**Institute for International Economic Policy Working Paper Series**
**Elliott School of International Affairs**
**The George Washington University**

# The Repeat Time-On-The-Market Index

**IIEP-WP-2015-8**

**Paul E. Carrillo**
**George Washington University**

**Benjamin Williams**
**George Washington University**

**February 2019**

Institute for International Economic Policy
1957 E St. NW, Suite 502
Voice: (202) 994-5320
Fax: (202) 994-5477
Email: **iiep@gwu.edu**
Web: **iiep.gwu.edu**

# The Repeat Time-On-The-Market Index*

Paul E. Carrillo†
George Washington University

Benjamin Williams‡
George Washington University

February 2019

**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heterogeneity exploiting *repeat* listings, b) their use of censored durations (listings that are expired and/or withdrawn from the market), and c) their computational simplicity.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 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 it is important to account for both censoring and unobserved heterogeneity in measuring housing market liquidity.

Keywords: Non-parametric models, proportional hazard, repeat sales
JEL Codes: C41, R31

*We would like to thank the Editor and two anonymous referees for their thoughtful suggestions. In addition, we thank John Clapp, Andra Ghent, Bryan Stuart, Scott Wentland, Anthony Yezer and participants at the Homer Hoyt Institute Workshop, the Richmond Fed Regional Workshop, the ASSA-AREUEA Annual Conference, and at seminars at BEA, CoreLogic, FHFA, Freddie Mac, George Washington, UNC-Charlotte and University of Georgia for helpful comments. 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 housing market clears in two dimensions: price and marketing time. Information asymmetries, transaction costs, costly matching between sellers of differentiated products and buyers with heterogenous tastes, and optimal buyers' and sellers' strategies determine both the sale price of a housing unit as well as its "time on the market" (TOM), the time it takes until it is sold.[1] To assess housing market conditions it seems important to track the evolution of both of these dimensions.

The literature has made significant progress developing alternative approaches to measuring real estate prices. Construction of home price indices include rigorous methods that use hedonic regressions, matching techniques, and repeat sales, among others. Several studies have analyzed and compared the properties of alternative price indices[2] and, perhaps due to its ability to control for unobserved housing heterogeneity, the classical approach to constructing home price indices in the U.S. focuses on *repeat* sales (Bailey et al., 1963; Case and Shiller, 1987).[3] Surprisingly, specification and estimation of housing liquidity indices have received much less attention. It is clear that TOM can measure housing liquidity risk (Lippman and McCall, 1986; Lin and Vandell, 2007), but very little is known about the properties of potential indices that track TOM's evolution.[4] This is probably the reason why there are no official indices that measure how TOM changes over time.

Given the prevalent approach to measuring prices using repeat sales, developing methods to measure the evolution of TOM while controlling for unobserved heterogeneity (i.e., by exploiting repeat sales) seems a first order issue. However, Carrillo and Pope (2012) show that expired and withdrawn listings (i.e., censored durations) are a common feature of real estate markets and that censoring drastically changes with market conditions: it remains low during housing booms and peaks during busts. Hence, any attempt to measure the distribution and evolution of TOM should explicitly take censoring into account. Measuring the evolution of TOM while accounting both for unobserved heterogenity and censored durations is not a trivial task and is the main objective and contribution of this paper. To track how TOM evolves over time we develop new indices

[1]A long and growing literature in urban economics that studies the microstructure of housing transactions has examined these issues. Han and Strange (2015) present a comprehensive survey of this topic; examples of recent studies include Carrillo (2012) and Merlo et al. (2015).

[2]The real estate price index literature is quite extensive; Eurostat (2013) presents a useful summary of these approaches.

[3]Home price indices in the U.S. that employ repeat sales include the Federal Housing Finance Agency (FHFA) quarterly price index and the S&P CoreLogic Case-Shiller index.

[4]Other papers have estimated TOM indices including Carrillo and Pope (2012) and Liu et al. (2016).

that a) control for unobserved heterogeneity by exploiting *repeat* listings and therefore can be directly compared with conventional and widely available repeat home price indices, b) account for censored durations, and c) are computationally straightforward. The paper brings a methodological innovation that has the potential to be used in a number of practical applications.

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 list properties along with 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 *conventional* statistics such as median or mean TOM. 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.[5] Expired and withdrawn listings, however, are common in all markets we study. For instance, in a suburb of Washington DC (Fairfax County, VA) as many 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 CBSAs in the U.S. that we analyze.[6] Hence, as it was the case in Carrillo and Pope (2012)'s application, any attempt to measure the evolution of TOM in our sample should account for censored durations. However, when some durations are censored, unobserved home heterogeneity cannot be differenced out using the classical repeat sales method.

We propose two models that can be used to correct for unobserved heterogeneity using *repeat* listings while, at the same time, accounting for censored durations. The first index, the "repeat proportional hazard index" (RPHI), is based on a proportional hazard specification and applies an

[5]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.

[6]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.

estimator developed by Ridder and Tunali (1989) to compute proportional displacements in the home sale hazard rate. The second index, the "repeat median TOM index" (RMTI), 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. These indices are based on two differencing strategies that adapt the conventional repeat sales method to account for the presence of random censoring.

We compute the RPHI and RMTI indices for TOM in each of the 15 CBSAs we study. We emphasize that our methods are computationally straightforward and can be easily implemented even with millions of observations. 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) significantly affects the estimates of changes over time in TOM.

Finally, we use our novel TOM indices to reassess three previous empirical findings in the literature on housing liquidity. First, consistent with the findings of Diaz and Jerez (2013), we find a high within-market correlation between our TOM indices and sales volumes. Second, we report stark seasonality patterns in our indices that are similar in spirit to those reported by Ngai and Tenreyro (2014). Third, as documented by Genesove and Han (2012), we illustrate a close relationship between TOM and market fundamentals – income, population, unemployment, and interest rates. In each case, we find important differences with results obtained using a conventional measure of TOM, though generally the results are qualitatively the same. We expect that the TOM indices developed in this paper will be 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 paying particular attention to censored observations. In Section 4 we develop our two novel TOM indices. Section 5 revisits a few stylized findings in the literature on housing liquidity. In Section 6 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.[7] 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, 1985; 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 models. Ridder and Tunali (1999) propose a stratified partial likelihood in a proportional hazards model.[8] 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. We extend this insight by showing that in our model of repeat listings the method is equivalent to a "repeat listings logit regression." This is a novel contribution of our paper, which allows the estimation of the RPHI. A second solution was proposed by Honoré et al. (2002) who adapt a method for fixed censoring (Honoré, 1992) by integrating over the distribution of the censoring variable. We modify and extend their approach

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

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

to construct the RMTI. 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). In a related paper, Carrillo and Pope (2012) provide an alternative measure of liquidity by estimating shifts in the distribution of TOM across time. They take censoring into account but ignore unobserved heterogeneity. Measuring the evolution of TOM while accounting both for unobserved heterogeneity (by focusing on repeat listings) and censored durations is the key contribution of our paper.

## 3 Data and Conventional Descriptive Statistics

Our data come from two sources. Metropolitan and Regional Information Systems (MRIS) provided us with MLS data from Fairfax County, VA.[9] This dataset contains information for all housing listings in this county that appeared 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 statistics at any level of geographic aggregation. More importantly, we can track if the same property is listed and/or sold in multiple periods.

Our second source of data is CoreLogic Solutions, LLC (CoreLogic). CoreLogic collects MLS data from more than 100 CBSAs, 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 CBSAs that were posted on the MLS between January 2004 and February 2013. The CBSAs in our sample include large and medium urban areas in the East, Midwest, and Western regions of the U.S.[10] CoreLogic data include information about pricing and the specific

[9]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.

[10]The CBSAs we analyze include Ann Arbor, MI, Boulder, CO, Durham, NC, Honolulu, HI, Las Vegas-Paradise,

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.

Following Carrillo and Pope (2012) 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. CBSAs. A list of the urban areas, the number of listings, and a description of the sample period is available in Table 1.[11] About 58 percent of all listings in the overall sample result 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 listings of such properties *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. Implementation of this definition, while straightforward, requires making certain ad-hoc choices. First, one needs to recognize that properties are sometimes listed many times without a sale. Multiple consecutive listings of the same home without an intervening sale are considered a single listing as long as the difference between the date when the last listing was withdrawn from the market and the date when the new listing was posted is less than two months.[12] The measure of TOM is based on the *cumulative* number of days that a property is for sale during such a set of consecutive listings. We can take this approach because we observe unique property identifiers in both datasets.[13] We also need to make an ad-hoc choice when choosing our sample of *repeat* listings. If two listings of the

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.

[11] Our data include listings of foreclosed units and "short sales" that were advertised on the MLS.

[12] When constructing cumulative TOM for these listings we do not include the time between consecutive listings when the property is off the market. The two month threshold is arbitrary but the indices barely change when we modify it.

[13] Most MLS Systems also provide information about the number of days that the listing and the property have been on the market. These two numbers are not the same when a property has been listed multiple times.

same home were not considered a single listing using the above criteria but were still close together (less than one year between list dates), then we dropped this pair of listings from the analysis.[14]

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 as the volume of sales. Importantly, the mean and median TOM are calculated just for the sample of properties listed in each quarter that result in a sale.[15] The top panels of Figures 1 and 2 present these statistics for Fairfax County, and five representative CBSAs from our U.S. areas: Las Vegas, San Diego, Miami, Honolulu and New Orleans.[16] 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. It is surprising 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.

