquidity, non-parametric models, proportional hazard, repeat-sales
JEL Codes: C41, R31

*We would like to thank John Clapp, Andra Ghent, Scott Wentland, Anthony Yezer and participants at the Homer Hoyt Institute Workshop, the Richmond Fed Regional Workshop, the 2016 ASSA-AREUEA Annual Conference, and at seminars at CoreLogic, Freddie Mac and George Washington for helpful comments and suggestions. A substantial portion of the empirical analysis was conducted using data from CoreLogic Solutions LLC provided under a CoreLogic Academic Research Council License Agreement. We are grateful to CoreLogic and to Metropolitan Regional Information Systems Inc. (MRIS) for sharing their data with us, and would also like to acknowledge financial support from the GWU Institute for International Economic Policy (IIEP).

†Email: pcarrill@gwu.edu.
‡Email: bdwilliams@gwu.edu.

1 Introduction

The home buying and selling process contains a number of frictions including information asymmetries, high transaction costs and costly matching between sellers of differentiated products and buyers with heterogenous tastes. Sellers can also influence the time their home is on the market by changing their asking and reservation price. These features of real estate markets have been considered by a long and growing literature in urban economics that employs search-and-matching models to study the microstructure of housing transactions [1]. In such settings the housing market clears in both prices and marketing time.

While it seems clear that housing markets clear in both of these dimensions, measurement of housing market conditions has been focused mostly on home prices. While a very large set of papers have proposed alternative methods to construct home price indices including approaches that use hedonic regressions and matching techniques, the classical approach to construct home price indices uses *repeat* sales to control for unobserved housing heterogeneity and estimate unbiased estimates of home price changes (Bailey et al., 1963; Case and Shiller, 1987).[2] Virtually all efforts made to measure the evolution of home prices focus on the *unconditional* distribution of prices and ignore the role of marketing time.[3]

In this paper we propose novel methods to estimate the evolution of the time it takes a home to sell or "time-on-the-market" (TOM). The indices we propose a) control for unobserved heterogeneity exploiting *repeat* listings and therefore can be directly compared with conventional and widely available repeat home price indices, b) use censored durations (listings that are expired and/or withdrawn from the market), and c) are computationally straightforward. We then use the repeat TOM indices as well as conventional repeat housing price indices to describe the *joint* distribution of prices and TOM.

Measuring the evolution of TOM in the housing market is important for several reasons. First of all, TOM is a market clearing mechanism and a key factor to assess housing liquidity risk (Lippman and McCall, 1986; Lin and Vandell, 2007). Changes in the liquidity of housing can affect its value and influence the optimal decisions of buyers, sellers and investors in the real estate

[1] Han and Strange (2015) present a comprehensive survey of this topic; examples of recent studies include Carrillo (2012) and Merlo et al. (2015)

[2] Eurostat (2013) presents a useful summary of these approaches.

[3] The exception is Fisher et al. (2003) who estimate a constant-liquidity commercial real estate price index.

market. Second, even crude measures of housing liquidity can substantially improve home price forecasts (Carrillo et al., 2015). The systematic production of measurements of housing liquidity that account for censoring and unobserved housing heterogeneity could be a valuable input to predict future housing market conditions. Third, despite the obvious importance of housing to local markets and the macroeconomy, there are no official indices that *rigorously* measure the evolution of TOM. Finally, because no rigourous measure of TOM is available for a panel of cities, the dynamic relationship between TOM and home prices remains largely unexplored. Our work takes a first step in this direction.

The lack of official measures of TOM (and housing liquidity) is probably not due to data constraints. In the U.S. and in many other developed countries, the marketing of real estate properties is centralized in a multiple listing system where sellers post their properties and their asking prices. This system typically records the date when a property is listed and the date when it is sold. This information allows one to compute the number of days that any home stays on the market. In fact, many real estate associations in the U.S. provide statistics such as median or mean TOM as part of their reports. However, these statistics fail to account for two important features – censoring of TOM due to the expiration or delisting of homes, and unobserved home heterogeneity. If all properties that were listed were to find a buyer (i.e., if there were no censored observations), the same conventional approach used to compute repeat sale price indices could be used to estimate a TOM index that accounts for unobserved heterogeneity.[4]

Expired and withdrawn listings are, however, a common feature of real estate markets. For instance, in a suburb of Washington DC (Fairfax County, VA) as much as 60 percent of listings expired and/or were withdrawn during the peak of the financial crisis. Similar patterns are found in San Diego, Las Vegas, Miami and 11 other MSAs in the U.S. that we analyze.[5] Descriptive statistics suggest that censoring drastically changes with market conditions: it remains low during housing booms and peaks during busts. Hence, conventional methods based on repeat sales need

[4]The classical approach to compute housing price indices uses repeat sales and a simple linear regression model to control for unobserved housing heterogeneity (Bailey et al., 1963). Influential extensions to this method are described in Case and Shiller (1987) and Case and Shiller (1989). The repeat sales approach has been used in recent studies to measure changes in income (Rosenthal, 2014), changes in rents (Ambrose et al., forthcoming) and changes in time-on-the market (Liu et al., 2016). Liu et al. (2016) construct a repeat time-on-the market index in their paper that is similar in spirit to ours but does not control for censoring.

[5]Our empirical analysis exploits individual residential real estate listing records in 15 separate US urban areas. The dataset contain about 1.8 million observations. Details about the sample are provided in Section 3.

to be modified to account for censoring. This is the main objective and contribution of our paper.

Intuitively, censoring of home listings can severely impact estimates of the mean or median TOM. TOM durations are only observed for homes where this duration is no larger than a censoring duration that represents the patience of the home seller. Not only does this censoring lead to a downward bias in estimates of mean or median TOM, but an increase in the distribution of TOM that is not matched by an increase in the distribution of sellers' patience will lead to a larger percentage of listings that expire or are withdrawn and hence more severe censoring. To illustrate empirically the effects of ignoring censoring issues when computing changes in TOM, we estimate (unconditional) median TOM with and without accounting for censoring in each of our regions of interest. Results strongly suggest that ignoring censoring leads to significantly different assessments about changes in the distribution of TOM. For example, the median TOM of properties that were listed in Fairfax County during 2007 and *sold* thereafter was about 60 days. This number is almost 3 times lower than the estimate of the median TOM of all 2007 listings when censored durations are included in the estimation. In all areas, the variance of median TOM of completed durations is much lower than the variance of median TOM when censored observations are accounted for. In our view, any method developed to produce a TOM index should incorporate censored observations.

Accounting for *observed* housing heterogeneity, such as home size, number of bathrooms, and other home characteristics in a duration model is relatively straightforward. Applying the reweighting procedure suggested by DiNardo et al. (1996), we estimate quality adjusted TOM distributions and show that, in Fairfax County, controlling for observed housing characteristics does not substantially change the estimate of the TOM distribution. Controlling for unobserved heterogeneity, however, has been shown to be important when estimating home price indices (Wallace and Meese, 1997) and may also be important when computing TOM indices. The unobserved heterogeneity in TOM, which is assumed to be constant over time, represents features of a home that make it more or less marketable.

We propose two models that can be used to correct for unobserved heterogeneity using *repeat* listings. The first index is based on a proportional hazard specification and applies an estimator developed by Ridder and Tunali (1989) to compute proportional displacements in the home sale baseline hazard rate. The second index is based on a novel procedure that accounts for censoring and unobserved heterogeneity in an accelerated failure time model of TOM; it estimates relative changes in (quality adjusted) median TOM over time. For both indices the main intuition is straightforward. Just as it is the case with repeat sales home price indices (Bailey et al., 1963), one can use the TOM of properties that have been on the market in more than one period to "difference out" the unobserved heterogeneity. In the presence of random censoring, however, the conventional methods employed to construct a repeat sale index no longer apply, and need to be adapted for this specific application.

Our first index amounts to a logit regression for whether the TOM corresponding to the second listing exceeded the TOM corresponding to the first listing.[6] We show that this approach provides a natural analogue to the repeat sales regression used to construct a home *price* index. The model underlying this construction assumes that the hazard rate evolves proportionately over time, and that a home's unobserved heterogeneity shifts its baseline hazard by the same amount in all time periods. This assumption essentially implies that unique features of a home make it always more (or less) likely to sell relative to the market's baseline hazard. This index controls for unobserved heterogeneity in a transparent manner, incorporating censored durations in the estimation process and estimating the "repeat proportional hazard index" using a computationally straightforward procedure. The repeat proportional hazard index (RPHI) provides an estimate of the gross percentage increase in the hazard rate relative to a base period.

The second index we propose in this paper is based on a log-linear specification for the TOM. This type of specification in a duration model is known as an accelerated failure time (AFT) model (see, e.g., Kalbfleisch and Prentice, 2011). The AFT specification models the median of (log) TOM as a function of period-specific shifters as well as a home-specific unobserved heterogeneity term, which is assumed to be time-invariant. The period-specific shifters are used to construct the "repeat median TOM index." The AFT model suggests that, absent censoring, we could simply estimate a fixed effects regression using the difference in the log TOM (i.e., a repeat sales regression). Instead we use the Kaplan-Meier procedure (Kaplan and Meier, 1958) to first estimate the difference in the median[7] log TOM between two listings *conditional on the home being listed in two particular periods*. That is for each pair of time periods, t and $t + k$, we select the set of repeat listings that were put on the market in period t for the first time and in period $t + k$ for the second time and,

[6] This logit regression is computed on a subsample of the homes. Conditioning on this subset of homes turns out to be innocuous. See Section 5.1 for details.

[7] We work with the median because the mean is typically not identified by the Kaplan Meier estimator.

using only this subset of observations, we estimate the median log TOM in each period using the Kaplan-Meier correction for censoring and take the difference. We repeat these calculations for every pair of periods in our sample. Then, in a second step, we run a simple OLS regression to estimate the "repeat median TOM index." The repeat median TOM index (RMTI) measures the relative change in the median TOM in the current period relative to a base period.

We compute TOM indices in each of the 15 MSAs we study. Results show that housing liquidity is subject to substantial variation over time. More importantly, we provide strong evidence that controlling for unobserved heterogeneity (i.e. using repeat listings) affects the estimate of TOM displacements. The indices described above provide an unbiased estimate of how the *unconditional* distribution of time on the market evolves over time.

Finally, we note that to fully describe the *joint* distribution of price and TOM $f(p, y) = f(p|y)f(y)$, in addition to the unconditional distribution of TOM $f(y)$, one needs to focus on the *conditional* distribution of prices given TOM $f(p|y)$. We estimate $f(p|y)$ and find some but small differences between the conditional and unconditional home price distributions. We then explore the relationship between the unconditional TOM and unconditional price indices. We show that both of our TOM indices anticipate changes in home prices and can hence improve home price forecasts in the short run. We also find a new stylized fact about these two series: TOM is cointegrated with home price *appreciation* but not with price levels. We hope that the TOM indices developed in this paper are useful in many other applications.

The rest of the paper is structured as follows. The next section discusses the relationship between our work and the econometrics literature. Section 3 describes the data and computed descriptive statistics. In Section 4, we discuss the censoring problem and investigate observable heterogeneity among homes. In Section 5 we develop our two new indices that incorporate censored observations as well and account for unobserved heterogeneity. Section 6 provides a framework to study the joint distribution of TOM and prices. In Section 7 we conclude.

2 Literature Review

In the introduction we discuss how our research relates to other studies in urban and real estate economics. In this section we clarify how our research fits within the econometrics literature. The indices developed in this paper build on an extensive literature on unobserved heterogeneity in duration models. Duration models with unobserved heterogeneity have been used in economics to model unemployment spells (Heckman and Borjas, 1980; Flinn and Heckman, 1982), auto accidents (Abbring et al., 2003), child mortality (Ridder and Tunali, 1989; Olsen and Wolpin, 1983), and brand-switching behavior (Gönül and Srinivasan, 1993), among other applications.[8] This literature traditionally focused on the distortions caused by unobserved heterogeneity when it takes the form of a random effect, independent of covariates (Lancaster, 1979; Heckman and Singer, 1984; Trussell and Richards, 1983; Heckman and Honoré, 1989). A classical random effect model in our application would assume that the distribution of unobserved housing heterogeneity does not vary over time. But it is precisely the potential for the quality of the homes listed to vary over time that we wish to control for in our analysis.

Instead the methods used in this paper assume a fixed effects model – where unobserved heterogeneity may be correlated with covariates. Application of a standard fixed effects regression (which would amount to a repeat sales regression in our context) is not possible because of the presence of random censoring. There are two solutions proposed in the literature that are relevant for our model. Ridder and Tunali (1999) propose a stratified partial likelihood in a proportional hazards model.[9] Their method has been applied in studying child mortality with family fixed effects (Ridder and Tunali, 1999) and to study spatial differences in unemployment duration using location fixed effects (Gobillon et al., 2011). As shown by Lancaster (2000), the stratified partial likelihood approach amounts to a logit regression for whether the TOM corresponding to the second listing exceeded the TOM corresponding to the first listing. We extend this insight by showing that in our model of repeat listings the method is equivalent to a repeat sales logit regression. This is a novel contribution of our paper, which allows the estimation of the repeat proportional hazard index.

Honoré et al. (2002) suggest adapting a method for fixed censoring (Honoré, 1992) by integrating over the distribution of the censoring variable. Both papers use an accelerated failure time model as an alternative to the proportional hazards model. Their approach however requires that the censoring variable be independent of all covariates. The implication of this assumption in our

[8] Duration models have been used to study *observed* heterogeneity in housing time on the market (see, e.g., Haurin, 1988; Glower et al., 1998).

[9] Ridder and Tunali (1999) extend an idea also discussed in Kalbfleisch and Prentice (1980), Chamberlain (1985) and Lancaster (2000).

context would be that the distribution of sellers' patience does not vary over time, which is not a reasonable assumption. We avoid this because the only covariates in our model are indicators for the time period in which each listing occurred. We plug in conditional Kaplan-Meier estimates rather than a single unconditional estimate. After an initial step that estimates the conditional median log TOM for each listing pair we employ the standard repeat sales regression. Lindgren (1997) also takes a conditional Kaplan-Meier approach but not in a repeated duration model and does not allow for unobserved heterogeneity. The present paper is apparently the first to apply such an approach to a fixed effects model.

The methods developed in this paper can be readily applied to estimate operational measures of housing liquidity in the literature (for example, Lippman and McCall, 1986; Lin and Vandell, 2007).

3 Data and Conventional Descriptive Statistics

In most developed countries, including the U.S., real estate agents collect rich information about the marketing process of housing sales. The data are collected in a database system known in the U.S. as Multiple Listing Services (MLS). These data contain details about each listing and each transaction. Besides the asking price, sale price and home characteristics, the specific dates when the listing was posted and when the home was sold (or when the listing was withdrawn from the market) are generally available. This allows researchers to compute the time that a property stays on the market (time-on-the-market TOM).

Our data come from two sources. Metropolitan and Regional Information Systems (MRIS) provided us with MLS data from Fairfax County, VA.[10] Data contain information for all housing listings in this county that were listed on the MLS between January 1, 1997 and December 31, 2010. Fairfax MLS data contain pricing, TOM as well as detailed characteristics about the properties such as the number of rooms, bathrooms, age, type of home and address. Because the location of each property is observed, one can compute aggregate statistics at any level of geographic aggregation. More importantly, we can track if the same property is listed and/or sold in multiple periods.

[10] Fairfax County is part of the Washington, D.C. metropolitan statistical area and is located in northern Virginia. According to the 2010 U.S. Census, Fairfax hosts more than one million residents and over 380,000 housing units. Fairfax also ranks as one of the richest and best-educated counties in the U.S.

Our second source of data is CoreLogic Solutions, LLC (CoreLogic). CoreLogic collects MLS data from more than 100 MSAs, verifies the consistency of the information and produces a series of indicators (available in its Real Estate Analytics Suite). Collecting MLS data from different U.S. regions is not easy. Besides legal agreements with each MLS regional association, a careful data validation process is needed because there are no set guidelines about database structure (variable names, etc.). CoreLogic provides this service. CoreLogic allowed us to work with their individual housing listings in 14 MSAs that were posted on the MLS between January 2004 and February 2013. The MSAs in our sample include large and medium urban areas in the East, Middle, and Western regions of the U.S.[11] CoreLogic data include information about pricing and the specific dates when the listing entered and exited the market. While we do not observe any of the property's characteristics, the data contain a unique property identifier. This allows us to track listings/sales of the same property over time.

We exclude from our two samples listings with unusually high or unusually low listing prices (top and bottom 1 percent during each year), observations that stayed on the market for more than two years, and observations with missing data. After this cleaning process, we are left with about 0.3 million listings in Fairfax County, and 1.4 million listings in the sample of 14 U.S. MSAs. A list of the urban areas, the number of listings, and a description of the sample period is available in Table 1. About 58 percent of all listings in the overall sample end up in a sale; the other listings either expire or are withdrawn from the market. Many properties are listed on the MLS more than once. We call these cases *repeat* listings and note that there are almost 1 million of them.

Before we present descriptive statistics, we need to discuss how time-on-the-market is defined. Both data sets include the date when a listing is first posted on the MLS, the date when the property was taken off the market (when the contract was signed) as well as the date when the transaction was closed (which is typically between 4 and 12 weeks after the contract agreement). We define TOM as the difference between the listing date and the contract date.