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 larger fraction of listings are withdrawn from the market. These observations suggest that accounting for censored durations may be consequential for estimation of TOM displacements.

We are not the first to make this point. To account for censoring, Carrillo and Pope (2012) estimate the unconditional distribution of TOM, using a Kaplan-Meier estimator where withdrawn and/or expired listings are treated as censored observations. We follow their approach to adjust the conventional statistics computed above. When a listing is withdrawn from the market or it expires without a sale, we compute the censored duration as the number of days between when the

[14] Failure to properly identify a) multiple consecutive listings as unique listings and b) the sample of repeat listings can theoretically bias the indices.

[15] 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.

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

listing was first posted and when it was withdrawn or expired. For each urban area, we calculate 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 compute 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 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.

Another way to adjust for 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 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 two ways: using only listings that ended in a sale, 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. [17]

In sum, conventional estimates that measure the evolution of TOM are misleading. Any statistic designed to track TOM should account for censored observations.

[17] 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.

# 4 Measuring TOM Using Repeat Listings

As discussed in the introduction, the classical approach to measure home prices is based on repeat sales. Using repeat sales allows researchers to estimate the evolution of home prices while controlling for each unit's heterogeneity (both observed and unobserved). The intuition is straightforward: any observed and/or unobserved heterogenity that is constant over time is "differenced out" when computing each unit's specific appreciation rate (Bailey et al., 1963; Case and Shiller, 1987). Perhaps due to its simplicity, the repeat sales approach is the leading method used in the U.S. to estimate home price indices (see, for example FHFA and S&P Corelogic Case-Shiller price indices). Given that price indices focus on repeat sales, developing a methodology to construct a comparable index that measures the evolution of TOM seems a first order concern.[18]

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, 1985; 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 of observations is very large (hundreds of thousands of observations). Our goal in this section is to propose housing TOM 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 estimate the evolution of TOM while correcting for unobserved heterogeneity using *repeat* listings. The main intuition is straightforward: Just as it is the case with repeat sales home price indices, one can use the TOM of properties that have been on the market in more than one period to control for and "difference out" the unobserved heterogeneity. This is precisely what Liu et al. (2016) do to compute a TOM index that focuses on

[18] An alternative to a TOM index that is based on repeat sales is an index that controls for *observed* heterogeneity (such as square footage, number of bedrooms and age of the unit.) Carrillo and Pope (2012) perform this approach to estimate changes in the distribution of TOM across time. We have applied these methods and computed such index for Fairfax County, the only area for which we have detailed housing unit's characteristics. Results, which are available upon request, suggest that controlling for observed heterogeneity does not affect the estimate of median TOM.

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.

To facilitate further exposition, let $Y_i^s, C_i^s$, and $t_i^s$ denote the TOM (time until the home is sold), censoring time (time until the home is withdrawn from the market), and listing period (quarter) for the $s^{th}$ listing of home $i$ ($s = 1, 2$).[19] It is also useful to define the duration of a marketing spell of listing $s$ for home $i$, $V_i^s$, which could end after the home is sold or after it is withdrawn from the market: $V_i^s = \min\{Y_i^s, C_i^s\}$. Furthermore, let $d_i^s$ take the value of one if the $s^{th}$ listing of home $i$ is not censored (i.e., the home was eventually sold) and zero otherwise: $d_i^s = \mathbf{1}(Y_i^s \leq C_i^s)$.

## 4.1 The Repeat Proportional Hazard Index

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 a home specific "fixed effect." This motivates a repeat sales logit regression method that we use to construct a repeat proportional hazard index. 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" (RPHI) using a computationally straightforward procedure. The RPHI provides an estimate of the gross percentage increase in the hazard rate relative to a base period.

Formally, 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 $t$.

$$\lambda_{it}(y) = \exp(\beta_t)\lambda_{0i}(y) \tag{4.1}$$

The first factor, $\exp(\beta_t)$, allows the hazard function to vary proportionately depending on the period

[19] 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.

listed and the second factor, $\lambda_{0i}(y)$, is a *home-specific* baseline hazard function. We normalize $\beta_{t_0} = 0$ for a baseline period $t_0$ so that $\lambda_{0i}(y)$ 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. These shifts are fully captured by the parameter $\beta_t$.

Assume that the baseline hazard can be written as

$$\lambda_{0i}(y) = \exp(\alpha_i)\lambda_0(y),$$

and note that unobserved heterogeneity in homes is fully captured by the parameter $\alpha_i$. This coefficient is a unit specific fixed effect that captures observed and unobserved features of units that shift the market hazard rate $\lambda_0(y)$. Intuitively, $\alpha_i$ is a time invariant fixed effect that is unique to each home $i$ and that determines the likelihood that it sells at any given period.

The structure of the proportional hazards model allows us to "difference out" the unobserved heterogeneity using repeat listings of the same home. Let $\Lambda_{0i}(y) := \int_0^y \lambda_{0i}(s)ds$ denote the baseline integrated hazard function. It follows then that

$$-\log(\Lambda_{0i}(Y_i^s)) = \beta_{t_i^s} + \alpha_i + \varepsilon_i^s \tag{4.2}$$

with $-\varepsilon_i^s \mid t_i^s, \lambda_0(\cdot) \sim EV1(0,1)$ 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).[20] Notice that the integrated hazard function in equation 4.2 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 $\log(\Lambda_{0i}(Y_i^2)) \geq$

[20] 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 $\Lambda_Z$ 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.

$\log(\Lambda_{0i}(Y_i^1))$ and the probability of the latter event is

$$Pr\left(\log(\Lambda_{0i}(Y_i^2)) \geq \log(\Lambda_{0i}(Y_i^1)) \mid t_i^1, t_i^2\right) = Pr(\varepsilon_i^2 - \varepsilon_i^1 \leq \beta_{t_i^1} - \beta_{t_i^2} \mid t_i^1, t_i^2) \tag{4.3}$$

$$= \frac{\exp(\beta_{t_i^1})}{\exp(\beta_{t_i^2}) + \exp(\beta_{t_i^1})} \tag{4.4}$$

since the difference of two type 1 independent extreme value distributions has a logistic distribution. Notice that when computing this probability, the fixed effect $\alpha_i$ was successfully "differenced out."

The result in equation 4.3 was derived assuming that there was no censoring in either period. The approach, however, also works in the general case when there are censored durations. If the two listings shared a common censoring time, $C_i^1 = C_i^2$, then the probability that $V_i^2$ exceeds $V_i^1$ conditional on at least one of these durations being uncensored is again given by the logit formula in equation 4.3:

$$\frac{\exp(\beta_{t_i^1})}{\exp(\beta_{t_i^2}) + \exp(\beta_{t_i^1})}.$$

Notice that the logit functional form here is implied by the proportional hazards model and the assumption of independent censoring; it is not an additional assumption.

While repeat listings are not censored by a common length, we can force a common censoring time on the data in order to use this powerful result. 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 $s = 1, 2$. Then if at least one of the two durations is uncensored by $\tilde{C}_i$, that is, if $\tilde{d}_i^1 + \tilde{d}_i^2 > 0$, then $V_i^2 \geq V_i^1$ holds if and only if $\tilde{V}_i^2 \geq \tilde{V}_i^1$ holds. Moreover, the condition $\tilde{d}_i^1 + \tilde{d}_i^2 > 0$ is equivalent to the condition that either (a) both listings are uncensored ($d_i^1 = d_i^2 = 1$) or (b) one listing is censored and this listing's censoring time exceeds the observed time on the market for the other listing. Therefore,

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

where $W_i$ is a dummy variable indicated those homes for which either condition (a) or (b) above holds. The left-hand side is what we observe in the data, and the right-hand side takes the logit form, $\frac{\exp(\beta_{t_i^1})}{\exp(\beta_{t_i^2})+\exp(\beta_{t_i^1})}$. Therefore,

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

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 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 4.1.**

*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 4.1. 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, $\lambda_{0i}$. 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 $t_i^2$.[21] The second condition

[21] 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.

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.

#### 4.1.1 Implementation: Proportional Hazard Index

The procedure to estimate the coefficients $\beta_t$ is straightforward and can be summarized as follows.

- Step 1: Identify the relevant sample of repeat listings. We first restrict our attention to homes with repeat listings. Among this sample of repeat listings, we select those properties with completed durations in both periods or if only the larger of the two durations is censored. This is the subsample for which $W_i = 1$.
- Step 2: Calculate the dependent variable. Using the sample defined in the previous step, we construct an indicator that takes the value of 1 if $V_i^2 \geq V_i^1$, 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 $X_i$.

#### 4.1.2 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 CBSAs in the sample and results are displayed in Table 2. In all areas the index has been 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.[22]

[22] This variation is statistically significant as well. We can compute standard errors for $\hat{\mu}_t = \exp(\hat{\beta}_t)$ by applying the delta method to the conventional standard errors for the logit regression coefficients. Standard errors for all CBSAs are available from the authors on request.

It is useful to compare the RPHI with the simple unconditional hazard ratio computed in Section 3. The hazard ratio discussed in Section 3 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 than 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.

### 4.2 A Repeat Median TOM Index

The second index 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 (log) TOM ($Y$) of home $i$ in quarter $t$ as a function of period-specific shifters $\beta_t$ as well as a home-specific unobserved heterogeneity term $\alpha_i$, which is assumed to be time-invariant

$$\log(Y_{it}) = \beta_t + \alpha_i + \varepsilon_{it}. \tag{4.5}$$

This model allows the distribution of $\varepsilon_{it}$ to be unrestricted, and the hazard rate in a given period is accelerated (or decelerated) relative to the baseline period, rather than just shifted proportionately (see, e.g., Kalbfleisch and Prentice, 2011).[23] The proportional hazard and the AFT models are not nested but coincide when the baseline hazard functions are constant.

[23] The model of equation (4.5) is derived from a hazard function of the form $\lambda_{it}(y) = \exp(\beta_t)\lambda_{0i}(\exp(\beta_t)y)$.

In the accelerated failure time model, the home fixed effect, $\alpha_i$, can be differenced out so that

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

where $X_i$ is as defined in Section 4.1 and $\tilde{\varepsilon}_i = \varepsilon_i^2 - \varepsilon_i^1$, suggesting the use of the usual fixed effects regression estimator if there were no censoring.

A similar approach is still possible in the presence of censoring. First, notice that the median TOM in a particular quarter, $Med(\log(Y_i^s) \mid X_i)$, can generally be identified when there are censored observations, as described in Section 3, via the method of Kaplan and Meier (1958). If we assume that once you control for the period specific shifter $\beta_t$ the median of TOM is stationary: $Med(\alpha_i + \varepsilon_i^1 \mid X_i) = Med(\alpha_i + \varepsilon_i^2 \mid X_i)$, then

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

Note that by taking the difference in the medians, we have "differenced out" the fixed effect $\alpha_i$ in equation 4.5. Our repeat median TOM index is based on this simple idea.

Before detailing the construction of the repeat median TOM index we state and discuss the assumptions required for this method to consistently estimate $\beta$.

**Assumption 4.2.**

*(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* $t_i^1, t_i^2$

Condition (i) is a conditional median stationarity condition that has been used in other econometric models with censoring (e.g., Khan et al., 2011). It combines a notion of strict exogeneity condition similar to Assumption 4.1(i) with a stationarity condition. The assumption holds as long as a) the selection of the listing period is not related to any other unobservable factors, and b) displacement in the conditional median can only be due to changes in $\beta$; the rest of the environment is stationary. Condition (ii) is a conventional independent censoring assumption.[24]

[24] Notice, however, that this is notably different from condition (iii) of Assumption 4.1 in that it is not conditional on unobserved home heterogeneity. For instance, if there are unobserved home characteristics that affect both the

This solution to censoring in the model of equation (4.5) 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.

### 4.2.1 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 construct the repeat median TOM index using the following procedure.

- 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 $\beta$.

Tthis 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.

For some pairs of listing periods, there is a value $\bar{y}$ such that all observations with durations exceeding $\bar{y}$ are censored and more than half of the listing durations exceed $\bar{y}$. In this case $Med(\log(Y_i^s) \mid t_i^1, t_i^2 = t_1, t_2)$ can not be estimated from the Kaplan-Meier survival curve. Since identification of $Med(\log(Y_i^s) \mid t_i^1, t_i^2 = t_1, t_2)$ for every pair of listing periods $t_i^1$ and $t_i^2$ is not necessary in order to identify $\beta$, we can safely exclude these observations from the regression in Step 2.

time to sale ($Y$) and the time until the listing is withdrawn ($C$) then condition (ii) of 4.2 is not satisfied. However, if these unobserved home characteristics are time-invariant then they are captured by alpha and condition (iii) of Assumption 4.1 is still satisfied.

### 4.2.2 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 $\exp(\beta_t)$ represents 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 CBSAs 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.[25]

In the bottom panels of Figures 5 and 6 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.

We compare the RMTI with a simple ratio of unconditional medians. The unconditional medians have been estimated using the Kaplan-Meier estimator and include censored durations but do not account for unobserved housing 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 was the case with the repeat proportional hazard index, controlling for unobserved heterogeneity is important in some markets when measuring the evolution of the median TOM.

[25] This variation is statistically significant as well. While an analytical formula for standard errors is possible, we instead suggest bootstrapping the standard errors. Standard errors for all CBSAs are available from the authors upon request.

## 4.3 Discussion

### The role of unobserved heterogenity

Controlling for unobserved heterogeneity has been shown to be relevant when estimating home price indices. As shown in the first panel of Figure 7, this seems to also be the case in our sample (for space constraints we show results for 3 urban areas). Panel A of this figure compares an unconditional mean (log) price index (using all sales) with a repeat sales index; both price indices are normalized to take the value of zero in the first quarter of 2010. The difference between these two indices, which is plotted in the bottom panel of the figure, depends on the housing cycle. For example, in Fairfax County, the repeat sales index is higher than the unconditional index when the market is tight (2004-2005) and vice versa. Differences between median, hedonic and repeat sales price indices can be explained by a variety of factors and have been widely documented in the literature (Wallace and Meese, 1997, for example).

Because the most relevant price indices available in the U.S. are based on repeat sales, it seems crucial to use a comparable measure when analyzing changes in liquidity. To illustrate this statement, in the second panel of Figure 7 we compare our (log) RPHI, which is based on repeat listings, with the unconditional index, which is estimated using all listings in the sample.[26] The differences between these indices also appear to depend on the housing cycle. More importantly, in the bottom panel of this figure, we clearly see that the gap between the two price indices (top panel) is positively correlated with the gap between the two TOM indices (bottom panel). [27] This suggestive evidence highlights the importance of developing a methodology to estimate displacements of TOM using *repeat* listings.

### 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

[26] The unconditional index is estimated from the sample of all listings using the Cox partial likelihood estimator for the proportional hazards model, accounting for censored durations.

[27] To formally test the significance of this relationship we have estimated a regression of the gap between the two prices indices on the gap between the two TOM indices. The results evidence a strong and positive correlation in almost all cities. Because the TOM indices measure the TOM for homes *listed* in a given period, while the price indices measure prices for homes *sold* during a given period, we might also expect the TOM gap to lag the price gap. For each MSA we also estimated five separate regressions to explore sensitivity to the choice of the lag. In 29 out of the 30 regressions the sign of the estimated coefficient is positive. Results are available upon request.

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.[28] 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 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: both those resulting in sales as well as those 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 to markets participants. 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.

**Controlling for sellers' idiosyncratic characteristics**

Does the seller's motivation or, more generally, the seller's characteristics affect the home selling strategy? The literature suggests this is the case: highly motivated sellers have lower reservation

[28] 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).

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 observed and unobserved housing characteristics to allow for the possibility that some units are more difficult to sell, perhaps due to unusual design features. However, as is the case with home price indices, we do not attempt to control for sellers' motivation and/or for sellers' characteristics. As a result, 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) is to enable comparison of the same good (the same bundle of home characteristics) over time when there is unobserved heterogeneity in home characteristics. Controlling for unobserved seller characteristics, on the other hand, would enable a comparison over time of the same bundle of home characteristics, sold by an individual with the same characteristics. This would fundamentally change the nature of the analysis since the market is defined by the bundle of home characteristics but not by the characteristics 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.

**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 three criterion on which we can base a comparison. First, the assumptions made. Second, the intuitive appeal. Third, the computational appeal.

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 and simultaneously an acceleration of the hazard rate (or deceleration if market conditions are worsening). However, the two are not nested; 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.[29] 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.[30]

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.

# 5 TOM: Some Stylized Facts

In this section we use our proposed TOM indices to establish several stylized facts that have been assessed by previous influential studies. It is important to reassess these relationships when a rigorous estimate, such as the indices developed in our paper, are used to measure housing liquidity. We first look at the *correlation* between home prices and sales volume. Second, we explore patterns

[29] This is conditional on the baseline hazard being an increasing function. More generally, if $\lambda_{it}(y) = \theta_t \lambda_{0i}(\theta_t y)$ then $\partial \lambda_{it}(y)/\partial \theta_t = \lambda_{0i}(\theta_t y) + \theta_t y \lambda'_{0i}(\theta_t y)$.

[30] We note that both indices assume that unobserved heterogeneity is constant over time. If homes experience renovations between listings this assumption no longer holds. Unfortunately, we cannot tell if a home was renovated in our sample.

of seasonality in TOM. Finally, we analyze if TOM is *correlated* with several fundamentals. In each case we find important differences between using our measures of TOM and the conventional measure.

### 5.1 Sales volume

The number of home sales is another important indicator of liquidity in the housing market. Sales volumes exhibit a lot of volatility and are positively correlated with home prices (Ríos-Rull and Sánchez-Marcos, 2007). Typically, in periods of low, or negative, home price appreciation, sales volume falls and time on the market rises (see, e.g., Krainer, 2008). Diaz and Jerez (2013) build a model that explains these correlations through a combination of search frictions and competitive forces. In this section, we show that our TOM indices also correlate with the number of sales.