We first show *conventional* descriptive statistics, that is, the kind of statistics that are typically computed and published by MLS associations. These include the mean and median TOM as well

[11] The MSAs we analyze include Ann Arbor, MI, Boulder, CO, Durham, NC, Honolulu, HI, Las Vegas-Paradise, NV, Medford, OR, Miami-Miami Beach-Kendall, FL, New Orleans-Metairie-Kenner, LA, Olympia, WA, San Diego-Carlsbad-San Marcos, CA, San Luis Obispo-Paso Robles, CA, Santa Barbara-Santa Maria, CA, Toledo, OH and Youngstown-Warren-Boardman, OH-PA.

as the volume of sales. Importantly, the mean and median TOM are calculated just for the sample of properties that have been sold.[12] The top panels of Figures 1 and 2 present these statistics for Fairfax County, and five representative MSAs from our U.S. areas: Las Vegas, San Diego, Miami, Honolulu and New Orleans.[13] The statistics have been computed for each quarter to illustrate within-year seasonality. The swings in expected duration in Fairfax County clearly coincide with the housing market boom and bust. Expected TOM in Fairfax decreased drastically in the late 1990s and remained rather low until the end of 2005. It increased back to the 1990's levels in 2007 and slowly dropped thereafter. What it is surprising is that, even during the midst of the financial crisis expected TOM is only about two and a half months, and median duration does not exceed 60 days. The other areas show similar patterns. For example, median TOM in San Diego increases about three times between the first quarter of 2004 and 2010, but is never less than 120 days; in Miami, median TOM after 2007 remains low (less than 4 months) and exhibits a decreasing trend; and, expected duration in Honolulu peaks in 2009 but is never above 5 months.

While interesting, the patterns shown in the top panels of Figures 1 and 2 can be misleading. The distribution of TOM of properties that are sold may be quite different than the unconditional distribution. The bottom panels in these figures display the share of listings that are withdrawn and/or expired in each of these areas. When the market is strong, most listings find a buyer. When the market slows down, however, a higher fraction of listings are withdrawn form the market. For example, in Fairfax County, while about 90 percent of properties listed in 2003 found a buyer, over half of properties listed in 2006 were withdrawn. Similar patterns are found in Las Vegas, San Diego and other areas we analyze. In the next section, we compute TOM statistics using information from both listings that were sold and listings that were withdrawn and/or expired.

[12]Real estate agents associations typically compute the average TOM of units that were sold during a particular period t. This statistic measures the mean TOM of the subset of listings that were sold during t and listed at any point before t. This statistic conveys information about current and past liquidity and cannot always identify spikes in current demand. Suppose, for example, that in previous quarters it was difficult to sell a home but demand has spiked and homes are much easier to sell today. Then among the homes that sell today will be homes that have sat on the market for a long time waiting for the market to pick up. Thus this conventional approach would not identify the sharp decrease in TOM of homes listed today.

[13]Descriptive statistics for other areas are available upon request.

4 Censoring and Observed Heterogeneity

In this section, we consider a model with only observed heterogeneity in time on the market. This allows us to introduce a formal setup for our problem that will be used in all sections hereafter. We first use the Kaplan-Meier estimator to compute the unconditional distribution of TOM. Then we estimate quality adjusted TOM statistics, where the characteristics of the housing stock are assumed to remain fixed at those from the properties listed in a base period.

4.1 Basic setup

Let Y_i denote the time home i would spend on the market between the list date and the contract date if it were not removed from the market before it went under contract. Let t_i denote the listing period. Our goal is to measure changes over time in time on the market. In this section we look specifically at how the distribution of $Y_i \mid t_i = t$ varies with t. We observe a sample of listings $\{Y_i\}_{i=1}^n$. Sometimes listings are removed without the home being sold. Indeed, as we noted in the previous section, during some periods, the fraction of listings that do not terminate in a sale can exceed 50%. To model this let C_i denote the censoring time for the i^{th} listing, that is, the length of time after which the home will be withdrawn or will expire if it has not been sold. Then a listing ends in a sale if and only if $Y_i \leq C_i$, and it otherwise is withdrawn or expires. Thus for each listing we observe $V_i = \min\{Y_i, C_i\}$ and $d_i = \mathbf{1}(Y_i < C_i)$, but not (Y_i, C_i).

4.2 Unconditional TOM

We first consider estimation of the unconditional distribution of $Y_i \mid t_i = t$ for each $t = 1, \ldots, T$, that is, not conditional on home or seller characteristics. The main difficulty is that some listings are censored and hence we only observe (V_i, d_i, t_i). Censoring, if not addressed, introduces a downward bias in median time on the market measures – homes that are sold before being removed from the market are less likely to be in the right tail of homes that take a long time to sell.[14] Both the median of the distribution of $V_i \mid d_i = 1$ and the median of the distribution of V_i will be biased downward relative to the median of Y_m.

[14]Here we refer to the bias relative to the true time on the market among the population of listings that appeared in the MLS. As pointed out by a referee, the bias is not necessarily downward if the relevant population includes so-called "pocket listings" that sell very quickly without ever being listed in the MLS. If the relevant population is all home (all potential listings) then, again, the direction of the bias is unclear.

To adjust for censoring we take the conventional approach of applying the Kaplan-Meier estimator. This can be viewed as a reweighting of the data based on the distribution of censored observations (Efron, 1967). The procedure effectively splits each censored observation into two partial observations at C_i and $+\infty$ each receiving weights according to the probability that the censored observation is above or below each given quantile.

Formally, if C_i is independent of Y_i conditional on t_i then

$$Pr(V_i = y, d_i = 1 \mid V_i \geq y, t_i) = Pr(Y_i = y \mid Y_i \geq y, t_i)$$

The left-hand side of this equation is observed in the data and the right-hand side is the hazard function for the distribution of $Y_i \mid t_i$, which we denote by $h_{Y_i|t_i}(y \mid t)$. The survival function is identified from the hazard function since $S_{Y_i|t_i}(y \mid t) = \exp(-\int_0^y h_{Y_m|t_m}(s \mid t)ds)$. The Kaplan-Meier estimator of the survival function is based on this argument. With an estimate of the survival function, an estimate of the median can be obtained as the median is the value at which the survival function hits 0.5. That is, $Med(Y_i \mid t_i = t)$ solves the equation $S_{Y_i|t_i}(Med(Y_i \mid t_i = t) \mid t) = 0.5$.

As we previously discussed in Section 3, for units that are sold, we compute $V_i = Y_i$ as the difference between the date when an offer was accepted and the date when the listing was first posted. When a listing is withdrawn from the market or it expires without a sale, we compute $V_i = C_i$ as the time between when the listing was first posted and when it was withdrawn or expired. Both are measured in days. For each sample, we calculated the Kaplan-Meier estimate of the survival function separately based on the quarter in which the listing was first posted. From the survival function estimates we can then estimate the medians. In the top panels of Figures 3 and 4 these estimates are compared to the median TOM among the homes that sold. Results strongly confirm our priors that accounting for censoring drastically affects the estimate of TOM statistics. For example, the median TOM in Fairfax County in 2007 is close to 6 months, about three times larger than the *conventional* estimate. In all other areas, the estimate of median TOM increases substantially when withdrawn listings are accounted for.

The Kaplan-Meier approach described above assumes only that censoring is independent. Another way to adjust for independent censoring is by using the Cox (1972) partial likelihood estimator. This provides estimates of proportional shifts over time in the hazard function, rather than changes over time in the median TOM, under a proportional hazard assumption. The Cox proportional hazard model makes no assumptions about the functional form of the baseline hazard and its coefficients can be easily used to compute displacement in the baseline hazard over time.

For each urban area in our sample, we estimate a Cox hazard model. The dependent variable is TOM and the covariates are dummy variables for each time period (quarter) in the sample. The coefficients in the Cox regression are estimates of the log of the hazard ratio between each period t and the base period (first quarter of 2010). A hazard ratio of 1.5 in period t would imply that the probability that a homeowner sells her home on a given day, given that the home is still on the market, is 50 percent higher than in the base period. We estimate the model both using only finished durations, and also incorporating censored observations. Results shown in the bottom panels of Figures 3 and 4 confirm that accounting for censoring can substantially change our assessment about the evolution of housing liquidity. For example, in Fairfax County, accounting for censoring leads to a less volatile estimate of the hazard ratio; and, in most other areas, there are significant differences between the two hazard ratio estimates.[15]

In sum, any statistic that measures the evolution of TOM should account for censored observations.

4.3 Conditional TOM

In the previous section no attempt was made to control for housing heterogeneity. The set of homes listed in one period may be much different than the set of homes listed in another period. Differences over time in TOM could reflect changes in the composition of homes for sale rather than changes in market conditions. This is the same concern that motivates the use of hedonic and repeat-sales methods to estimate housing price indices. In this section, we follow the methods proposed by Carrillo and Pope (2012) to estimate changes in median TOM and shifts in the hazard rate while controlling for *observed* housing heterogeneity.

Carrillo and Pope (2012) show how to compute (quality adjusted) time on the market distributions and hazard functions using MLS data. In particular, the duration distribution and hazard

[15]Notice that in all areas there is an increase in the unadjusted hazard ratio during the last sample period. This is mechanically produced by the sampling procedure. Data were collected at the end of the sample period. Hence, properties that are listed and sold during the last quarter in the sample (recent listings) must have been sold quickly. These patterns are corrected when censored observations are accounted for.

function during each period is simulated assuming that housing units have the same characteristics as homes in a base period. The simulation is based on the decomposition methods proposed by DiNardo et al. (1996) and the Kaplan-Meier estimator (Kaplan and Meier, 1958). This method permits estimation of the distribution of TOM while controlling both for censoring and observed heterogeneity. To keep our exposition self-contained, technical details of the method are provided in an appendix.

For each period t in Fairfax County, we simulate the distribution of time-on-the-market assuming that the characteristics of housing units remain constant as those prevalent in the first quarter of 2000 (the base period).[16] We denote this counterfactual distribution as $\hat{F}_t$. We then estimate the counterfactual median TOM in each period. Results are shown in Figure 5. The dashed and solid lines display the median and counterfactual median, respectively. Controlling for observed heterogeneity does not substantially change our estimate of the median TOM. This is not surprising since previous studies have found that housing characteristics do not explain much of the variation of housing marketing time.[17]

5 Censoring and Unobserved Heterogeneity

We have shown so far that accounting for censoring affects TOM statistics by a large amount, while controlling for observed housing characteristics does not. In this section, we estimate changes in TOM while explicitly controlling for *unobserved* heterogeneity. As we mentioned before, there is an extensive literature on unobserved heterogeneity in duration models where it is generally assumed that the unobserved heterogeneity is a random effect (Heckman and Singer, 1984; Trussell and Richards, 1983; Heckman and Honoré, 1989). This approach is not appealing in our context for the following reasons. First, if we model unobserved heterogeneity as a random effect we would implicitly assume that its distribution does not vary over time, a strong assumption in the context of the housing market. Second, estimation results could be affected by the modeling choices (for example, parametric vs. non parametric specifications of the unobserved heterogeneity). Finally, these methods tend to be computationally intensive and may not work when the number

[16] The characteristics included in the model are the home's age (9 categories), number of bathrooms, bedrooms and indicators for the home's type.

[17] Typically, the coefficient of determination in log-linear TOM regressions is typically very low (for example, Levitt and Syverson, 2008), and structural models have a hard time predicting TOM (Horowitz, 1992; Carrillo, 2012).

of observations is very large (hundreds of thousands of observations). Our goal in this section is to propose housing liquidity indices that control for unobserved heterogeneity, account for censored durations and, more importantly, are computationally easy to implement.

We propose below two models that can be used to correct for unobserved heterogeneity using *repeat* listings. The main intuition is straightforward: Just as it is the case with repeat sales home price indices (Bailey et al., 1963; Case and Shiller, 1987), one can use the TOM of properties that have been on the market in more than one period to "difference out" the unobserved heterogeneity. This is precisely what Liu et al. (2016) do to compute a TOM index that focuses on completed transactions. Because TOM is subject to random censoring, however, the conventional methods employed to construct a repeat sale index no longer apply, and need to be adapted for this specific application. For $s = 1, 2$, let Y_i^s, C_i^s, and t_i^s denote the TOM, censoring time, and listing period for the s^{th} listing of home i and let $V_i^s = \min\{Y_i^s, C_i^s\}$ and $d_i^s = \mathbf{1}(Y_i^s \leq C_i^s)$.[18]

5.1 The Repeat Proportional Hazard Index

A mixed proportional hazards model for the duration of time on the market is defined by the following hazard function for home i if listed in period t.

$$\lambda_{it}(y) = \exp(\beta_t)\lambda_{0i}(y) \tag{5.1}$$

The first factor, $\exp(\beta_t)$, allows the hazard function to vary proportionately depending on the time period listed and the second factor, $\lambda_{0i}(\cdot)$, is a home-specific baseline hazard function. If we normalize $\beta_{t_0} = 0$ for a baseline period t_0 then $\lambda_{0i}(y)$ is the hazard function for home i if it is placed on the market in this baseline period. Thus, according to this model, fluctuations in the housing market contribute to variation in time on the market through proportional shifts in the hazard functions.

The baseline hazard function is allowed to vary across homes. This represents unobserved heterogeneity in homes. If Y_{it} is observed for every (i, t) pair then this is a standard mixed proportional hazards model and can be consistently estimated on pooled data using the Cox (1972) partial likelihood estimator, as discussed briefly in Section 4.2. If $\lambda_{0i}(\cdot)$ and t_i^s are independent then

[18] We observe more than two listings for a small fraction of the homes in our data. The methods discussed here can be extended to allow this but we do not do this here to avoid the cumbersome notation this would entail.

this estimator is also valid under our sampling scheme where we observe home i only in periods t_i^1 and t_i^2. This is because in that case the distribution of $Y_i^s \mid t_i^s = t$ is characterized by the hazard function λ_{it}. If $\lambda_{0i}(\cdot)$ and t_i are correlated then the distribution of $Y_i^s \mid t_i^s = t$ is distorted by selection effects. These results are all analogous to similar well-known results for a linear panel data model with fixed effects. In this section we employ an extension of Cox (1972) due to Ridder and Tunali (1989, 1999) that uses variation among separate listings for the same home to eliminate the fixed effect, $\lambda_{0i}(\cdot)$. This motivates a repeat sales logit regression method that we use to construct a repeat proportional hazard index.

The structure of the proportional hazards model allows us to avoid bias due to censoring and simultaneously difference out the unobserved heterogeneity using repeat listings of the same home. The hazard function in equation (5.1) is the hazard corresponding to the conditional distribution $Y_i^s \mid t_i^s = t, \lambda_{0i}(\cdot)$. Let $\Lambda_{0i}(y) := \int_0^y \lambda_{0i}(s)ds$ denote the baseline integrated hazard function. It follows then that

$$-\log(\Lambda_{0i}(Y_i^s)) = \beta_{t_i^s} + \varepsilon_i^s$$

with $-\varepsilon_i^s \mid t_i^s, \lambda_{0i}(\cdot) \sim EV1(0,1)$ where $EV1$ represents the type one extreme value distribution. This is a standard result for the proportional hazard model (see, e.g., Kalbfleisch and Prentice, 2011).[19] If the unobserved heterogeneity can be written as $\lambda_{0i}(y) = \exp(\alpha_i)\lambda_0(y)$ then we have $-\log(\Lambda_0(Y_i^s)) = \beta_{t_i^s} + \alpha_i + \varepsilon_i^s$, which suggests differencing as a solution. However, even in this case, the baseline hazard function is generally not known so a standard differencing strategy is not feasible. Instead we obtain identifying information from observing which listing was on the market longer since $Y_i^2 \geq Y_i^1$ if and only if $\log(\Lambda_{0i}(Y_i^2)) \geq \log(\Lambda_{0i}(Y_i^1))$ and the probability of the latter event is

$$\begin{aligned} Pr\left(\log(\Lambda_{0i}(Y_i^2)) \geq \log(\Lambda_{0i}(Y_i^1)) \mid t_i^1, t_i^2\right) &= Pr(\varepsilon_i^2 - \varepsilon_i^1 \leq \beta_{t_i^1} - \beta_{t_i^2} \mid t_i^1, t_i^2) \\ &= \frac{\exp(\beta_{t_i^1})}{\exp(\beta_{t_i^2}) + \exp(\beta_{t_i^1})} \end{aligned}$$

[19] First, the proportional hazard model implies that $\log(\Lambda_{it}) = \beta_t + \log(\Lambda_{0i})$. The result then follows from the following fact about cumulative hazard functions. For a random variable Z with density f_Z and cumulative distribution function F_Z, the hazard rate is $h_Z(z) = \frac{f_Z(z)}{1-F_Z(z)}$. Integrating both sides, we find that $\Lambda_Z(z) = -\log(1 - F_Z(z))$ where Λ_Z represents the cumulative hazard function. Then, for any $\ell > 0$, $Pr(-\log(\Lambda(Z)) \leq \ell) = Pr(F_Z(Z) \leq e^{-e^{-\ell}}) = e^{-e^{-\ell}}$, where the second equality is the probability integral transform.

since the difference of two type 1 extreme value distributions has a logistic distribution.

However, we cannot always determine whether $Y_i^2 \geq Y_i^1$ because some listings are censored. We start by showing that we can make our problem equivalent to one with common censoring times by restricting our sample. Let $\tilde{C}_i = \min\{C_i^1, C_i^2\}$ be the common censoring time for listing i, let $\tilde{V}_i^s = \min\{Y_i^s, \tilde{C}_i\}$ for $s = 1, 2$, and let $\tilde{d}_i^s = \mathbf{1}(Y_i^s \leq \tilde{C}_i)$ for $s = 1, 2$. Let W_i be a dummy indicating homes where either (a) both listings are uncensored or (b) one listing is censored and this listing's censoring time exceeds the observed time on the market for the other listing. Then

$$Pr(V_i^2 \geq V_i^1 \mid W_i = 1, t_i^1, t_i^2) = Pr(\tilde{V}_i^2 \geq \tilde{V}_i^1 \mid \tilde{d}_i^1 + \tilde{d}_i^2 > 0, t_i^1, t_i^2)$$

The left-hand side is what we observe in the data. The last step is to show that the right-hand side is equal to $\frac{\exp(\beta_{t_i^1})}{\exp(\beta_{t_i^2})+\exp(\beta_{t_i^1})}$. First, among homes with $\tilde{d}_i^1 + \tilde{d}_i^2 > 0$ the condition $\tilde{V}_i^2 \geq \tilde{V}_i^1$ is equivalent to $Y_i^2 \geq Y_i^1$. This is because in this case either $\tilde{C}_i$ is between Y_i^1 and Y_i^2, in which case the censored listing must be the one with a longer (potential) time on the market, or $\tilde{C}_i$ is larger than both durations, meaning that neither is censored. Second, with a common censoring time, removing the homes for which both listings were censored does not cause a sample selection bias.

We will now demonstrate this second fact. Under the independent censoring condition in Assumption 5.1 below,

$$\begin{aligned}
&Pr(Y_i^2 \geq Y_i^1 \mid \tilde{C}_i \geq \min\{Y_i^1, Y_i^2\}, t_i^1, t_i^2, \lambda_{0i}) \\
&= \frac{Pr(Y_i^2 \geq Y_i^1, \tilde{C}_i \geq Y_i^1, t_i^1, t_i^2, \lambda_{0i})}{Pr(Y_i^2 \geq Y_i^1, \tilde{C}_i \geq Y_i^1, t_i^1, t_i^2, \lambda_{0i}) + Pr(Y_i^1 \geq Y_i^2, \tilde{C}_i \geq Y_i^2, t_i^1, t_i^2, \lambda_{0i})} \\
&= \frac{\int (1 - F_2(y_1)) F_{\tilde{C}}(y_1) f_1(y_1) dy_1}{\int (1 - F_2(y_1)) F_{\tilde{C}}(y_1) f_1(y_1) dy_1 + \int (1 - F_1(y_2)) F_{\tilde{C}}(y_2) f_2(y_2) dy_2}
\end{aligned}$$

where F_1, F_2, and $F_{\tilde{C}}$ are short-hand notation for the distribution functions for $Y_i^1 \mid t_i^1, t_i^2, \lambda_{0i}$, $Y_i^2 \mid t_i^1, t_i^2, \lambda_{0i}$, and $\tilde{C}_i \mid t_i^1, t_i^2, \lambda_{0i}$, and f_1, f_2 are the corresponding density functions. Because of the proportional hazard assumption, $f_1 = \exp(\beta_1)\lambda_{0i}(1 - F_1)$ and $f_2 = \exp(\beta_2)\lambda_{0i}(1 - F_2)$. Plugging this in above, we can cancel out the common factor, $\int (1 - F_1(y))(1 - F_2(y)) F_{\tilde{C}}(y) \lambda_{0i}(y) dy$. Since the resulting formula does not depend on λ_{0i} we can conclude that

$$\begin{aligned} Pr(Y_i^2 \geq Y_i^1 \mid \tilde{C}_i \geq \min\{Y_i^1, Y_i^2\}, t_i^1, t_i^2) &= Pr(Y_i^2 \geq Y_i^1 \mid \tilde{C}_i \geq \min\{Y_i^1, Y_i^2\}, t_i^1, t_i^2, \lambda_{0i}) \\ &= \frac{\exp(\beta_{t_i^1})}{\exp(\beta_{t_i^2}) + \exp(\beta_{t_i^1})} \end{aligned}$$

Our final result is then

$$Pr(V_i^2 \geq V_i^1 \mid W_i = 1, t_i^1, t_i^2) = \frac{\exp(\beta_{t_i^1})}{\exp(\beta_{t_i^2}) + \exp(\beta_{t_i^1})}$$

This suggests estimation based on the conditional likelihood

$$\sum_{i=1}^{n} W_i \mathbf{1}(V_i^2 > V_i^1) \log\left(\frac{\exp(\beta' X_i)}{1 + \exp(\beta' X_i)}\right) + W_i \mathbf{1}(V_i^1 > V_i^2) \log\left(\frac{1}{1 + \exp(\beta' X_i)}\right)$$

where X_i is a vector of dummy variables, X_{it} for $t = 2, \ldots, T$, where $X_{it} = 1$ if $t_i^2 = t$, $X_{it} = -1$ if $t_i^1 = t$ and $X_{it} = 0$ otherwise and W_i is equal to 1 if neither duration is censored or if only the smaller of the two durations is censored, and is equal to 0 otherwise. This is the likelihood function for the logit regression of the binary indicator $\mathbf{1}(V_i^2 > V_i^1)$ on X_i *on the subsample of observations with* $W_i = 1$.

Before detailing how the implentation we state and discuss the formal assumption required for the above derivation. These conditions are simplified versions of the assumptions used in Ridder and Tunali (1999) for a more general model.

Assumption 5.1.

For each pair of listings, s and $s+1$,

(i) $Y_i^1 \mid t_i^1, t_i^2, \lambda_{0i}(\cdot) =_d Y_{it_i^1} \mid \lambda_{0i}(\cdot)$ *and* $Y_i^2 \mid t_i^1, t_i^2, \lambda_{0i}(\cdot) =_d Y_{it_i^2} \mid \lambda_{0i}(\cdot)$

(ii) $Y_i^1 \perp\!\!\!\perp Y_i^2$ *conditional on* $t_i^1, t_i^2, \lambda_{0i}(\cdot)$

(iii) $(Y_i^1, Y_i^2) \perp\!\!\!\perp (C_i^1, C_i^2)$ *conditional on* $t_i^1, t_i^2, \lambda_{0i}(\cdot)$

The first condition is what is known as a strict exogeneity condition in panel data models. This ensures that the conditional distribution, $Y_i^1 \mid t_i^1, t_i^2, \lambda_{0i}(\cdot)$, is characterized by the hazard function on the left hand side of equation (5.1). This assumption requires that the selection of the listing period is not related to other unobservable factors not captured by the home-specific baseline hazard, λ_{0i}. So it rules out, for example, seller-specific fixed effects as determinants of the listing date. It also rules out dynamics where the t_i^1 affects Y_i^2 or Y_i^1 affects t_i^2.[20] The second condition states that any dependence between the time on the market of the two listings of the same home is accounted for by the home-specific hazard, $\lambda_{0i}(\cdot)$. The third condition is the standard assumption of independent censoring similar to the assumption underlying the Kaplan-Meier estimator. Here, however, censoring is assumed to be conditionally independent given the home-specific unobserved heterogeneity, rather than unconditionally independent.

5.1.1 Implementation: Proportional Hazard Index

The procedure to estimate the coefficients β_t is straightforward and can be summarized as follows.

- Step 1: Identify the relevant sample of repeat listings (observations where $W_i = 1$). We focus first on listings that appear in more than one period. Among this set of repeat listings, we select those properties with completed durations in both periods or if only the larger of the two durations is censored. That is, pick observations where $\tilde{C}_i \geq \min\{Y_i^1, Y_i^2\}$.
- Step 2: Calculate the dependent variable. Using the sample defined in the previous step, we estimate an indicator that takes the value of 1 if $Y_i^2 \geq Y_i^1$, and zero otherwise.
- Step 3: Estimation of a logistic model. The dependent variable is the indicator computed in the previous step, and the covariates are the variables in vector X_i.

5.1.2 Results: Repeat Proportional Hazard Index (RPHI)

In the proportional hazard model we take $\mu_t = \exp(\beta_t)$ as the repeat proportional hazard index (RPHI) which can be interpreted as the gross percentage increase in the hazard rate relative to the baseline period. The index is pegged at $\mu_{t_0} = 1$. In other words, if $\mu_t = 1.5$ the probability that a home listed in period t will sell on any given day, conditional on still being on the market, is 50% larger than it would be if it had been listed in the baseline period. The RPHI is estimated in all 15 MSAs in the sample and results are displayed in Table 2. In all areas the index has been

[20] It is plausible that t_i^2 depends on Y_i^1; the longer the listing is on the market before it sells, the later the next listing date will be. If this was a serious concern we would expect the results to be sensitive to restrictions of the sample to exclude observations with $t_i^2 - t_i^1 < c$, where c is an arbitrary cutoff. We have found that this is not the case.

normalized so that it equals 1 in the first quarter of 2010. There is significant variation in the RPHI both across areas and over time.

It is useful to compare the RPHI with the simple unconditional hazard ratio computed in Section 4.2. The hazard ratio discussed in Section 4.2 has the same interpretation as the RPHI and is estimated incorporating both censored and non-censored durations; however, it does not account for unobserved heterogeneity. The evolution of both of these variables is displayed in the top panels of Figures 5 and 6. While the overall trend of both indices is much alike, there are some important differences. For example, in Fairfax County, the RPHI during a booming market (in 2000) is much higher than the unconditional hazard ratio. Such patterns would arise if the types of homes that were listed in this particular period were especially hard to sell (less liquid) due to their unique features. On the contrary, in a slow market (2007), the unconditional hazard ratio is higher than the RPHI suggesting that homes being listed during "bad times" were, due to their unobserved characteristics, more liquid than the average home in the sample. This translates into a RPHI that is more volatile that the unconditional hazard ratio. These patterns seem to be persistent in the other urban areas.

In sum, just as it is the case with repeat-sales home price indices, controlling for unobserved heterogeneity is key to measure the evolution of the baseline hazard rate.

5.2 A Repeat Median TOM Index

An alternative method is based on the accelerated failure time model

$$\log(Y_i^s) = \beta_{t_i^s} + \alpha_i + \varepsilon_i^s \tag{5.2}$$

This generalizes a proportional hazards model with a constant hazard function by allowing the distribution of ε_i^s to be unrestricted. This model allows the hazard rate in a given period to be accelerated (or decelerated) relative to the baseline period, rather than just shifted proportionately (see, e.g., Kalbfleisch and Prentice, 2011).

In the accelerated failure time model, the home fixed effect, α_i, can be differenced out.

$$\log(Y_i^2) - \log(Y_i^1) = \beta' X_i + \tilde{\varepsilon}_i$$

where X_i is as defined in Section 5.1 and $\tilde{\varepsilon}_i = \varepsilon_i^2 - \varepsilon_i^1$. If $E(\tilde{\varepsilon}_i \mid X_i) = 0$ and there is no censoring then the usual fixed effects regression estimator will consistently estimate β.

A similar approach is still possible in the presence of censoring. First we make the following assumption.

Assumption 5.2.

(i) $Med(\alpha_i + \varepsilon_i^1 \mid t_i^1, t_i^2) = Med(\alpha_i + \varepsilon_i^2 \mid t_i^1, t_i^2)$

(ii) Y_i^1 *and* C_i^1 *are independent conditional on* t_i^1, t_i^2 *and* Y_i^2 *and* C_i^2 *are independent conditional on* t_i^1, t_i^2

The first condition has been used in other econometric models (e.g., Khan et al., 2011). It combines something like the strict exogeneity condition of Assumption 5.1(i) and a stationarity condition. To see this, note that it follows if (a) $\varepsilon_i^1 \mid \alpha_i, t_i^1, t_i^2 =_d \varepsilon_i^1 \mid \alpha_i$ and $\varepsilon_i^2 \mid \alpha_i, t_i^1, t_i^2 =_d \varepsilon_i^2 \mid \alpha_i$ and (b) $\varepsilon_i^1 \mid \alpha_i =_d \varepsilon_i^2 \mid \alpha_i$. Under condition (i) β is identified if $Med(\log(Y_i^s) \mid X_i)$ is identified for $s = 1, 2$ because

$$Med(\log(Y_i^2) \mid X_i) - Med(\log(Y_i^1) \mid X_i) = \beta' X_i$$

Moreover, $Med(\log(Y_i^s) \mid X_i)$ will generally be identified under independent censoring (condition (ii)), as described in Section 4.2, via the method of Kaplan and Meier (1958). An important caveat is relevant if there is a value $\bar{y}$ such that all observations with durations exceeding $\bar{y}$ are censored. In this case the u quantile of the distribution is only identified for $u \leq \bar{u}$ where $\bar{u}$ is the proportion of the durations that exceed $\bar{y}$. If $\bar{u} < 1/2$ then the median is not identified. However, identification of $Med(\log(Y_i^s) \mid X_i)$ for every value of X_i (i.e., every pair (t_i^1, t_i^2) is not necessary in order to identify β.

This solution to censoring in the model of equation (5.2) is more straightforward than other suggestions in the literature (most notably, Honoré et al., 2002) because we are able to produce reliable conditional Kaplan-Meier estimates since the only covariates are discrete time periods (quarters in the results reported in the paper). If we wanted to condition on continuous home characteristics (square footage, for example) then our approach would not be feasible and the more complicated method of Honoré et al. (2002) would be necessary.

5.2.1 Implementation: Median Index

The model is estimated in a two step procedure. First we select a set of repeat listings: those that were put on the market in period t for the first time and in period $t+k$ for the second time. Using only this subset of observations (that includes both completed and censored durations) we estimate the median log TOM separately in *each* period by carrying out the Kaplan-Meier estimator. We can then compute the difference in the median log TOM between periods t and $t+k$. Note that unobserved heterogeneity dissapears when the difference in median log TOM is computed (as long as assumption 5.2 holds.) We repeat these calculations for every other pair of periods in our sample. Then, in a second step, we run a simple OLS regression to estimate β. The procedure to estimate the repeat median TOM index is summarized below:

- Step 1: We estimate $Med(\log(Y_i^2) \mid X_i) - Med(\log(Y_i^1) \mid X_i)$ by carrying out the Kaplan-Meier procedure conditional on $t_i^1, t_i^2 = t_1, t_2$ for each pair of t_1, t_2 such that $t_1 < t_2$. For each observation i we then have an estimate $\widehat{\delta M}_i$ of $Med(\log(Y_i^2) \mid X_i) - Med(\log(Y_i^1) \mid X_i)$.
- Step 2: We run an OLS regression of $\widehat{\delta M}_i$ on X_i to estimate β.

As we mentioned in the introduction, this procedure can be viewed a as a repeat sales quantile (median) regression that makes use of an accelerated failure time (AFT) assumption and uses a conditional Kaplan-Meier estimator to correct for censoring in a first stage.

5.2.2 Results: Repeat Median TOM Index (RMTI)

A repeat sales index for home prices, such as the Case-Shiller index, is created by taking $\hat{\mu}_t = \exp(\hat{\beta}_t)$. This transformation is natural in that context because $\exp(\beta_t)$ represent the gross market return between the initial period and period t. Here we will use the same construction for the index though the interpretation as a gross return is less salient.

In the accelerated failure time model we take $\mu_t = \exp(\beta_t)$ and define it as the repeat median TOM index (RMTI). Notice that a larger index value represents an increase in the time on the market. For example, if $\mu_t = 1.5$ then the median time on the market would have been 50% larger in period t compared to the base period. The RMTI is estimated in all 15 MSAs in the sample and results are displayed in Table 3. In all areas the index has been normalized so that it equals

1 in the first quarter of 2010. As it was the case with the RPHI, the RMTI exhibits significant variation both across areas and over time.

In the bottom panels of Figures 6 and 7 we show the inverse of the RMTI in selected urban areas. We plot the inverse (rather than the level) of the RMTI to facilitate a direct comparison with the RPHI above. These two indices can be compared as they both represent rates at which the time on the market changes over time. As expected, in all areas the RMTI and the RPHI seem extremely highly correlated. In fact, if the baseline hazard function is constant then the two models coincide and the two indices estimate the same thing.

We compare the RMTI with a simple ratio of unconditional medians. The unconditional medians have been estimated using the Kaplan-Meier estimator and censored durations but do not account for housing unobserved heterogeneity. The RMTI and the unconditional median ratio follow a similar trend. In some areas, the trends are almost identical (Fairfax County) while in others there are substantial differences (Miami and New Orleans, for example). As it was the case with the repeat proportional hazard index, controlling for unobserved heterogeneity is important when measuring the evolution of the median TOM.

5.3 Discussion

5.3.1 Sample selection bias

The home selling process requires sellers to make a series of optimal choices. Home owners choose when (and if) to put their home on the market. Once a unit is listed, sellers choose a list price and a reservation value. List prices may change during the marketing period. When an offer (or multiple offers) arrive, sellers bargain with potential buyers over the transaction price and decide whether or not to trade.[21] All of these sellers' optimal choices will directly determine the sample of observations available for construction of price and TOM indices. Conventional price indices such as the Case-Shiller and FHFA use transaction prices and reflect market conditions at the time of the transaction of the subset of homes that were traded in each period. These estimates do not necessarily reflect price changes of the stock of homes (that includes those that are not for sale). Some attempts to correct for sample selection bias in the construction of real estate price indices

[21] At any time the home can be withdrawn from the market. For an excellent study that describes the home selling process, see Merlo et al. (2015).

have been made, but they typically rely on strong parametric assumptions (see for example Fisher et al. (2003)). A price index constructed with transaction-level data will trace changes in market prices of homes that have been sold, while an index that features a sample-selection correction will feature changes in prices of all homes (both in and out of the market). In both cases, the index reflects market conditions at the time of the transaction.