Because our data is a representative sample of listings, not a complete record of all listings, the number of sales observed in the data is not an appropriate measure to link with our constructed TOM indices. However, Standard and Poor's reports the number of (repeat sales) transactions along with the Case-Shiller home price index for major metropolitan areas. Of the metropolitan areas represented in our Core Logic data, three (Las Vegas, Miami, and San Diego) are included in the S&P data. For these three MSAs we find substantial correlation between sales volume and TOM.

First, look at Figure 8. In each of these metropolitan areas, variations over time in TOM are evidently closely related to fluctuations in the number of transactions. The first row shows how sales volumes vary over time with the conventional measure of TOM. The second row shows how sales volumes vary over time with the RPHI TOM index. In Las Vegas, for example, from the first quarter of 2004 until 2008, sales volumes plummet as time on the market increases dramatically. Interestingly, while both the conventional TOM and the RPHI show a recovery for liquidity over the remaining 6 years, sales volume does not. Similar patterns can be observed in Miami and San Diego.

Table 4 demonstrates regression results that quantify the correlation between sales volumes and TOM. We estimate fixed effects regressions of time on the market measures on the log of the number of transactions. The relationship between sales volume and TOM is statistically significant, even when accounting for serial correlation, and it persists when including quarter dummies to account for seasonality. Interestingly, the coefficients are larger using our proposed TOM indices. A 10% increase in the number of transactions is associated with a $7-8\%$ decrease in time on the market using these indices, but only a $4-5\%$ decrease using the conventional measure.

### 5.2 Seasonality

Ngai and Tenreyro (2014) document seasonality in prices and sales volumes in the housing market, arguing that seasonal variation in housing demand, possibly small, can be amplified by the sellers' response due to search frictions. This leads to more transactions and higher prices in the "hot" season than in the "cold" season. In this section we explore seasonality patterns of out TOM indices.

In Table 5 we estimate regressions of time on the market on quarter dummies. We estimate these regressions for the conventional measure as well as our two proposed TOM indices. Rather than estimate these regressions separately for each MSA, we pool the data and estimate a panel regression with MSA fixed effects because the number of time periods is not large enough to obtain precise MSA-specific estimates. Three different specifications are considered to show that results are robust to different ways of dealing with imbalance in our panel.

Results confirm that there is seasonality in time on the market. Results for the conventional measure show that the median home sold in the second quarter sells significantly more quickly than the median home sold in other quarters. Homes sold in the third quarter sell the second most quickly. This is in line with the results of Ngai and Tenreyro (2014) who find that price appreciation and sales volumes are higher in the summer (2nd and 3rd quarter) than in the winter. Because the RPHI and RMTI measure time on the market as a function of the quarter listed, the timing is slightly different – both indices are significantly lower (indicating an increase in the hazard rate or a reduction in median time on the market) in the first and second quarters than in the third and fourth.

In addition to this difference in timing, the magnitude of the seasonality effects are substantially smaller for the RPHI and RMTI, at least in the third specification that drops data from Fairfax County and the Miami, FL MSA to create a balanced panel.

### 5.3 Determinants of TOM

Here we investigate some determinants of the time on the market using our proposed TOM indices. First, Genesove and Han (2012) provide some evidence that time on the market is negatively influenced by demand factors such as income and population. When demand is high time on the market is low. Second, Eerola and Maattanen (2018) argue that volatility in time on the market can be explained, in part, by variation in borrowing constraints. In this section we show that we are able to replicate these findings in some cases with both the conventional measure of TOM and our new indices. Moreover, in most cases the findings are stronger for our TOM indices.

Figure 9 demonstrates comovement of time on the market and income within MSA for six of our MSAs using the RPHI index. It is evident that as home prices peaked and then crashed in 2007 and continuing through 2008, time on the market and income were actually positively correlated. This appears to be related to dynamics of the housing market crash and the Great Recession. Starting in 2009, as the economy returned to normal times, time on the market and income move in opposite directions.

We pool the MSAs together and estimate fixed effects regressions in order to assess the correlation between TOM and several economic fundamentals: income, population and unemployment. For this analysis we focus on the years after the Great Recession (2009-2013). We include MSA level income, population and unemployment as regressors. The top panel of Table 6 shows the results. While the coefficients are not estimated very precisely, the signs are all in the expected direction (Genesove and Han, 2012) and many are significant at a 10% level. The coefficients on income and population are larger for the regressions using the RPHI and RMTI than when the conventional TOM measure is used. These patterns do not hold in the 2004-2008 period (see bottom panel of Table 6).

Lastly, we can use the real federal funds interest rate as a proxy for borrowing constraints. This does not vary across MSAs so we average the TOM across MSAs in each quarter. We then regress this average on the real federal funds interest rate. Table 7 shows results for each of the TOM measures. We find that when interest rates are high time on the market is too. The effects again are more pronounced for the RPHI and RMTI than for the conventional TOM.

# 6 Conclusions

This paper develops the first housing market TOM indices that are based on repeat listings that also account for censored TOM durations. The two indices we propose control for unobserved heterogeneity exploiting *repeat* listings and 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 in 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 controlling for unobserved heterogeneity has a significant effect on TOM.

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.

## 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. 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.

Diaz, A. and B. Jerez (2013): "House Prices, Sales, and Time on the Market: A Search-Theoretic Framework," *International Economic Review*, 54, 837–872.

Eerola, E. and N. Maattanen (2018): "Borrowing constraints and housing market liquidity," *Review of Economic Dynamics*, 27, 184 – 204.

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.

Genesove, D. and L. Han (2012): "Search and matching in the housing market," *Journal of urban economics*, 72, 31–45.

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 (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.

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.

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.

Krainer, J. (2008): "Falling house prices and rising time on the market," *FRBSF Economic Letter*.

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.

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

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.

Ngai, L. R. and S. Tenreyro (2014): “Hot and cold seasons in the housing market,” *American Economic Review*, 104, 3991–4026.

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.

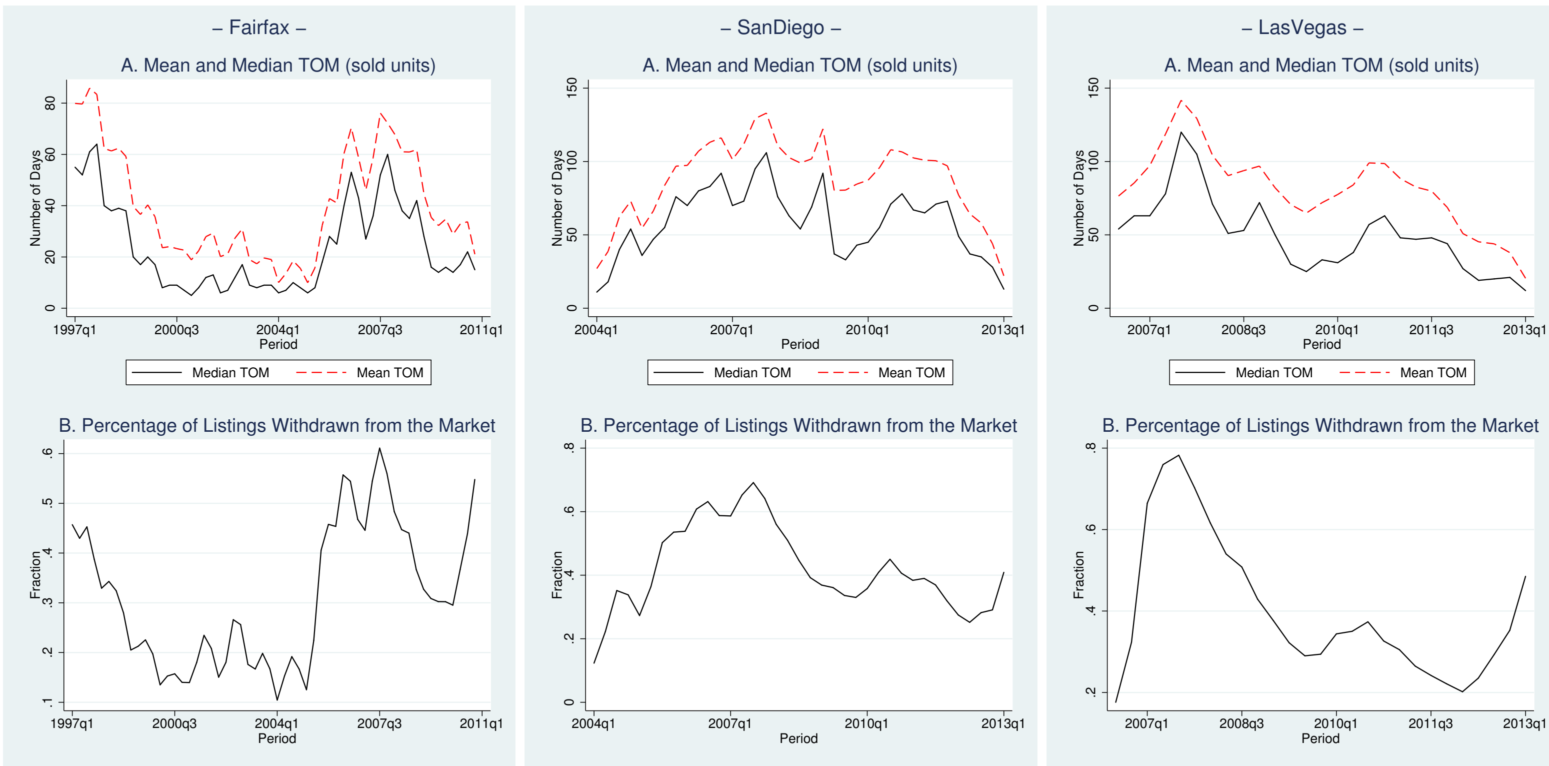
Ríos-Rull, J. and V. Sánchez-Marcos (2007): “House price movements,” Tech. rep., Mimeo.

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 (1985): “Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure,” *Sociological methodology*, 15, 242–276.

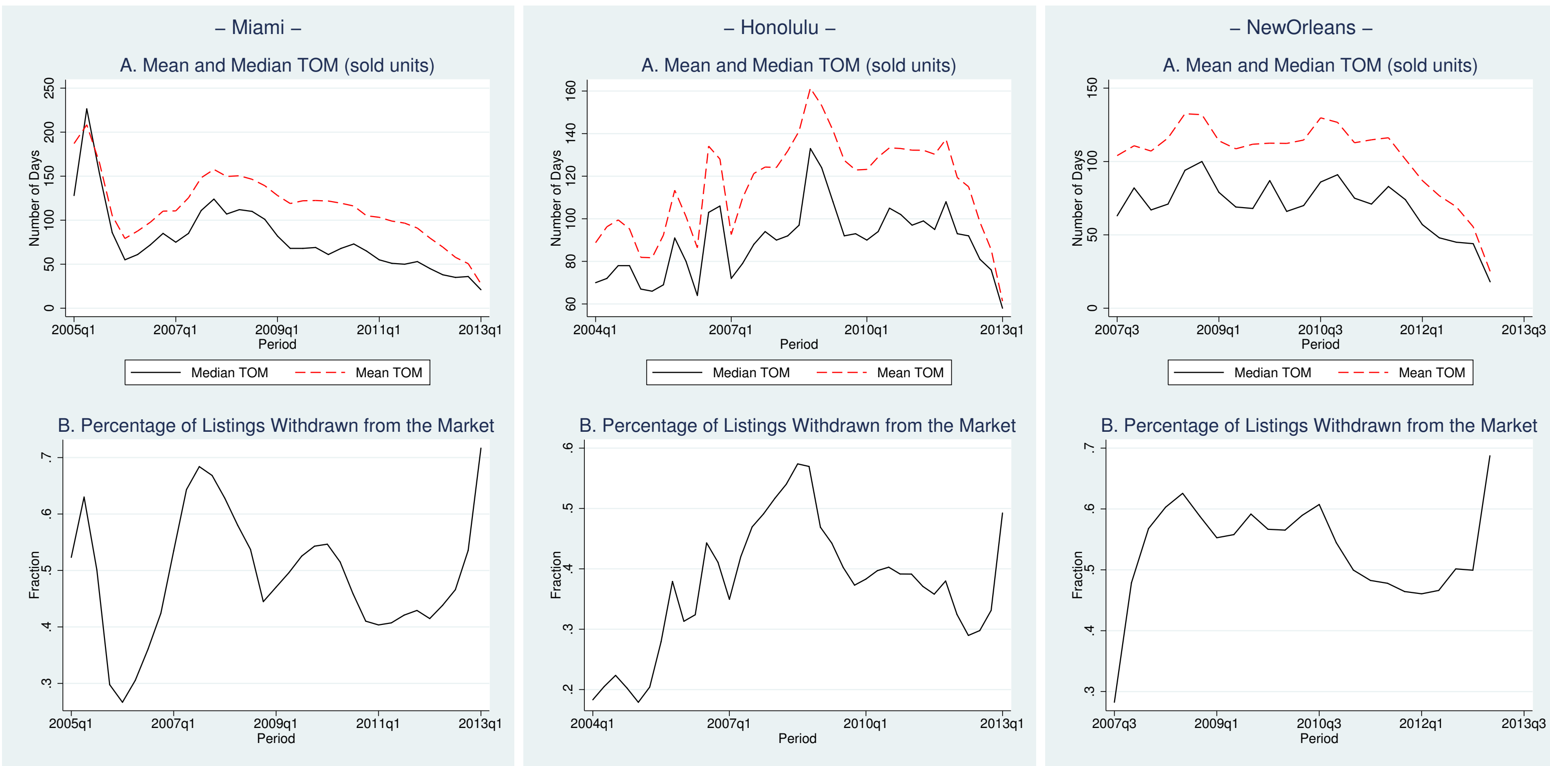
Wallace, N. and R. Meese (1997): “The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches,” *The Journal of Real Estate Finance and Economics*, 14, 51–73.

Figure 1:
Descriptive "Conventional" Statistics (part 1)



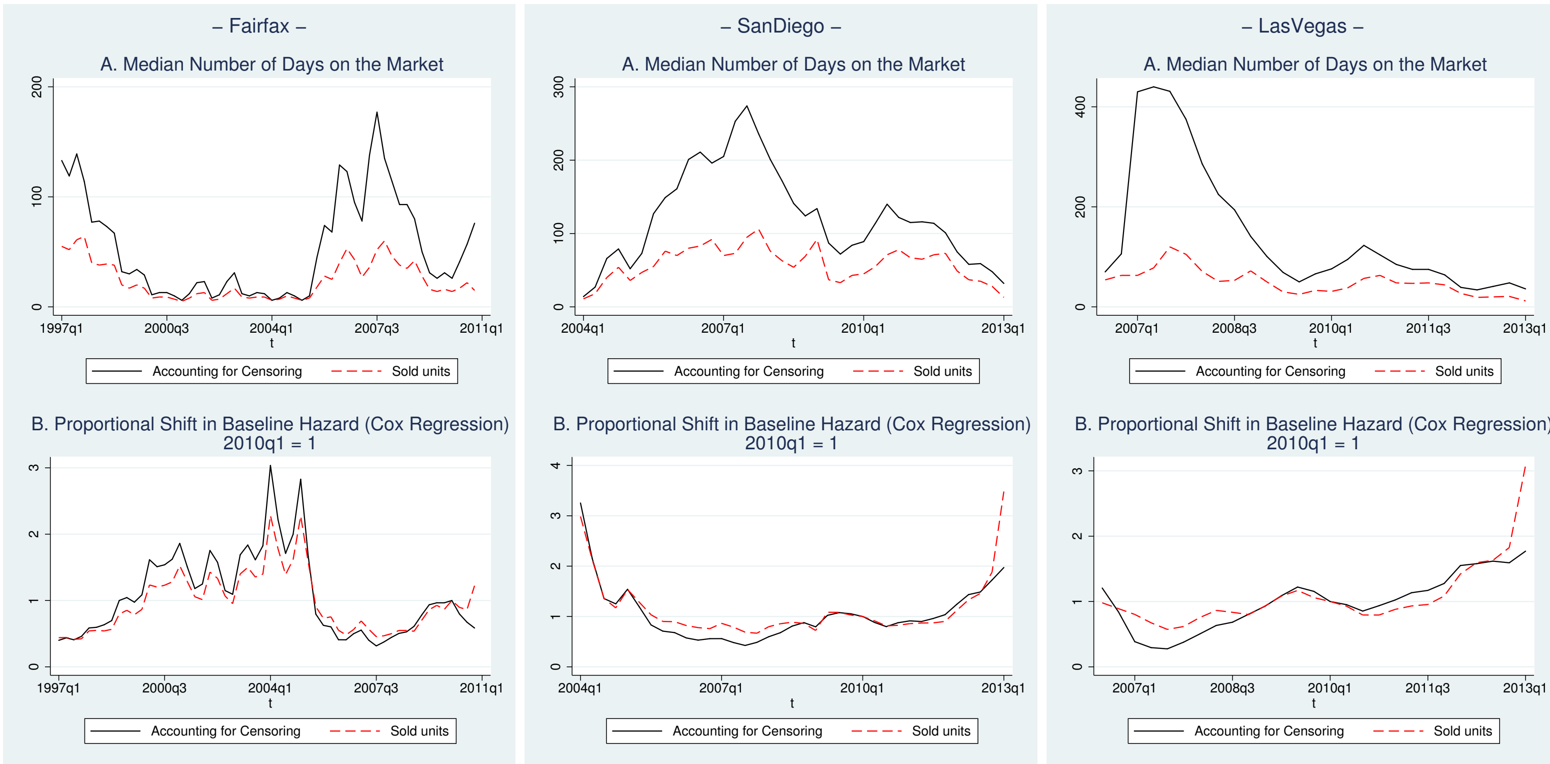

*Notes:* This figure presents descriptive statistics of the sample. Panel A computes the mean and median number of days that a home stays on the market (TOM). This is a "conventional" estimate that simply computes the mean and median TOM of finished durations (sold units). The second panel shows the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

Figure 2:
Descriptive "Conventional" Statistics (part 2)