To estimate TOM indices in our application we use all listings including those that end up in a sale as well as those that have been withdrawn from the market. By treating withdrawn listings as censored observations, we estimate TOM indices that reflect market conditions of homes that were put on the market at the time they were initially listed. In other words, the indices provide information to answer the following question: given that a home seller has decided to sell her home, how long will it stay on the market before it is sold? This seems to be the relevant question that markets participants care about. Of course, our methods are still subject to sample selection bias because we cannot control for the decision to list. Our index provides information about expected TOM conditional on the listing decision, and cannot be interpreted as a measure of TOM of the overall stock of housing. As it is also the case with home price indices, computing TOM indices that are robust to sample-selection bias is an important topic for future research.

5.3.2 Controlling for sellers' idiosyncratic characteristics

Does seller's motivation or, more generally, seller's characteristics affect the home selling strategy? The literature suggests this is the case: highly motivated sellers have lower reservation prices, set lower asking prices and, on average, sell their units faster (see, for example, Glower et al. (1998) and Carrillo (2012)). Our TOM index controls for housing observed and unobserved characteristics to account for the fact that some units are always easier/harder to sell. However, as it is the case with all home price indices, we do not attempt to control for sellers' motivation and/or for sellers' characteristics. Hence, the index itself captures changes in sellers' motivation and other factors that influence their willingness to trade. For example, an increase in the RMTI reflects a slower market, which may have been determined in part by changes in sellers' motivation and other of their idiosyncratic conditions.

The motivation of a repeat sales approach (for prices or TOM) to control for unobserved heterogeneity in home characteristics is to enable comparison of the same good/market (the same bundle of home characteristics) over time. Controlling for unobserved seller characteristics, on the other hand, would not be motivated by such a concern. The market is defined by the bundle of (observed and unobserved) home characteristics but not by the identity of the market participants. Assessing how changes in TOM indices are related to changes in home buyers and sellers economic conditions (including motivation to trade) is an interesting topic that deserves further research.

5.3.3 Comparing the RPHI vs. RMTI

We have proposed two indices to measure changes in TOM. It is important to discuss virtues and limitations of each index to guide researchers who may need to choose one over the other. There are four criterion on which we can base a comparison. First, the assumptions made. Second, the intuitive appeal. Third, the computational appeal. Fourth, performance for forecasting (or other uses of the index).

The main difference in the underlying statistical assumptions is that the RPHI assumes that changes over time in the distribution of TOM are reflected in proportional shifts in the hazard rate. The RMTI, on the other hand, assumes that changes over time in the distribution of TOM operate through changes in the median log TOM. The RMTI is based on an accelerated failure time model where changes over time in market conditions cause a proportional shift in the hazard but at the same time an acceleration of the hazard rate (or deceleration if market conditions are worsening). However, the two are not nested as the AFT model assumes a parametric structure that links these two changes. The exception is when the hazard rate is constant over time.

Based only on a comparison of the assumptions, the RPHI may be preferred. According to the model underlying the RMTI, homes that will remain on the market longer disproportionately benefit (in terms of the hazard) from better market conditions. This seems counterintuitive as one might expect motivated sellers of easily marketable homes to be in a better position to take advantage of a sudden increase in demand. Further, the proportional hazard assumption is more commonly used in other applications.

The RPHI makes use only of knowledge of which of the two listings of the same home was longer, but not the relative length of the two listings. The reasoning is that the relative length of the two listings is not the proper way to difference out the unobserved heterogeneity in the proportional hazards model. Instead, the unobserved heterogeneity can only be differenced out by taking a certain transformation of the two TOMs and then differencing. Because this transformation is not generally known, the only information that is unbiased by unobserved heterogeneity is which duration was longer. The logic of the RMTI on the other hand fits with the traditional understanding of a fixed effects model (in logs). For this reason, the RMTI does seem to have greater intuitive appeal. Remarkably, this seemingly stark difference between what patterns in the data each construction is taking advantage of does not lead to substantially different indices.

Computationally, RPHI is preferred because it requires less time. The first stage of the RMTI requires constructing a Kaplan-Meier survival curve for each pair of potential listing periods. With T quarters, this requires as many as $T(T-1)$ Kaplan-Meier constructions.

Lastly, in our forecasting exercise described in Section 6.2 both indices seem to perform equally well.

6 Joint Distribution of TOM and Home Prices: Some Stylized Facts

In this section we use our proposed TOM indices to establish several stylized facts about the relationship between TOM and home prices. We first look at how the joint distribution of price and TOM changes over time. We then describe the correlation between our TOM indices and conventional repeat sales price indices in the regions represented in our sample.

6.1 Conditional vs. Unconditional Distribution

In order to understand how the housing market evolves over time it is important to highlight again that residential real estate prices and TOM are jointly determined; that is, housing markets clear in both TOM and price dimensions. Analysis of both dimensions is necessary to understand housing market conditions. Tracking separately the distribution of prices and that of TOM misses out on important features of the joint distribution of price and TOM. In particular, as market conditions change, home sellers change their listing strategy and both TOM and price change as a result. What market participants and policymakers ideally want to know is how long it will take today to sell a home at different price points, or, conversely, how much a home will sell in a given amount of time. Carefully disentangling the various mechanisms at play would involve jointly modeling not just TOM and transaction price, but also list price and list price adjustments.[22] While this is beyond the scope of the present paper, we provide a starting point for such an investigation in this section.

The *joint* distribution of price (p) and TOM (y) can be factored into the conditional distribution of price given TOM and the marginal distribution of TOM. To fully describe changes in the joint distribution of price and TOM $f(p, y) = f(p|y)f(y)$, one can evaluate changes in the *unconditional* distribution of TOM $f(y)$ as well as changes in the *conditional* distribution of prices given TOM $f(p|y)$. The main contribution of this paper has been to properly measure the latter. However, if the conditional distribution of price given TOM changes over time differently at different levels of TOM then this suggests that it is important to consider changes over time in the full joint distribution.

Almost all home price indices proposed in the literature focus on measuring the *unconditional* distribution of home prices $f(p)$ rather than the conditional distribution $f(p|t)$, ignoring that liquidity (TOM) is notoriously variable over time. The exception is Fisher et al. (2003) where a constant-liquidity real estate price index (for commercial real estate) is proposed. They find that constant-liquidity price indices display greater volatility and cycle amplitude than standard transaction based price indices. Indices that are based on real estate transaction prices alone will also incorporate underlying changes in liquidity.

We estimate a repeat-sales constant-liquidity home price index and compare it with the conventional repeat-sales index. The constant-liquidity index we propose is based on homes that are sold quickly in every period (high-liquidity index). In each of our MSAs, we identified "highly liquid" homes in our repeat sales sample as those which went under contract within 15 days of being listed both times the property was listed. We then constructed repeat sales price indices using only these homes. In some cases we used a cutoff of 30 or 60 days on the market to ensure a large enough sample remained. In Figures 8-12, we compare these high liquidity price indices to repeat sales price indices constructed using all homes with at least two sales in the sample. Unlike Fisher et al. (2003), we do not find substantial differences between the unconditional and conditional price indices. To the extent that "high liquidity" homes are priced differently, this can be explained by a

[22] List prices can have non-linear effects on sale prices and TOM. Higher list prices are generally associated with higher transaction prices and TOM, but asking prices that are unusually low can generate "bidding wars" resulting in quick transactions at high prices (Han and Strange, 2016, 2014).

factor that is time-invariant, at least over the time period covered by our data.

Constant-liquidity home price indices can be estimated in a variety of ways. The approach proposed by (Fisher et al., 2003) is not feasible in our application because it requires estimation of a sample selection equation. By conditioning on TOM we are able to estimate the constant-liquidity price indices without relying on any parametric identification. In the appendix, we also discuss an alternative way to construct a price index that is conditional on TOM that is also based on repeat-sales but includes both quick and slow transactions. We find similar results using this alternative method.

6.2 ADL Regressions and Forecasting

Carrillo et al. (2015) found that crude measures of housing liquidity are correlated with future home price appreciation. In this section, we explore this relationship using the novel repeat time-on-the-market indices computed above. Such exercise can help researchers assess the relative strength of each one of these indices. We first look at the time series relationship between measures of TOM and home prices within a market using an autoregressive distributed lag (ADL) regression model. Let $\hat{\mu}_{j,t}$ denote a TOM index for region j in quarter t and let $\hat{\mu}^P_{j,t}$ denote the FHFA price index for region j in quarter t. Let $\Delta^P_{j,t}$ denote the year-over-year home price appreciation, that is, $\hat{\mu}^P_{j,t} - \hat{\mu}^P_{j,t-4}$. The ADL regression model for region j is

$$\Delta^P_{j,t} = \alpha_j + \sum_{\ell=1}^{L} \beta_{\ell j} \Delta^P_{j,t-\ell} + \gamma_i \hat{\mu}_{j,t-1} + \varepsilon_{jt}$$

We estimated various specifications of this model. In Table 4, we report results from specifications where $L = 3$. We estimate the model separately using each of two TOM indices: the RPHI and RMTI developed in this paper. The RPHI and RMTI indices are constructed in such a way that a lower value on the index represents a shift upward in the distributiof of TOM (a lower hazard rate or higher median, respectively). Therefore we find, with only a few exceptions, across regressions using both indices, that an increase in TOM is associated with a reduction in home price appreciation controlling for current and past price appreciation. This is in line with previous evidence from Carrillo et al. (2015). We also find that the coefficient is statistically significant at a 10% level in the majority of CBSAs. The positive correlation between TOM and future appreciation rates seems to be stronger when the RPHI is used.

We also estimate a panel ADL model, specified in equation (6.1) below.

$$\Delta^P_{j,t} = \alpha_j + \sum_{\ell=1}^{L} \beta_\ell \Delta^P_{j,t-\ell} + \gamma \hat{\mu}_{j,t-1} + \varepsilon_{jt} \tag{6.1}$$

This model allows for region-specific effects but imposes common coefficients. As reported in Table 5, we again find a positive and statistically significant correlation between TOM and home price appreciation. The R^2 in columns 2 and 3 are almost identical, suggesting that either one of the indices can account for about the same portion of variance of future home price appreciation. Both findings – the panel ADL and individual ADL analysis – are robust to slight changes in the number of lags.

The above evidence suggests that the TOM indices might improve home price forecasts. We now use a vector autoregression (VAR) with the price and TOM indices to forecast home prices. Unlike the previous exercises we replace $\hat{\mu}_{jt}$ with a real-time measure of TOM at period t. That is, to construct $\hat{\mu}_{jt}$ we artificially censor the time-on-the-market for any homes that were initially listed during period t and remained on the market at the end of the period. Taking $M_{jt} = [\hat{\mu}^P_{jt}, \hat{\mu}_{jt}]'$ we can specify the VAR for region j as follows.

$$M_{jt} = \alpha_j + \sum_{\ell=1}^{L} \beta_{\ell j} M_{\ell,t-1} + \varepsilon_{jt}$$

In our main specification, reported in Table 6, we impose $L = 1$. More lags generally do not seem to improve the forecast, likely due to the fact that our series are rather short. We use VARs estimated recursively starting with the $t = 20$, corresponding to fourth quarter of 2008, to create one-period and two-period ahead forecasts. We evaluate the forecasts by computing the root mean squared error of the forecast over time and across region. We also compute the percent of the time that the realized price appreciation fell within the 90% prediction interval. From the results we see that including TOM measures improves forecasts of home price appreciation.

As far as comparing the performance of the RMTI with the RPHI in this application we find somewhat mixed results. While the in-sample RMSE is slightly smaller when the RPHI is used, forecasting performance is higher with the RMTI. Depending on the nature of the application, researchers could choose one index over the other.

6.3 Long Run Trends

The ADL regression results above show that shocks to TOM have a dynamic impact on future home price appreciation, suggesting the possibility of a long run cointegrating relationship between home prices and TOM. Motivated by this, we test for the presence of cointegration between our TOM indices and the FHFA price indices.

In most markets $\Delta^P_{j,t}$ and $\hat{\mu}_{j,t}$ are both $I(1)$, as demonstrated in Table 7. This precludes the possibility of cointegration between $\hat{\mu}^P_{j,t}$ and $\hat{\mu}^P_{j,t}$ but suggests that TOM measures might instead be cointegrated with home price appreciation. To test if home price appreciation is cointegrated with the TOM indices, we run a two-step test. We first estimate the following regression model for each market j

$$\Delta^P_{jt} = a_j + g_j \hat{\mu}_{jt} + \nu_{jt}.$$

We then form the residual series, $\hat{\nu}_{jt}$, and test the unit root null using the augmented Dickey-Fuller test. This test amounts to a test of the null that $\tilde{g}_j = 0$ in the regression

$$\Delta\hat{\nu}_{jt} = \tilde{a}_j + \tilde{g}_j \hat{\nu}_{j,t-1} + \sum_{\ell=1}^{L} \tilde{b}_{j\ell} \Delta\hat{\nu}_{j,t-\ell} + e_{jt}.$$

When $\Delta^P_{j,t}$ is defined as the year-to-year appreciation, evidence in favor of cointegration is weak and not very robust to changes in the specification of the number of lags, L, used in the second stage. This is likely due to combination of two factors – the serial correlation induced by taking year-over-year differences and the small sample size used for our tests (based only in 36 observations for each region). The ADF test is known to have low power especially for shorter series, and more so when there is a need to include a large number of lags in the second step. When $\Delta^P_{j,t}$ is instead defined as the quarter-to-quarter appreciation, $\hat{\mu}^P_{j,t} - \hat{\mu}^P_{j,t-1}$ we find much stronger results, as reported in Table 8. There we see that we can reject the null in favor of cointegration for 8 out of the 14 regions.

In sum, our results suggest that TOM is cointegrated with home price appreciation rather than with home price levels.

7 Conclusions

This paper develops the first housing market TOM indices that are based on repeat listings. Important features of the two indices we propose are their ability to control for unobserved heterogeneity exploiting *repeat* listings, and their use of censored durations. The first index, the RPHI, computes proportional displacements in the home sale baseline hazard rate. This index is based on an econometric model that has been used on other contexts; the application of this method to the measurement of real estate liquidity is one of the main contributions of our paper. The second index, the RMTI, estimates the relative change in (quality adjusted) median TOM. The RMTI uses a novel econometric procedure that contributes to the literature in econometrics that attempts to relax the proportional hazards assumption. We compute the indices using listings data from 15 US urban areas including Miami, San Diego, Las Vegas, and a suburb of Washington D.C. The combined empirical evidence suggests that to measure housing liquidity it is key to account for observations that are censored when the home is withdrawn from the market and to control for unobserved heterogeneity using repeat listings.

We also highlight the computational transparency and simplicity of both indices. The RPHI can be estimated using a simple logistic regression, while the RMTI can be estimated with a simple two step procedure that combines the estimation of median TOM in each period and an OLS regression. Given the availability of MLS data, we hope that the application of our methods is a straightforward task. Periodic reporting of such indices should be useful to investors, regulators, home buyers and home owners to assess housing market conditions and make informed decisions.

Estimation of home price indices has been the focus of the literature and practical applications. After we construct our TOM indices, we also develop a framework to study how the *joint* distribution of prices and TOM evolves over time. This is a topic that deserves further scrutiny given that the housing market clears in both price and TOM dimensions.

References

Abbring, J. H., P.-A. Chiappori, and J. Pinquet (2003): "Moral hazard and dynamic insurance data," *Journal of the European Economic Association*, 1, 767–820.

Ambrose, B. W., N. E. Coulson, and J. Yoshida (forthcoming): "The Repeat Rent Index," *The Review of Economics and Statistics.*

Bailey, M. J., R. F. Muth, and H. O. Nourse (1963): "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association*, 58, 933–942.

Carrillo, P. E. (2012): "An empirical stationary equilibrium search model of the housing market," *International Economic Review*, 53, 203–234.

Carrillo, P. E., E. R. de Wit, and W. Larson (2015): "Can Tightness in the Housing Market Help Predict Subsequent Home Price Appreciation? Evidence from the United States and the Netherlands," *Real Estate Economics*, 43, 609–651.

Carrillo, P. E. and J. C. Pope (2012): "Are homes hot or cold potatoes? The distribution of marketing time in the housing market," *Regional Science and Urban Economics*, 42, 189–197.

Case, K. E. and R. J. Shiller (1987): "Prices of Single-Family Homes Since 1970: New Indices for Four Cities," *New England Economic Review*, 45.

——— (1989): "The Efficiency of the Market for Single-Family Homes," *The American Economic Review*, 79, 125.

Chamberlain, G. (1985): "Heterogeneity, omitted variable bias, and duration dependence," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman and B. Singer, Cambridge U. Press.

Cox, D. R. (1972): "Regression models and life-tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, 187–220.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044.

Efron, B. (1967): "The two sample problem with censored data," *Proc. 5th Berkeley Sympos. Math. Statist. Prob., Prentice-Hall: New York.*

Eurostat (2013): *Handbook on Residential Property Prices Indices*, no. 17280 in World Bank Publications, The World Bank.

Fisher, J., D. Gatzlaff, D. Geltner, and D. Haurin (2003): "Controlling for the Impact of Variable Liquidity in Commercial Real Estate Price Indices," *Real Estate Economics*, 31, 269–303.

Flinn, C. and J. Heckman (1982): "New methods for analyzing structural models of labor force dynamics," *Journal of Econometrics*, 18, 115–168.

Glower, M., D. R. Haurin, and P. H. Hendershott (1998): "Selling time and selling price: The influence of seller motivation," *Real estate economics*, 26, 719–740.

Gobillon, L., T. Magnac, and H. Selod (2011): "The effect of location on finding a job in the Paris region," *Journal of Applied Econometrics*, 26, 1079–1112.

Gönül, F. and K. Srinivasan (1993): "Consumer purchase behavior in a frequently bought product category: estimation issues and managerial insights from a hazard function model with heterogeneity," *Journal of the American Statistical Association*, 88, 1219–1227.

Han, L. and W. C. Strange (2014): "Bidding Wars for Houses," *Real Estate Economics*, 42, 1–32.

——— (2015): "Chapter 13 - The Microstructure of Housing Markets: Search, Bargaining, and Brokerage," in *Handbook of Regional and Urban Economics*, ed. by J. V. H. Gilles Duranton and W. C. Strange, Elsevier, vol. 5 of *Handbook of Regional and Urban Economics*, 813 – 886.

——— (2016): "What is the role of the asking price for a house?" *Journal of Urban Economics*, 93, 115 – 130.

Haurin, D. (1988): "The Duration of Marketing Time of Residential Housing," *Real Estate Economics*, 16, 396–410.

Heckman, J. and B. Singer (1984): "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica: Journal of the Econometric Society*, 271–320.

Heckman, J. J. and G. J. Borjas (1980): "Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence," *Economica*, 47, 247–283.

Heckman, J. J. and B. E. Honoré (1989): "The identifiability of the competing risks model," *Biometrika*, 76, 325–330.

Honoré, B., S. Khan, and J. L. Powell (2002): "Quantile regression under random censoring," *Journal of Econometrics*, 109, 67–105.

Honoré, B. E. (1992): "Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects," *Econometrica: journal of the Econometric Society*, 533–565.

Horowitz, J. L. (1992): "The Role of the List Price in Housing Markets: Theory and an Econometric Model," *Journal of Applied Econometrics*, 7, 115–29.

Kalbfleisch, J. D. and R. L. Prentice (1980): *The statistical analysis of failure time data*, John Wiley & Sons.

——— (2011): *The statistical analysis of failure time data*, vol. 360, John Wiley & Sons.

Kaplan, E. L. and P. Meier (1958): "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, 53, 457–481.

Khan, S., M. Ponomareva, and E. Tamer (2011): "Identification of Panel Data Models with Endogenous Censoring," Discussion paper 98, Duke University.

Lancaster, T. (1979): "Econometric methods for the duration of unemployment," *Econometrica: Journal of the Econometric Society*, 939–956.

——— (2000): "The incidental parameter problem since 1948," *Journal of econometrics*, 95, 391–413.

Levitt, S. D. and C. Syverson (2008): "Market Distortions When Agents are Better Informed: The Value of Information in Real Estate Transactions," *The Review of Economics and Statistics*, 90, 599–611.

Lin, Z. and K. D. Vandell (2007): "Illiquidity and Pricing Biases in the Real Estate Market," *Real Estate Economics*, 35, 291–330.

Lindgren, A. (1997): "Quantile regression with censored data using generalized L_1 minimization," *Computational Statistics & Data Analysis*, 23, 509–524.

Lippman, S. A. and J. J. McCall (1986): "An Operational Measure of Liquidity," *The American Economic Review*, 76, 43–55.

Liu, C. H., A. Nowak, and S. S. Rosenthal (2016): "Housing price bubbles, new supply, and within-city dynamics," *Journal of Urban Economics*, 96, 55 – 72.

Merlo, A., F. Ortalo-Magne, and J. Rust (2015): "THE HOME SELLING PROBLEM: THEORY AND EVIDENCE," *International Economic Review*, 56, 457–484.

Olsen, R. J. and K. I. Wolpin (1983): "The impact of exogenous child mortality on fertility: a waiting time regression with dynamic regressors," *Econometrica: Journal of the Econometric Society*, 731–749.

Ridder, G. and I. Tunali (1989): "Analysis of related durations: a semi-parametric approach with an application to a study of child mortality in Malaysia," Discussion paper 1, NORC and the University of Chicago.

——— (1999): "Stratified partial likelihood estimation," *Journal of Econometrics*, 92, 193–232.

Rosenthal, S. S. (2014): "Are Private Markets and Filtering a Viable Source of Low-Income Housing? Estimates from a "Repeat Income" Model," *American Economic Review*, 104, 687–706.

Trussell, J. and T. Richards (1983): "Correcting for unobserved heterogeneity in hazard models: an application of the Heckman-Singer procedure to demographic data." [Unpublished] 1983. Presented at the 52nd Annual Meeting of the Population Association of America Pittsburgh Pennsylvania April 14-16 1983.

Wallace, N. and R. Meese (1997): "The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches," *The Journal of Real Estate Finance and Economics*, 14, 51–73.

A Appendix

A.1 Derivation of Measures

We estimate quality-adjusted time-on-the-market distributions for each quarter-area combination in our sample. The method used to compute the distributions follow Carrillo and Pope (2012), who combine the Dinardo, Fortin and Lemieux (1996) (DFL) with the Kaplan-Meier (1958) estimator. This allows the DFL decomposition to work in cases where the dependent variable is subject to random censoring. To keep our exposition self-contained, we carefully review the decomposition method.

Let Y be our variable of interest (the time a listing stays on the market) and t_0 and t_1 refer to the two mutually exclusive periods (quarters) in each of the areas we analyze. The cumulative probability function of Y in period t_0 is defined as

$$F(y|T = T_0) = \Pr(Y \leq y|T = t_0) = \int F(y|x, T = t_0)h(x|T = t_0)dx \tag{A.1}$$

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 of housing characteristics 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 would look if the individual attributes x (number of bathrooms, bedrooms and age, for example) were the same as in period t_0 (the base quarter). We denote this counterfactual as $F_{t_1 \to t_0}$ and express it symbolically as[23]

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

Using Bayes' rule, DFL recognized that

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

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

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

Notice that this expression differs from Equation A.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.

Carrillo and Pope (2012) note that the DFL method described above cannot be directly used in this application because marketing time is subject to random censoring; that is, some properties are not sold and withdrawn from the market. Because the random variable Y (marketing time) is

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

subject to random censoring, the counterfactual distribution can be computed using the Kaplan-Meier estimator, with sampling weights given by $\tau_{t_1 \to t_0}(x)$.

To be specific, we summarize the estimation algorithm for the counterfactual given that a random sample of N_0 and N_1 observations for periods t_0 and t_1 is available. Notice that in all steps described below the sample includes all censored and non-censored observations.

- Step 1: Estimate $P(T = t_0)$ using the share of observations where $T_i = t_0$; that is, compute: $\hat{\Pr}(T_i = t_0) = N_0/(N_0 + N_1)$.
- 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 $\hat{\Pr}(T_i = t_0|X = x_i) = \exp(x_i\hat{\beta})/\left(1 + \exp(x_i\hat{\beta})\right)$, where $\hat{\beta}$ is the parameter vector from the logit regression. Then, compute the estimated weights $\hat{\tau}_{t_1 \to t_0}(x)$.
- 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 $\hat{\tau}_{t_1 \to t_0}(x)$.

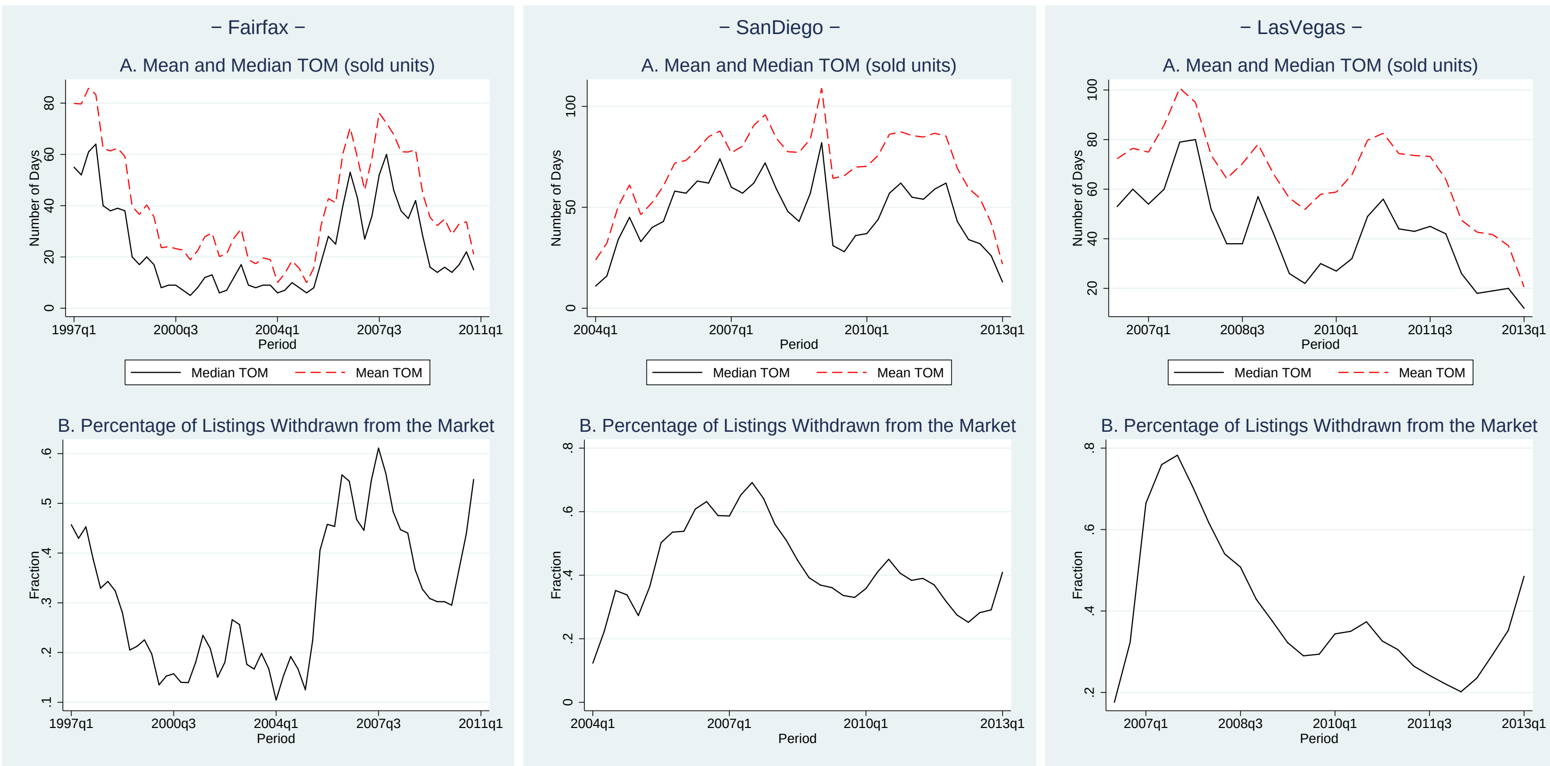
A.2 An alternative conditional price index

In Section 6.1 we construct a price index that is conditional on TOM by restricting the sample to "highly liquid" homes that sold within 15 (or 30 or 60) days. We can instead estimate the following price regression using repeat sales of the same home.

$$log(price_{it}) = a_t + b_t tom_{it} + \theta_i + \varepsilon_{it}$$

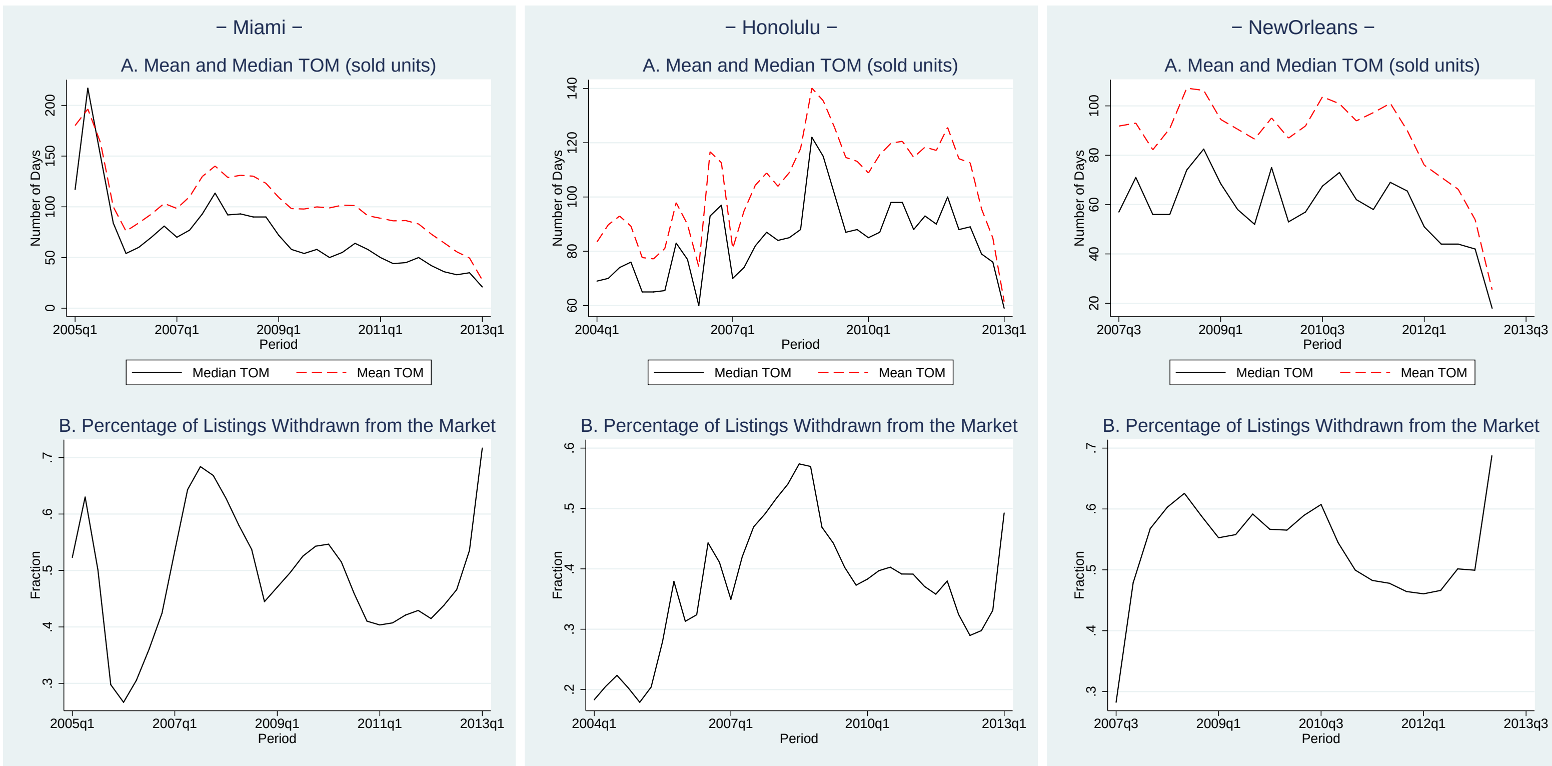
This can be estimated via a repeat sales regression where the indicators for which quarters the home was sold are also interacted with tom_{it}. Then estimates of a_t can be used to construct a price index that is conditional on TOM. Figure x compares this to a repeat sales price index that does not include TOM. As in Section 6.1 we find that the conditional price index does not vary substantially from the unconditional index.