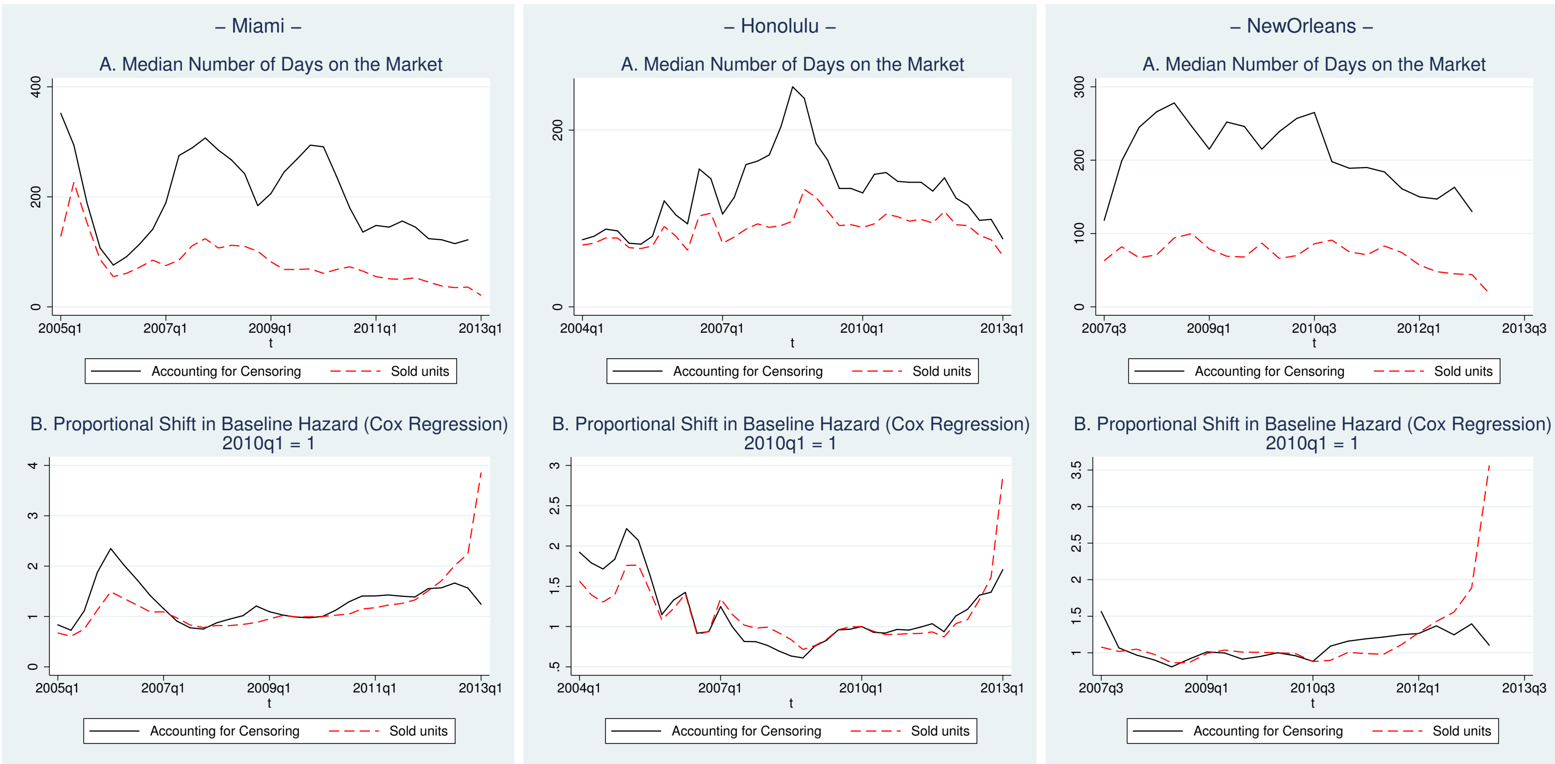

*Notes:* This figure presents descriptive statistics of the sample. Panel A computes the mean and median number of days that a home stays on the market (TOM). This is a "conventional" estimate that simply computes the mean and median TOM of finished durations (sold units). The second panel shows the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

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



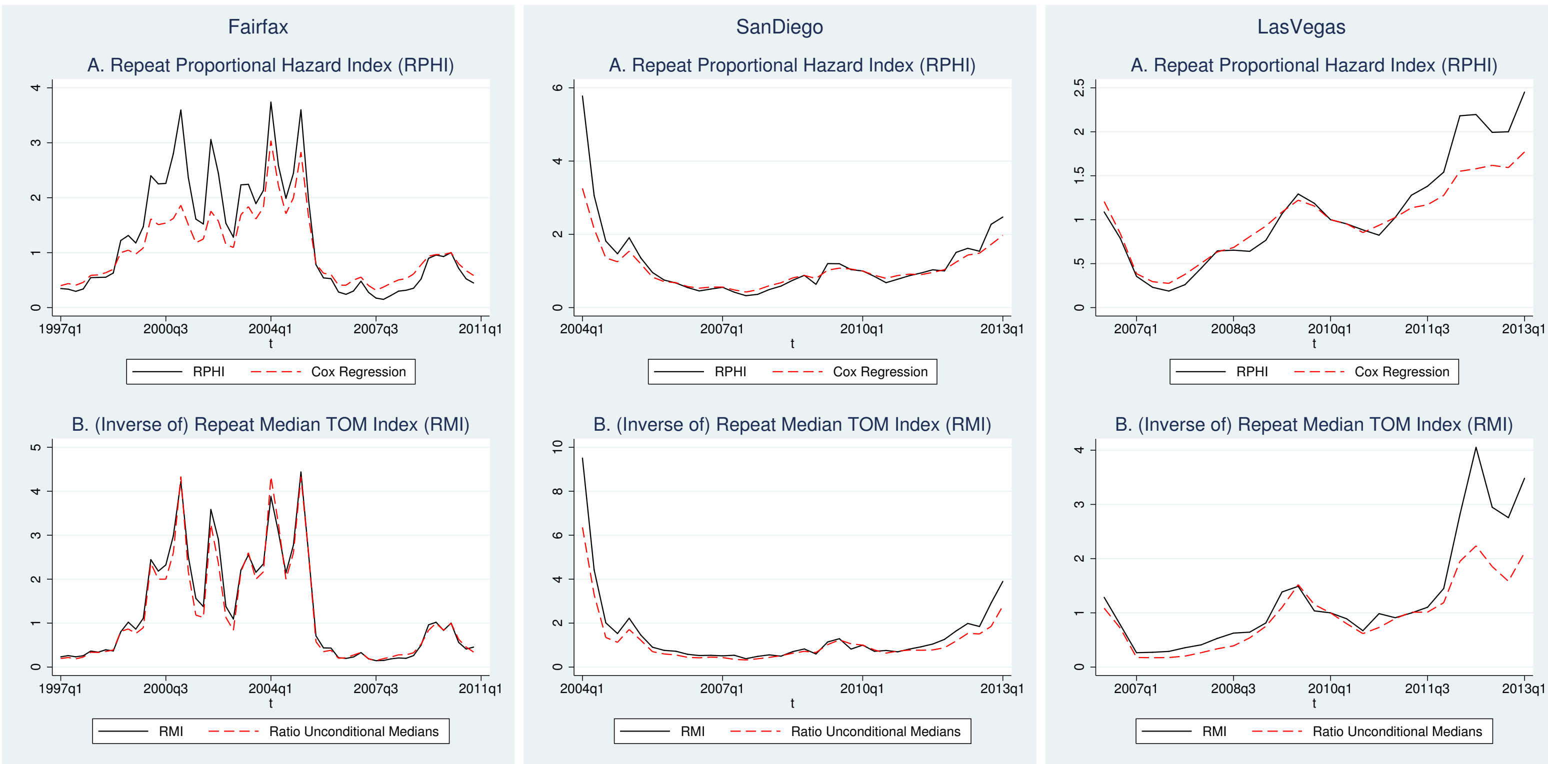

*Notes:* Panel A computes the median number of days that a home stays on the market (TOM). The "conventional" estimate simply computes the median TOM of finished durations (sold units). To account for censoring, a Kaplan-Meier estimator is used. For units that are sold, TOM is defined as the difference 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)