Figure 1:
Descriptive “Conventional” Statistics (part 1)



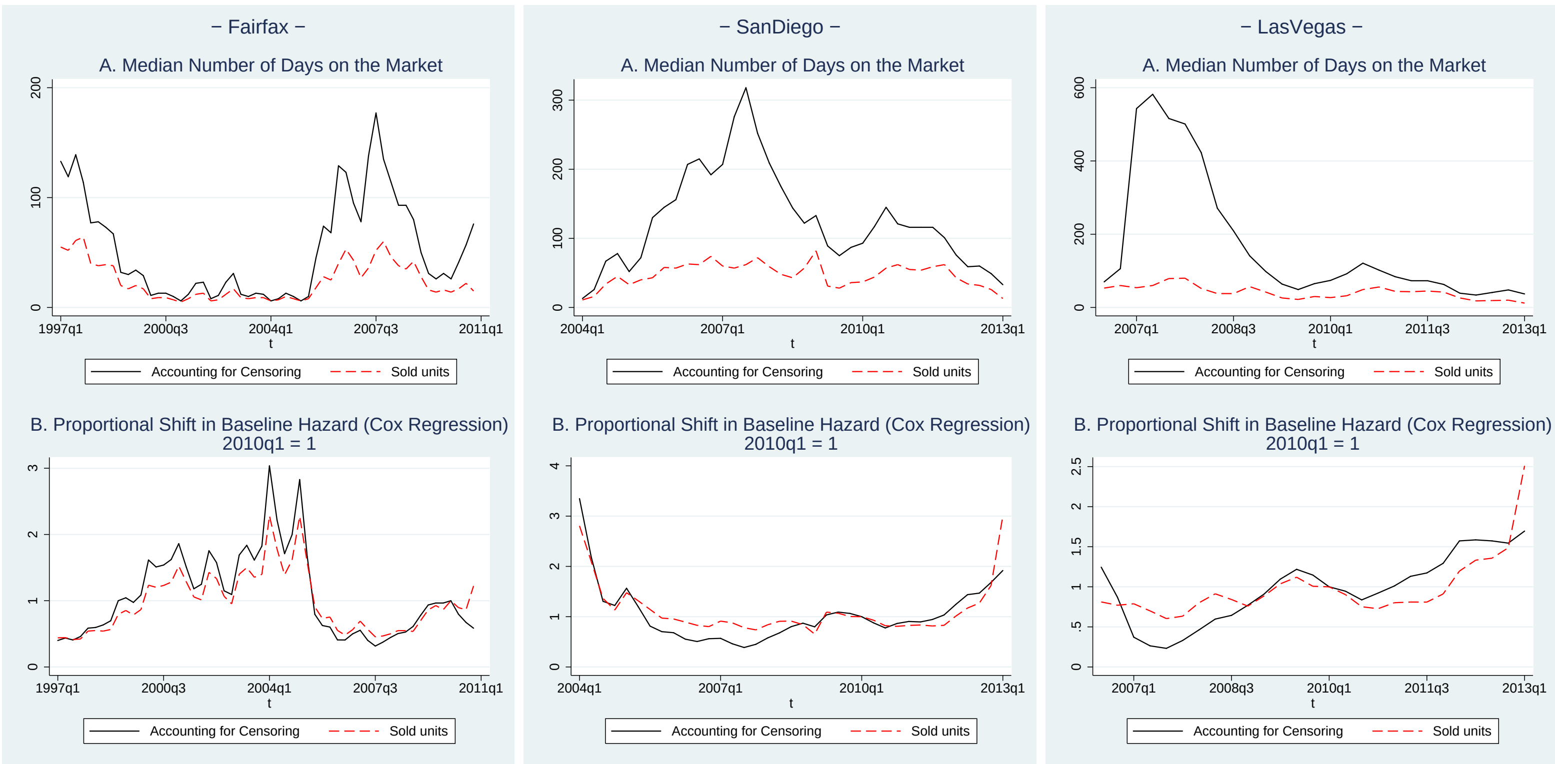
Notes: This figure presents descriptive statistics of the sample. Panel A computes the mean and median number of days that a home stays on the market (TOM). This is a “conventional” estimate that simply computes the mean and median TOM of finished durations (sold units). The second panel shows the share of total listings that are withdrawn from the market.

Figure 2:
Descriptive “Conventional” Statistics (part 2)



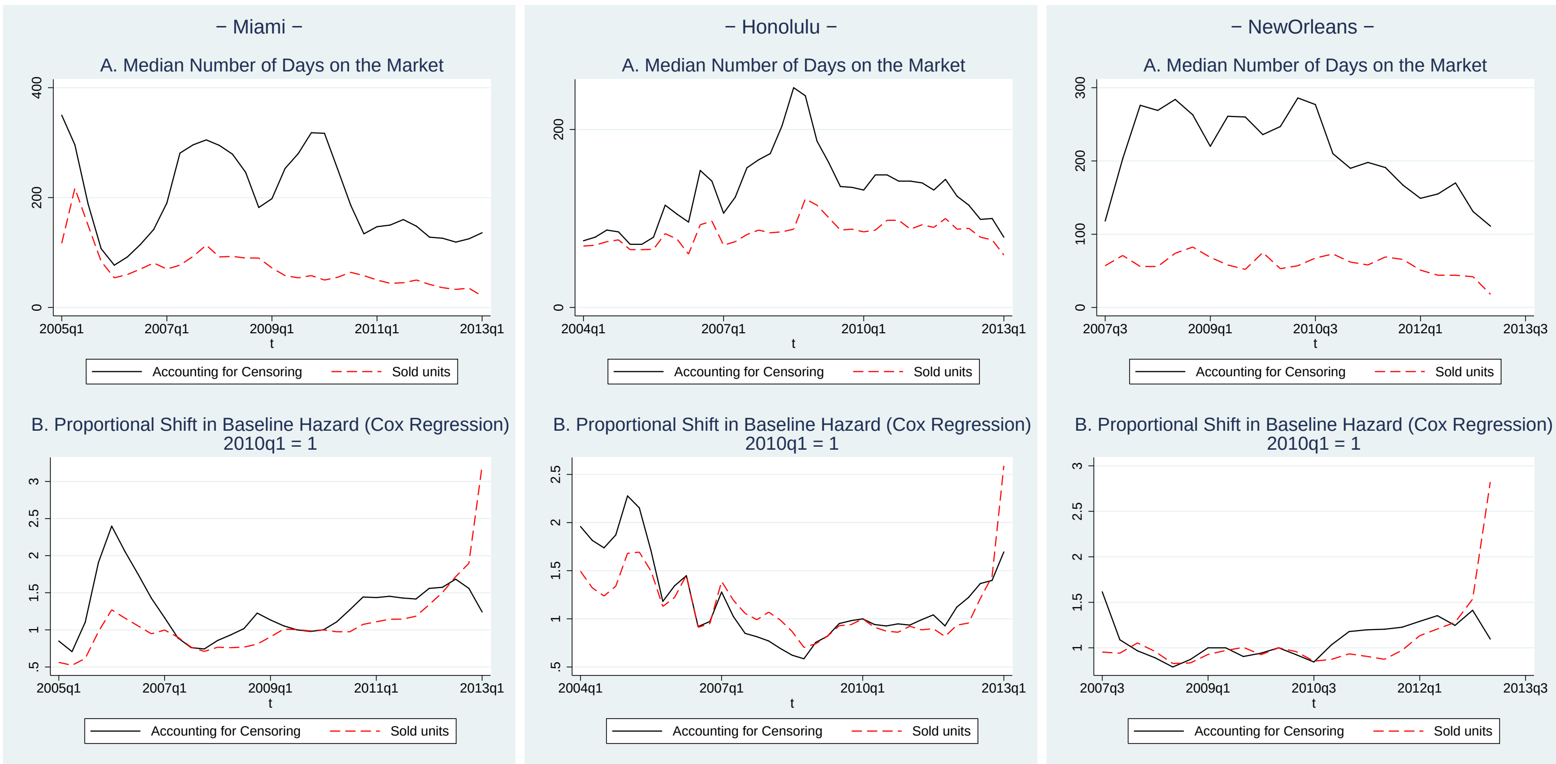
Notes: This figure presents descriptive statistics of the sample. Panel A computes the mean and median number of days that a home stays on the market (TOM). This is a “conventional” estimate that simply computes the mean and median TOM of finished durations (sold units). The second panel shows the share of total listings that are withdrawn from the market.

Figure 3:
Adjusting for Censoring When Computing TOM Statistics (part 1)



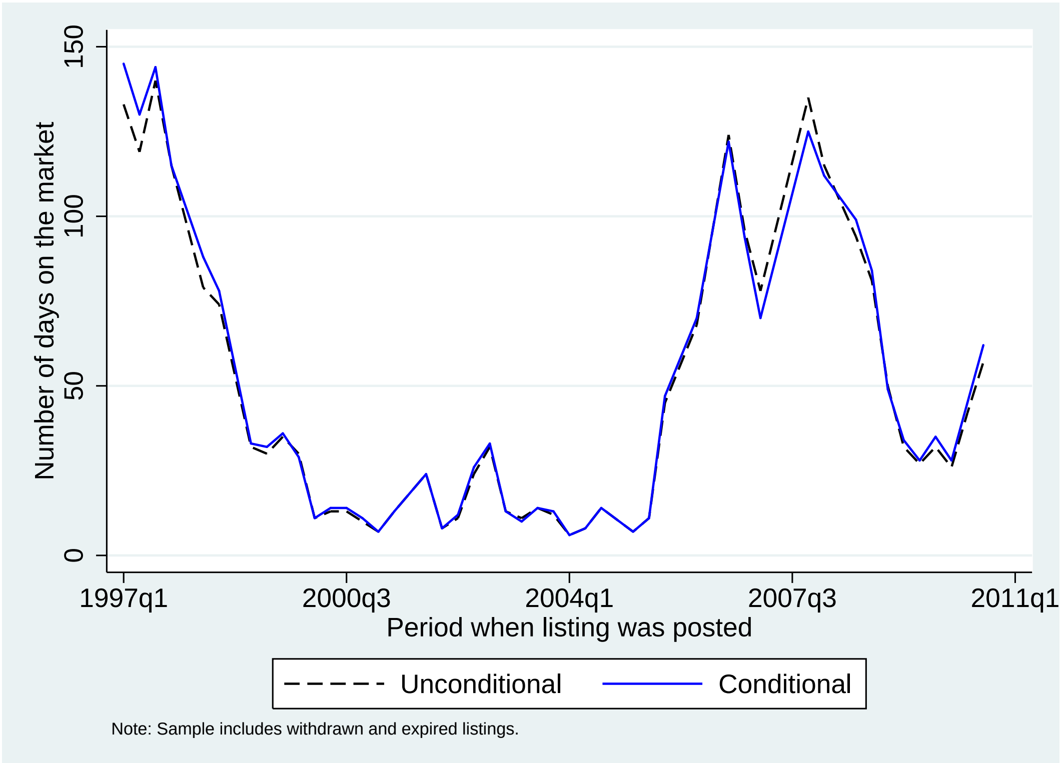
Notes: Panel A computes the median number of days that a home stays on the market (TOM). The "conventional" estimate simply computes the median TOM of finished durations (sold units). To account for censoring, a Kaplan-Meier estimator is used. For units that are sold, TOM is defined as the difference between the date when an offer was accepted and the date when the listing was posted. For censored observations, we compute duration as the difference between the date when the listing was posted and the date when it was withdrawn. In Panel B, a COX proportional hazard model is used to estimate changes in the baseline hazard relative to a base period (2010 q1). The "conventional" approach uses only the sample of finished durations (sold units). To account for censoring, proportional hazard models are estimated using both finished and censored durations.

Figure 4:
Adjusting for Censoring When Computing TOM Statistics (part 2)



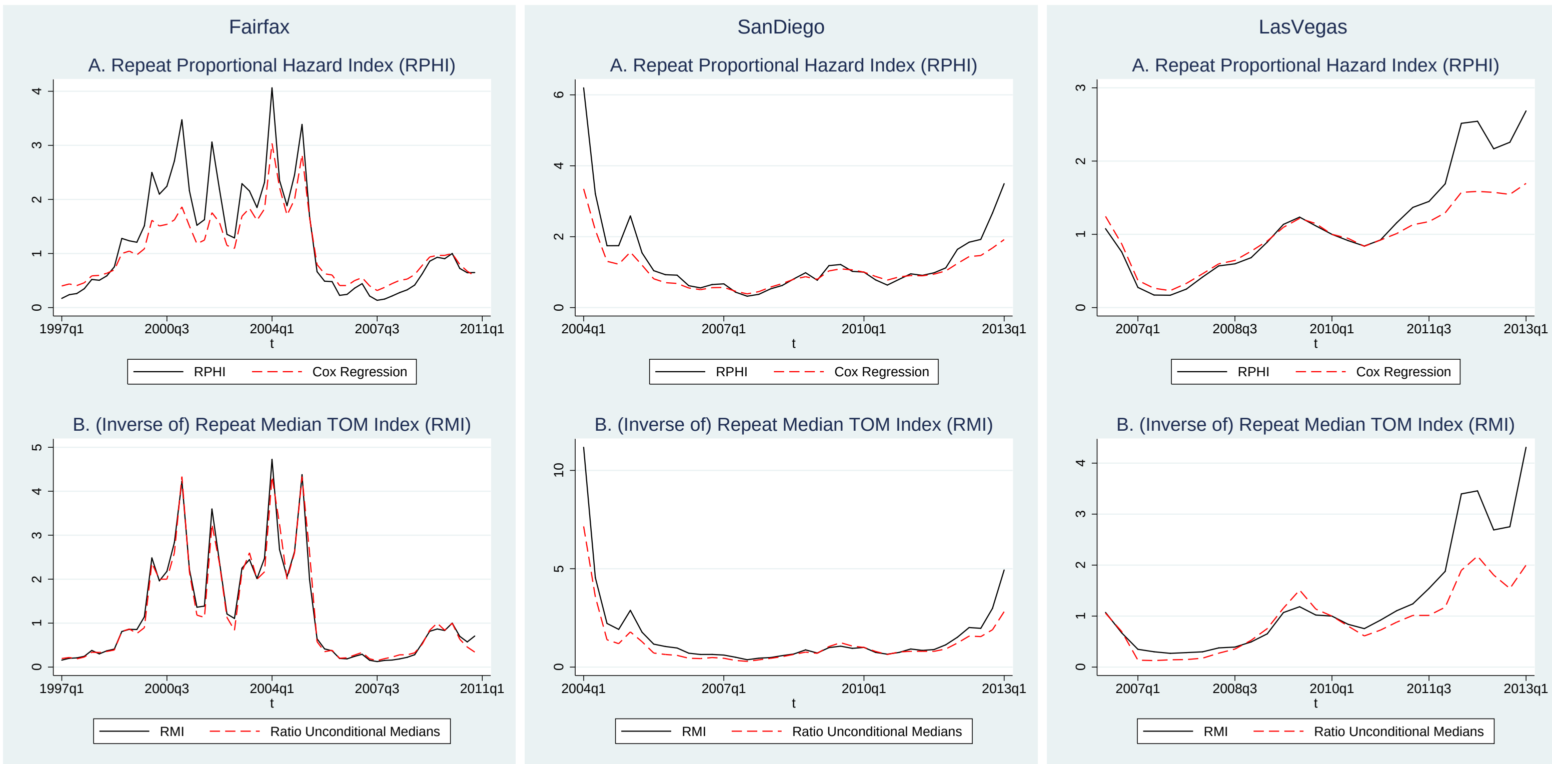
Notes: Panel A computes the median number of days that a home stays on the market (TOM). The "conventional" estimate simply computes the median TOM of finished durations (sold units). To account for censoring, a Kaplan-Meier estimator is used. For units that are sold, TOM is defined as the difference between the date when an offer was accepted and the date when the listing was posted. For censored observations, we compute duration as the difference between the date when the listing was posted and the date when it was withdrawn. In Panel B, a COX proportional hazard model is used to estimate changes in the baseline hazard relative to a base period (2010 q1). The "conventional" approach uses only the sample of finished durations (sold units). To account for censoring, proportional hazard models are estimated using both finished and censored durations.

Figure 5:
Controlling for Observed Heterogeneity: Conditional and Unconditional Median TOM
- Fairfax County, VA -



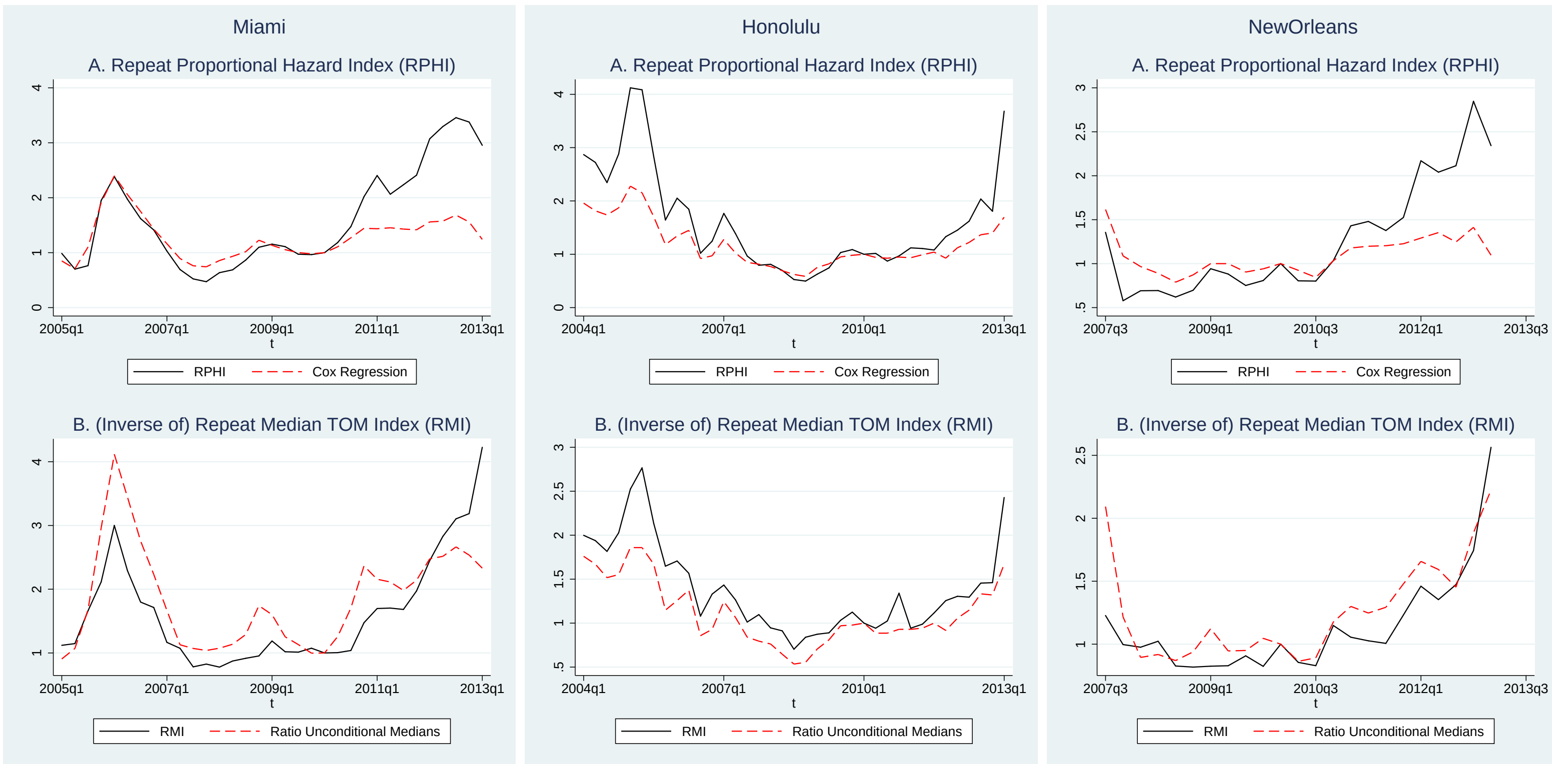
Notes: This Figure plots the evolution of the unconditional and conditional median TOM in Fairfax County. The unconditional median is computed using completed and censored durations and the Kaplan-Meier estimator. The conditional median is calculated using the re-weighting approach proposed by Dinardo, Fortin and Lemieux (1996). The conditional median TOM is simulated assuming that the characteristics of homes (age, structure type, number of bedrooms and bathrooms) remain constant as in a base period (2000 q1).

Figure 6:
Repeat Time-On-The-Market Indices (part 1)



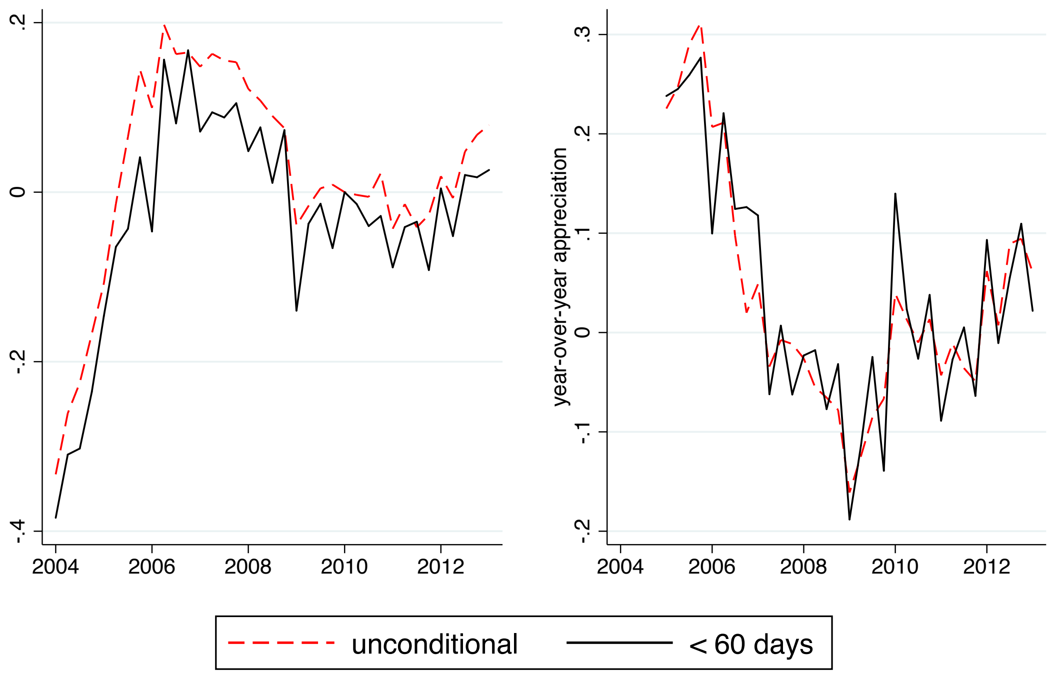
Notes: Panel A computes the Repeat Proportional Hazard Index (RPHI). The index measures relative shifts of the baseline hazard after controlling for unobserved home's heterogeneity. For example, and index value of 1.5 in period t reflects a 50 percent increase in the home sale baseline hazard in period t relative to the base period (2010 q1). The RPHI is compared with a similar index based on a Cox- regression that does not correct for unobserved heterogeneity. Panel B shows the (inverse of) the Repeat Median TOM Index (RMI). We report the inverse of the RMI to facilitate comparison with the RPHI. The RPHI is compared with a simple index that compares the relative shift of the unconditional median TOM in period t relative to the base period.

Figure 7:
Repeat Time-On-The-Market Indices (part 2)



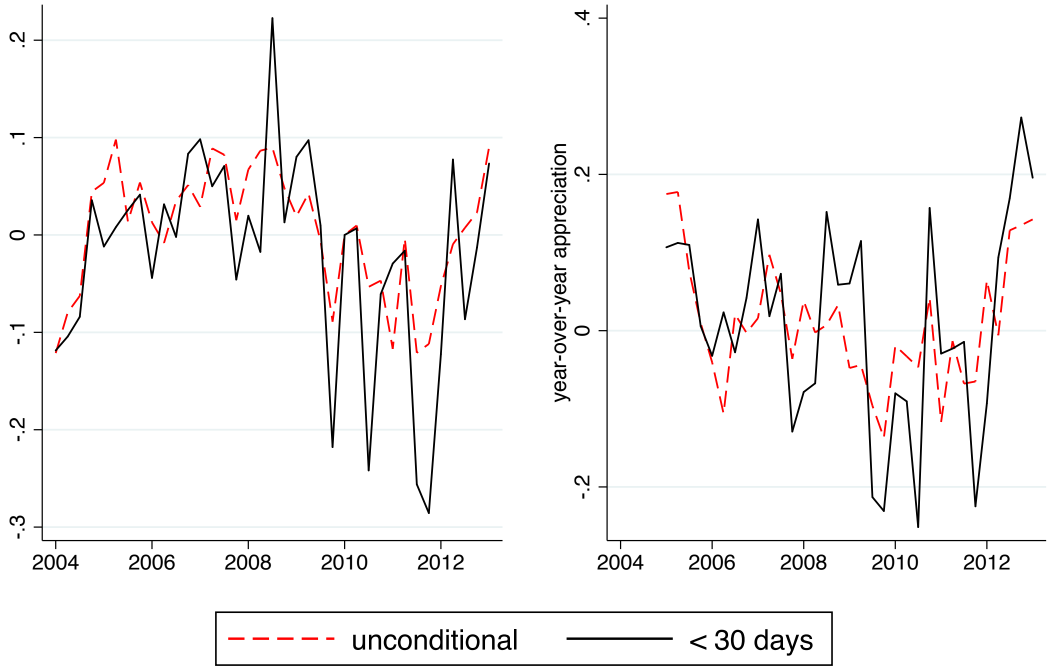
Notes: Panel A computes the Repeat Proportional Hazard Index (RPHI). The index measures relative shifts of the baseline hazard after controlling for unobserved home's heterogeneity. For example, and index value of 1.5 in period t reflects a 50 percent increase in the home sale baseline hazard in period t relative to the base period (2010 q1). The RPHI is compared with a similar index based on a Cox- regression that does not correct for unobserved heterogeneity. Panel B shows the (inverse of) the Repeat Median TOM Index (RMI). We report the inverse of the RMI to facilitate comparison with the RPHI. The RPHI is compared with a simple index that compares the relative shift of the unconditional median TOM in period t relative to the base period.

Figure 8. Repeat sales price indices by duration on the market
Honolulu



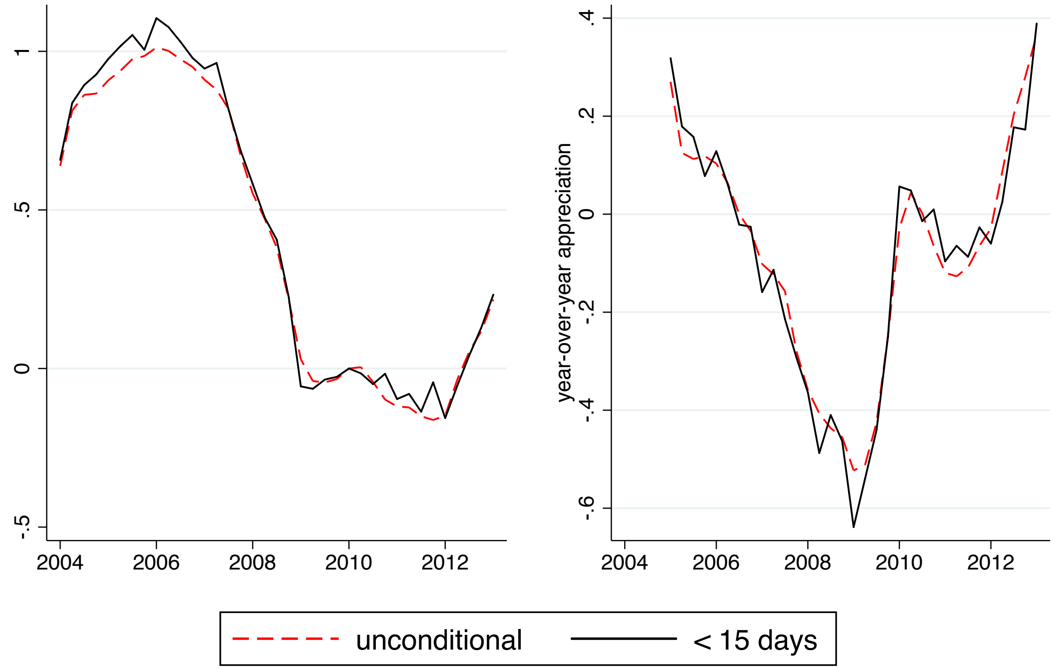
Notes: The left panel computes both an unconditional repeat sales price index and a conditional repeat sales price index. The conditional index is computed conditional on the listing ending within 60 days. The baseline period where the index is normalized to 0 is the first quarter of 2010. The right panel shows the year-over-year changes in these two price indices.

Figure 9. Repeat sales price indices by duration on the market
New Orleans



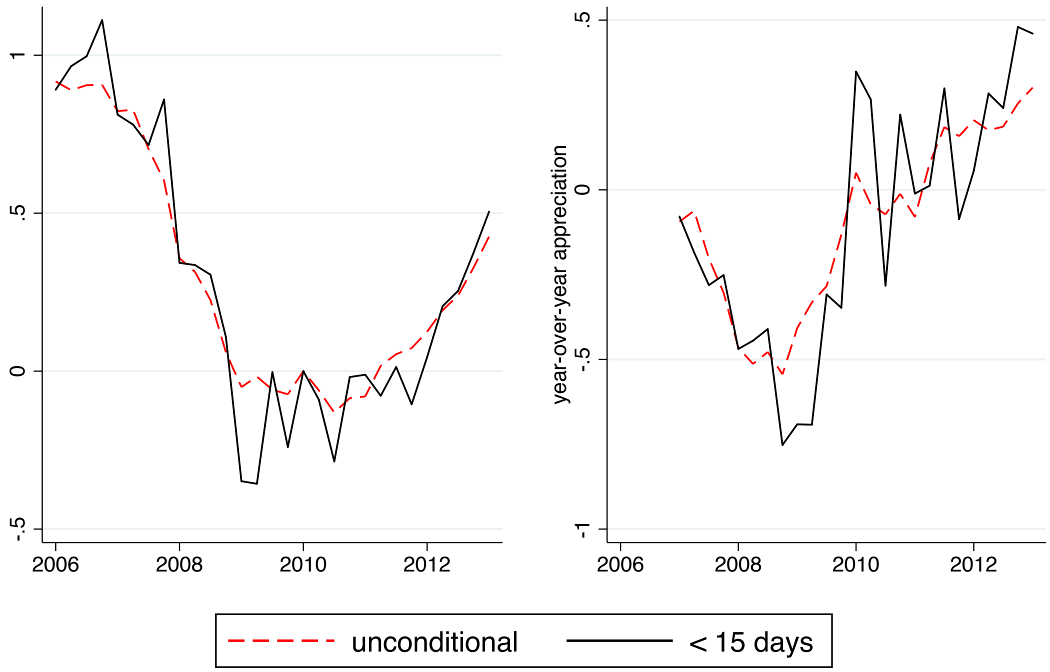
Notes: The left panel computes both an unconditional repeat sales price index and a conditional repeat sales price index. The conditional index is computed conditional on the listing ending within 30 days. The baseline period where the index is normalized to 0 is the first quarter of 2010. The right panel shows the year-over-year changes in these two price indices.

Figure 10. Repeat sales price indices by duration on the market
Las Vegas



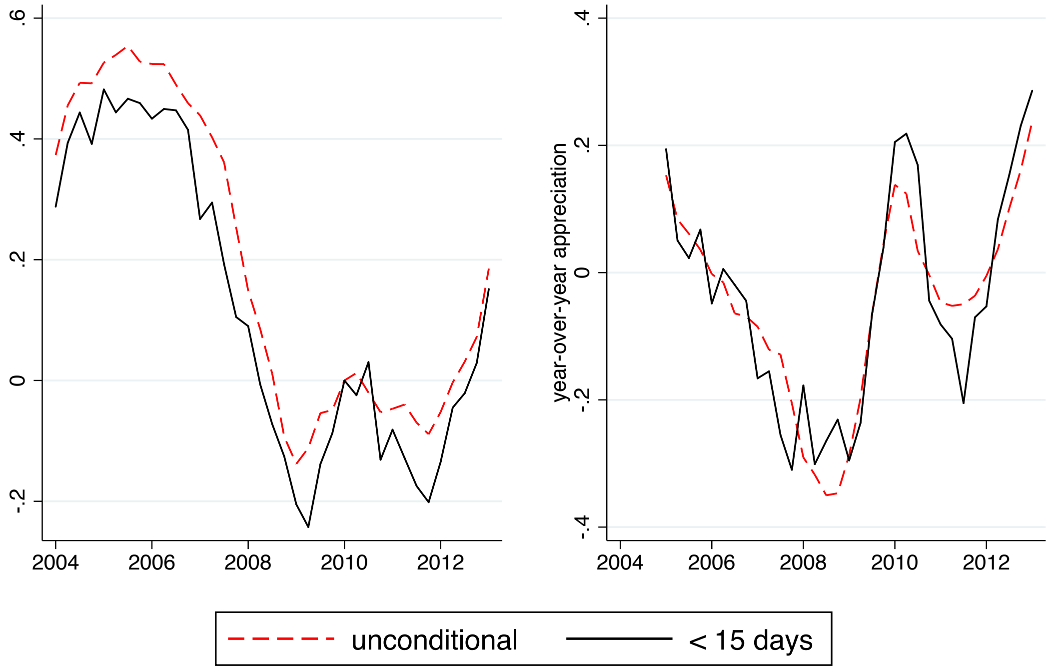
Notes: The left panel computes both an unconditional repeat sales price index and a conditional repeat sales price index. The conditional index is computed conditional on the listing ending within 15 days. The baseline period where the index is normalized to 0 is the first quarter of 2010. The right panel shows the year-over-year changes in these two price indices.

Figure 11. Repeat sales price indices by duration on the market
Miami



Notes: The left panel computes both an unconditional repeat sales price index and a conditional repeat sales price index. The conditional index is computed conditional on the listing ending within 15 days. The baseline period where the index is normalized to 0 is the first quarter of 2010. The right panel shows the year-over-year changes in these two price indices.

Figure 12. Repeat sales price indices by duration on the market San Diego



Notes: The left panel computes both an unconditional repeat sales price index and a conditional repeat sales price index. The conditional index is computed conditional on the listing ending within 15 days. The baseline period where the index is normalized to 0 is the first quarter of 2010. The right panel shows the year-over-year changes in these two price indices.

Table 1:
Geographic and Time Coverage of Sample

	Urban Area	# Obs. All Listings	# Obs. All Home Sales	# Obs. Rep. Listings	Period Begin	Period End
01	Ann Arbor, MI	45,044	21,101	27,642	2004q1	2013q1
02	Boulder, CO	47,177	26,923	27,510	2004q1	2013q1
03	Durham, NC	59,234	34,125	31,112	2004q1	2013q1
04	Fairfax County, VA	357,515	244,961	232,382	2007q2	2010q4
05	Honolulu, HI	85,511	54,350	46,670	2004q1	2013q1
06	Las Vegas-Paradise, NV	262,267	153,577	140,181	2006q3	2013q1
07	Medford, OR	26,138	16,315	13,247	2004q1	2013q1
08	Miami-Miami Beach-Kendall, FL	219,210	112,069	96,982	2005q1	2013q1
09	New Orleans-Metairie-Kenner, LA	79,845	36,869	43,446	2007q3	2013q1
10	Olympia, WA	35,416	21,733	18,351	2004q1	2013q1
11	San Diego-Carlsbad-San Marcos, CA	367,122	207,162	228,676	2004q1	2013q1
12	San Luis Obispo-Paso Robles, CA	27,506	19,044	11,747	2004q1	2013q1
13	Santa Barbara-Santa Maria, CA	29,054	20,085	13,929	2004q1	2013q1
14	Toledo, OH	65,873	35,673	33,599	2004q1	2013q1
15	Youngstown-Warren-Boardman, OH-PA	52,825	27,098	26,099	2004q1	2013q1

Notes: This table tabulates the number of observations in each area we study. The first column shows the total number of real estate listings reported on the MLS during the sample period. The second column shows the sale volume: the number of listings that end up in a sale. Column three shows the number of repeat listings: the number of properties that were listed more than once during the sample period.