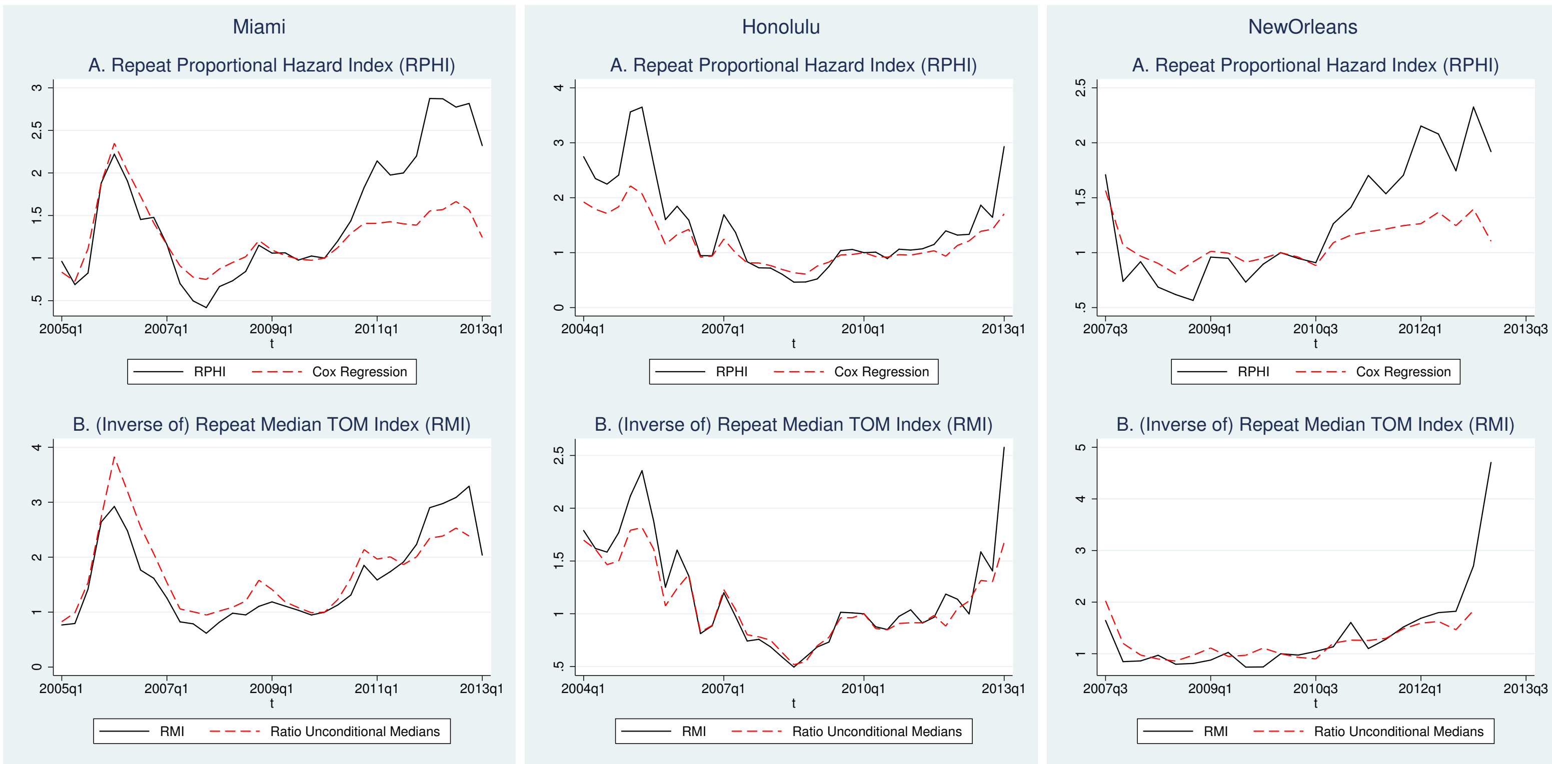

*Notes:* Panel A computes the median number of days that a home stays on the market (TOM). The "conventional" estimate simply computes the median TOM of finished durations (sold units). To account for censoring, a Kaplan-Meier estimator is used. For units that are sold, TOM is defined as the difference 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:
Repeat Time-On-The-Market Indices (part 1)



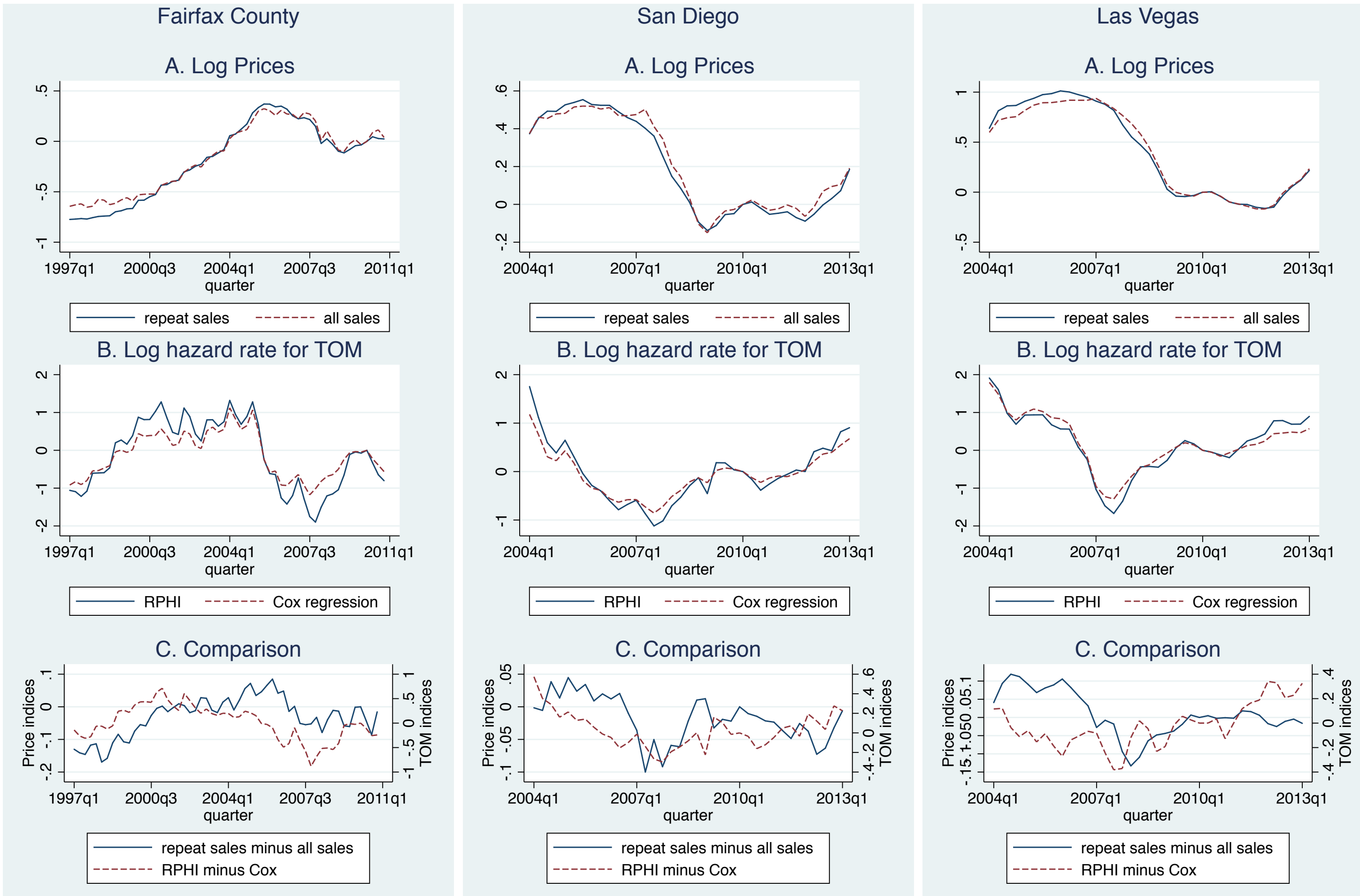

*Notes:* Panel A computes the Repeat Proportional Hazard Index (RPHI). The index measures relative shifts of the baseline hazard after controlling for unobserved home's heterogeneity. For example, and index value of 1.5 in period $t$ reflects a 50 percent increase in the home sale baseline hazard in 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 (RMTI). We report the inverse of the RMTI 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 6:
Repeat Time-On-The-Market Indices (part 2)