Table 2:
Repeat Proportional Hazard Index (RPHI) in Selected US MSAs.

	Geographic Area														
Period	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
2002q1				3.062											
2002q2				2.191											
2002q3				1.355											
2002q4				1.288											
2003q1				2.291											
2003q2				2.156											
2003q3				1.851											
2003q4				2.317											
2004q1	1.425	0.844	1.442	4.064	2.870		2.930			1.636	6.199	3.040	3.878	2.829	1.146
2004q2	1.625	0.843	1.523	2.352	2.723		2.966			2.910	3.212	2.777	2.613	2.595	1.021
2004q3	1.409	0.830	1.376	1.885	2.345		3.317			2.625	1.745	1.708	2.111	1.852	1.094
2004q4	1.295	1.068	1.169	2.454	2.882		3.217			3.731	1.747	1.837	2.708	1.673	1.424
2005q1	1.696	1.300	2.133	3.388	4.122		3.956	0.987		4.228	2.585	2.242	1.967	2.568	1.488
2005q2	1.222	1.125	1.936	1.686	4.085		3.803	0.700		3.492	1.543	1.935	1.690	2.269	1.441
2005q3	0.781	1.020	1.463	0.665	2.833		2.344	0.763		3.682	1.042	1.673	1.278	2.058	1.066
2005q4	0.699	1.236	1.643	0.488	1.643		1.266	1.950		2.361	0.928	1.027	0.885	1.622	1.180
2006q1	0.734	1.208	2.171	0.481	2.052		1.690	2.384		2.897	0.914	1.035	0.950	2.132	1.406
2006q2	0.712	0.985	2.066	0.226	1.845		1.114	1.972		1.969	0.617	0.801	0.703	1.628	1.071
2006q3	0.495	0.843	1.646	0.245	1.020	1.078	0.759	1.619		1.693	0.556	0.870	0.610	1.244	0.938
2006q4	0.481	1.037	1.977	0.359	1.246	0.760	0.862	1.413		1.342	0.650	0.791	0.650	1.140	1.064
2007q1	0.534	1.511	2.270	0.443	1.767	0.276	1.152	1.030		1.774	0.670	1.108	0.748	0.947	1.110
2007q2	0.438	0.990	1.994	0.214	1.388	0.171	1.011	0.693		1.091	0.436	0.784	0.485	0.775	0.983
2007q3	0.561	0.883	1.154	0.135	0.968	0.169	0.580	0.523	1.356	0.945	0.321	0.579	0.465	0.564	0.871
2007q4	0.637	1.209	1.380	0.158	0.794	0.253	0.774	0.472	0.578	0.926	0.375	0.459	0.380	0.706	0.741
2008q1	0.756	1.324	1.318	0.215	0.812	0.417	0.487	0.636	0.692	1.096	0.528	0.637	0.651	0.942	0.968
2008q2	0.650	1.246	1.431	0.278	0.698	0.569	0.751	0.687	0.694	0.813	0.624	0.576	0.667	0.817	0.849
2008q3	0.688	0.914	1.119	0.328	0.527	0.597	0.887	0.865	0.620	0.875	0.811	0.689	0.699	0.777	0.836

Continuation of Table 2:
Repeat Proportional Hazard Index (RPHI) in Selected US MSAs

	Geographic Area														
Period	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
2008q4	0.844	0.793	0.977	0.417	0.497	0.681	0.804	1.098	0.698	0.965	0.983	0.860	1.029	0.846	0.947
2009q1	0.917	0.870	1.145	0.623	0.627	0.894	0.688	1.155	0.942	0.937	0.772	0.910	0.787	1.143	0.944
2009q2	1.003	0.997	1.130	0.858	0.744	1.136	0.928	1.110	0.882	0.827	1.184	0.751	0.921	1.024	0.907
2009q3	1.249	1.064	1.250	0.930	1.033	1.235	1.039	0.972	0.752	1.127	1.216	1.099	0.862	0.942	0.937
2009q4	0.961	0.957	1.216	0.904	1.088	1.116	1.351	0.965	0.808	1.082	1.023	1.115	0.673	0.929	1.265
2010q1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2010q2	0.801	0.812	0.794	0.725	1.017	0.914	0.813	1.184	0.804	0.879	0.781	0.765	0.703	0.909	0.688
2010q3	0.779	0.572	0.713	0.645	0.872	0.840	1.078	1.474	0.802	1.052	0.635	0.905	0.595	0.766	0.657
2010q4	1.080	0.824	1.000	0.648	0.969	0.923	1.517	2.019	1.034	0.841	0.796	0.849	0.626	1.240	1.084
2011q1	1.019	1.030	0.879		1.120	1.157	1.350	2.404	1.430	1.071	0.952	0.977	0.639	1.308	0.990
2011q2	0.939	0.996	0.994		1.108	1.367	1.392	2.064	1.480	0.853	0.909	1.239	0.767	1.188	1.085
2011q3	1.065	0.785	1.091		1.080	1.449	1.419	2.236	1.377	0.869	0.982	1.128	0.885	1.054	0.869
2011q4	1.290	1.093	1.392		1.330	1.689	2.431	2.410	1.525	1.052	1.120	1.126	1.080	1.227	1.334
2012q1	2.002	2.252	1.520		1.453	2.515	2.437	3.072	2.170	1.472	1.643	1.922	1.126	2.018	2.254
2012q2	1.644	2.527	1.345		1.621	2.544	2.625	3.295	2.041	1.388	1.845	1.858	1.749	1.762	1.712
2012q3	1.821	1.805	1.530		2.036	2.168	2.371	3.457	2.114	1.041	1.924	1.864	2.044	1.675	1.516
2012q4	1.987	3.366	2.128		1.807	2.257	2.817	3.377	2.847	1.289	2.666	2.478	2.837	2.463	1.721
2013q1	1.730	3.903	2.485		3.684	2.685	3.101	2.954	2.342	2.515	3.495	3.255	3.193	3.001	1.774

Notes: Notes: The RPHI has been estimated in area 01: Ann Arbor, MI; 02: Boulder, CO; 03: Durham, NC; 04: Fairfax, VA; 05: Honolulu, HI; 06: Las Vegas, NV; 07: Medford, OR; 08: Miami, FL; 09: New Orleans; 10: Olympia, WA; 11: San Diego, CA; 12: San Luis Obispo, CA; 13: Santa Barbara, CA; 14: Toledo, OH; and in area 15: Youngtown, OH. The index measures the (quality adjusted) shift in the baseline hazard relative to the base period (2010q1).

Table 3:
Repeat Median TOM Index (RMTI) in Selected US MSAs.

	Geographic Area														
Period	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
2002q1				0.278											
2002q2				0.417											
2002q3				0.827											
2002q4				0.904											
2003q1				0.444											
2003q2				0.409											
2003q3				0.496											
2003q4				0.404											
2004q1	0.565	0.942	0.841	0.212	0.500		0.572			0.666	0.089	0.318	0.275	0.476	0.893
2004q2	0.565	1.292	0.801	0.374	0.516		0.616			0.519	0.220	0.261	0.301	0.458	0.751
2004q3	0.804	1.403	0.867	0.488	0.551		0.511			0.605	0.450	0.478	0.447	0.724	0.905
2004q4	0.311	1.141	1.083	0.381	0.493		0.712			0.450	0.521	0.452	0.455	0.800	0.814
2005q1	1.139	0.982	0.772	0.228	0.396		0.453	0.893		0.504	0.347	0.413	0.466	0.582	0.806
2005q2	0.766	1.100	0.712	0.503	0.362		0.525	0.872		0.469	0.564	0.404	0.501	0.496	0.897
2005q3	1.097	1.180	0.918	1.553	0.469		0.644	0.602		0.454	0.856	0.549	0.669	0.582	1.001
2005q4	1.380	1.243	0.946	2.415	0.607		0.884	0.473		0.657	0.955	0.850	0.754	0.677	0.868
2006q1	1.093	1.026	0.767	2.697	0.586		0.793	0.333		0.561	1.031	0.859	1.040	0.722	0.869
2006q2	1.038	1.096	0.705	5.079	0.638		0.942	0.436		0.630	1.414	1.073	1.293	0.868	0.844
2006q3	1.367	1.346	0.885	5.336	0.926	0.933	1.168	0.556		0.624	1.561	1.104	0.908	0.738	1.084
2006q4	1.314	1.265	0.767	4.093	0.752	1.490	1.126	0.584		0.665	1.555	1.214	2.013	1.170	0.828
2007q1	1.271	0.915	0.694	3.429	0.698	2.866	0.980	0.857		0.634	1.633	0.842	1.411	1.414	0.986
2007q2	1.295	1.054	0.610	6.453	0.791	3.329	0.965	0.932		0.943	1.993	1.001	1.195	1.439	0.962
2007q3	1.390	0.982	0.768	7.973	0.988	3.705	0.914	1.279	0.815	0.700	2.672	1.389	1.589	0.991	1.048
2007q4	1.531	1.437	0.779	6.575	0.911	3.524	1.084	1.212	1.003	1.065	2.202	1.348	2.738	1.649	0.995
2008q1	1.507	1.027	0.867	6.252	1.057	3.337	0.811	1.285	1.025	0.806	2.060	1.125	1.720	0.844	1.104
2008q2	1.128	0.976	0.607	5.371	1.096	2.654	0.901	1.145	0.977	0.858	1.715	1.254	1.188	0.958	1.019
2008q3	1.599	1.156	1.374	4.501	1.422	2.548	1.462	1.092	1.210	0.831	1.500	1.220	1.005	1.072	0.964

Continuation of Table 3:
Repeat Median TOM Index (RMTI) in Selected US MSAs

	Geographic Area														
Period	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
2008q4	0.762	1.217	0.928	3.545	1.191	2.030	1.911	1.049	1.223	0.802	1.143	0.954	1.095	1.066	1.026
2009q1	2.270	1.241	0.976	1.838	1.144	1.532	1.504	0.842	1.212	1.142	1.397	1.143	1.088	1.123	0.926
2009q2	1.220	1.035	1.011	1.228	1.124	0.935	1.112	0.981	1.207	0.937	1.010	1.063	1.013	1.010	0.919
2009q3	1.155	1.259	0.994	1.153	0.968	0.845	1.109	0.987	1.102	1.075	0.939	0.859	0.889	0.991	0.886
2009q4	1.952	1.267	1.255	1.202	0.889	0.979	1.037	0.931	1.214	1.044	1.049	0.807	1.679	1.001	1.206
2010q1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2010q2	1.324	0.887	1.044	1.427	1.061	1.192	1.167	0.995	1.169	0.832	1.343	1.084	1.291	0.923	1.107
2010q3	1.077	2.083	1.179	1.745	0.976	1.326	1.096	0.963	1.206	1.415	1.521	1.176	1.797	1.210	1.168
2010q4	2.074	1.082	1.179	1.414	0.746	1.086	1.064	0.678	0.871	1.054	1.343	1.463	1.432	0.983	1.015
2011q1	1.271	1.401	1.343		1.061	0.905	1.261	0.589	0.948	1.289	1.086	0.841	1.510	0.959	1.221
2011q2	1.154	1.240	1.121		1.012	0.806	1.170	0.587	0.974	1.273	1.177	0.764	1.311	1.004	1.044
2011q3	1.453	1.230	1.194		0.896	0.648	1.001	0.594	0.994	1.378	1.121	0.945	1.203	1.073	1.126
2011q4	1.120	1.510	1.137		0.796	0.533	0.952	0.507	0.812	1.293	0.882	1.166	1.105	1.028	1.116
2012q1	1.151	0.958	1.013		0.766	0.294	0.802	0.408	0.685	1.077	0.658	0.750	1.207	0.946	0.925
2012q2	0.931	0.724	1.031		0.772	0.289	0.732	0.354	0.739	0.861	0.497	0.545	0.646	0.950	0.911
2012q3	0.784	0.722	0.964		0.687	0.372	0.780	0.322	0.679	1.111	0.507	0.513	0.585	0.955	0.778
2012q4	0.970	0.928	1.174		0.685	0.363	0.623	0.314	0.574	1.283	0.335	0.553	0.502	0.809	0.914
2013q1	0.909	0.551	0.656		0.412	0.232	0.512	0.237	0.390	0.728	0.203	0.345	0.251	0.598	0.558

Notes: Notes: The RMTI has been estimated in area 01: Ann Arbor, MI; 02: Boulder, CO; 03: Durham, NC; 04: Fairfax, VA; 05: Honolulu, HI; 06: Las Vegas, NV; 07: Medford, OR; 08: Miami, FL; 09: New Orleans; 10: Olympia, WA; 11: San Diego, CA; 12: San Luis Obispo, CA; 13: Santa Barbara, CA; 14: Toledo, OH; and in area 15: Youngtown, OH. The index measures the shift in the (quality adjusted) median TOM relative to the base period (2010q1).

Table 4:
ADL Summary

Model 1:	Positive, p<0.1	Positive, p>0.1	Negative, p<0.1	Negative, p>0.1
RPHI	10	4	0	0
Model 2:	Positive, p<0.1	Positive, p>0.1	Negative, p<0.1	Negative, p>0.1
RMTI	8	5	0	1

Notes: For each of the 14 CBSAs we estimated an autoregressive distributed lag model of home price appreciation on its 3 lags and the lag of a TOM index. This was done separately using both of our TOM indices. The first row shows the results from using the RPHI and the second row shows the results from using the RMTI. The table summarizes the sign and significance of the coefficient on the TOM index across the 14 CBSAs.

Table 5:
Panel ADL

	(1)	(2)	(3)
RPHI		0.006***	
		(0.002)	
RMI			0.026***
			(0.006)
Price appreciation, first lag	1.391***	1.359***	1.342***
	(0.05)	(0.05)	(0.049)
Price appreciation, second lag	-0.384***	-0.383***	-0.384***
	(0.081)	(0.08)	(0.079)
Price appreciation, third lag	-0.088	-0.069	-0.069
	(0.046)	(0.046)	(0.045)
CBSA fixed effects?	yes	yes	yes
R-squared	0.934	0.936	0.937
Observations	419	419	419

Notes: This table reports the results of three panel ADL regressions estimated using data from all 14 CBSAs, where the dependent variable is home price appreciation. In the first column we did not include a TOM index and in columns (2) and (3) we included the RPHI and the RMTI.

Table 6:
Forecast Analysis

	In-sample RMSE	1-step forecast RMSE	1-step realiz. in 90 % PI	2-step forecast RMSE	2-step realiz. in 90 % PI
Model 1:					
univariate forecast	0.019	0.033	0.836	0.06	0.672
Model 2:					
RPHI	0.017	0.031	0.761	0.057	0.693
Model 3:					
RMTI	0.018	0.03	0.811	0.055	0.714

Notes: For each of the 14 CBSAs we estimated a VAR model for home price appreciation and TOM with one lag and used this to forecast home price appreciation. For each period, starting with Q4 of 2008 we recalculated "real-time" TOM indices using home-level data only up to that date. We then estimated the VAR using only the TOM and price indices from that date back to Q1 2004 and used this to forecast one and two periods ahead. This was done separately using both the RPHI and the RMTI. We also estimated a univariate model using home price appreciation only. The first column reports, for comparison purposes, the RMSE of the VAR estimates across the sample used to estimate the VAR. The second and fourth columns reports the RMSE of the one- and two-quarter ahead forecasts. Along with the forecasts, we also computed 90% prediction intervals. The third and final columns report how often the realized rate of home price appreciation fell within the prediction interval.

Table 7:
ADF tests - summary across CBSAs

	Levels	First Difference	Second Difference
log price	0	1	14
log of RPHI index	0	10	14
log of RMTI index	2	12	14

Notes: For each of the 14 CBSAs we carried out an augmented Dickey-Fuller (ADF) test on the levels, first differences, and second differences of each of the series. The tests allowed for an intercept and a linear trend and 2 lags. Each series contains 37 observations. The table reports for how many CBSAs we rejected the null of a unit root.

Table 8:
Cointegration tests - summary across CBSAs

	(1)	(2)
log of RPHI index	8	14
log of RMTI index	9	14

Notes: For each of the 14 CBSAs we estimated the regressions described in the text. Column (1) corresponds to the regression using first differences in log price and levels of log TOM and column (2) corresponds to the regression using second differences in log price and first differences in log TOM. Then we carried out an augmented Dickey-Fuller (ADF) test on the residuals. The ADF tests allowed for an intercept and a linear trend and 2 lags. Each series contains 37 observations. The table reports for how many CBSAs we rejected the null of no cointegration.