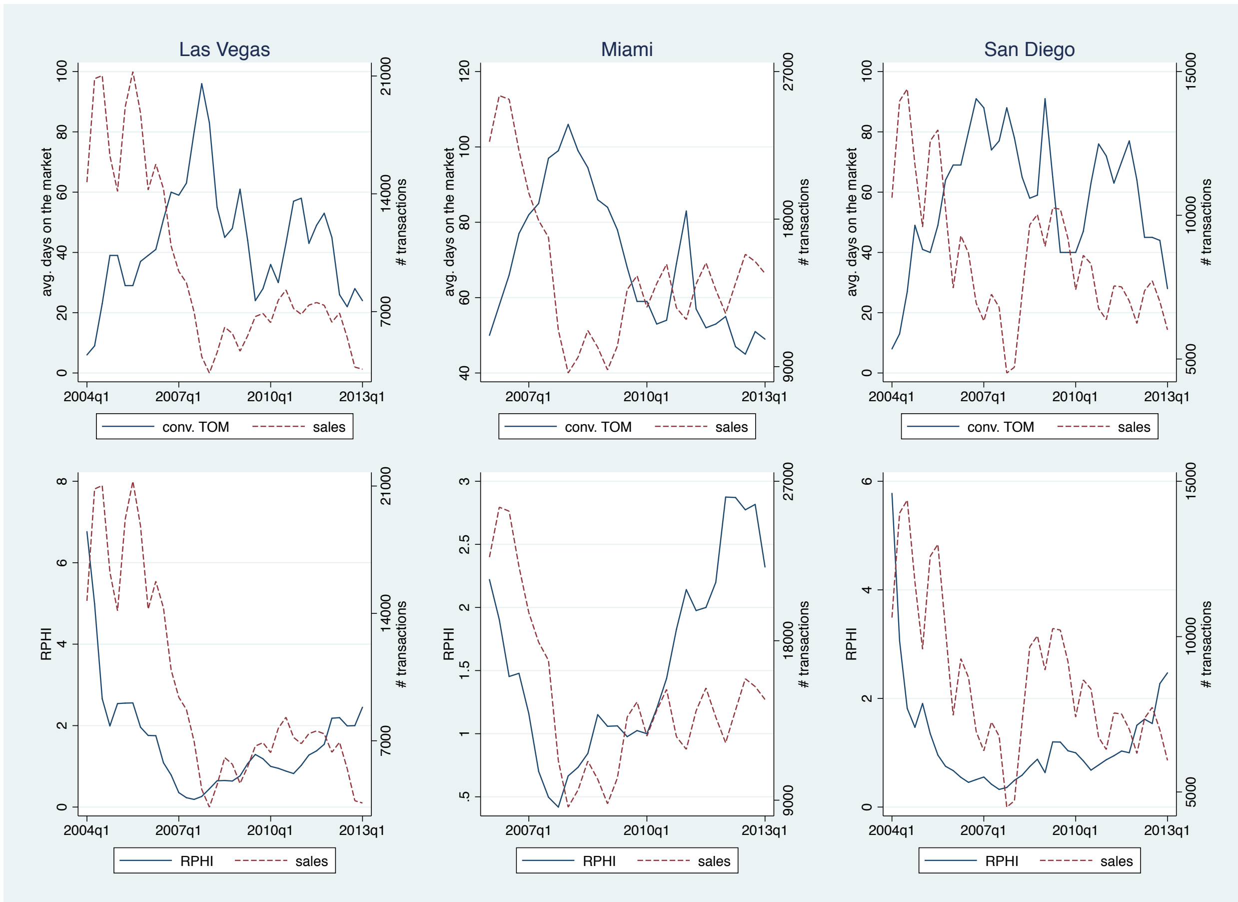

*Notes:* Panel A computes the Repeat Proportional Hazard Index (RPHI). The index measures relative shifts of the baseline hazard after controlling for unobserved home's heterogeneity. For example, and index value of 1.5 in period $t$ reflects a 50 percent increase in the home sale baseline hazard in 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 (RMTI). We report the inverse of the RMTI 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:
Unobserved Heterogeneity in Prices and TOM



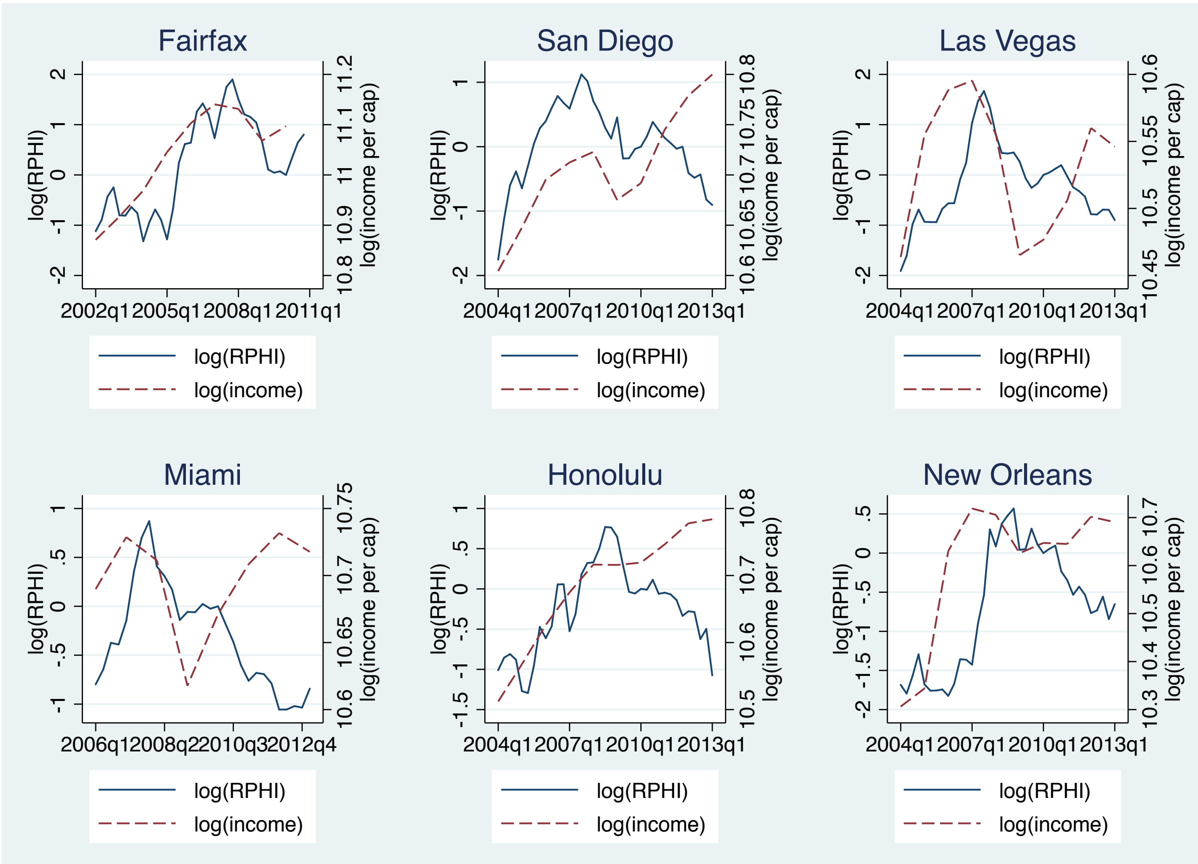

*Notes:* Panel A computes an unconditional mean price index, using all sales, and a repeat sales price index. Panel B compares the RPHI index with a similar index based on a Cox regression that does not correct for unobserved heterogeneity. Unlike in figures 5 and 6, both of these indices are log-transformed in order to highlight their differences. Panel C compares the difference between the two prices indices shown in panel A (solid line) and the difference between the two TOM indices shown in panel B (dashed line).

Figure 8:
TOM measures and sales volumes by urban area




*Notes:* The top row of figures plots sales volumes over time along with the conventional time on the market measure. The figures in the bottom row show the same sales volume numbers along with the RPHI over time. Sales volumes were obtained from Standard and Poor's.

Figure 9:
TOM measures and incomes by urban area




*Notes:* Each figure plots the log of the RPHI TOM measure along with log of per capita personal income for different metropolitan areas over time. Per capita personal income was obtained from the U.S. Bureau of Economic Analysis.

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 | 16,166 | 2004q1 | 2013q1 |
| 02 | Boulder, CO | 47,177 | 26,923 | 19,485 | 2004q1 | 2013q1 |
| 03 | Durham, NC | 59,234 | 34,125 | 22,285 | 2004q1 | 2013q1 |
| 04 | Fairfax County, VA | 357,515 | 244,961 | 232,382 | 2007q2 | 2010q4 |
| 05 | Honolulu, HI | 85,511 | 54,350 | 36,567 | 2004q1 | 2013q1 |
| 06 | Las Vegas-Paradise, NV | 262,267 | 153,577 | 112,111 | 2006q3 | 2013q1 |
| 07 | Medford, OR | 26,138 | 16,315 | 11,050 | 2004q1 | 2013q1 |
| 08 | Miami-Miami Beach-Kendall, FL | 219,210 | 112,069 | 73,471 | 2005q1 | 2013q1 |
| 09 | New Orleans-Metairie-Kenner, LA | 79,845 | 36,869 | 30,724 | 2007q3 | 2013q1 |
| 10 | Olympia, WA | 35,416 | 21,733 | 13,416 | 2004q1 | 2013q1 |
| 11 | San Diego-Carlsbad-San Marcos, CA | 367,122 | 207,162 | 166,427 | 2004q1 | 2013q1 |
| 12 | San Luis Obispo-Paso Robles, CA | 27,506 | 19,044 | 9,664 | 2004q1 | 2013q1 |
| 13 | Santa Barbara-Santa Maria, CA | 29,054 | 20,085 | 11,304 | 2004q1 | 2013q1 |
| 14 | Toledo, OH | 65,873 | 35,673 | 22,778 | 2004q1 | 2013q1 |
| 15 | Youngstown-Warren-Boardman, OH-PA | 52,825 | 27,098 | 18,227 | 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 CBSAs.

| | Geographic Area | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Period | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 |
| 2002q1 | | | | 3.060 | | | | | | | | | | | |
| 2002q2 | | | | 2.446 | | | | | | | | | | | |
| 2002q3 | | | | 1.535 | | | | | | | | | | | |
| 2002q4 | | | | 1.281 | | | | | | | | | | | |
| 2003q1 | | | | 2.234 | | | | | | | | | | | |
| 2003q2 | | | | 2.244 | | | | | | | | | | | |
| 2003q3 | | | | 1.892 | | | | | | | | | | | |
| 2003q4 | | | | 2.133 | | | | | | | | | | | |
| 2004q1 | 2.191 | 1.111 | 1.537 | 3.740 | 2.746 | | 2.940 | | | 2.204 | 5.778 | 3.333 | 4.152 | 3.256 | 1.396 |
| 2004q2 | 2.082 | 1.242 | 1.690 | 2.583 | 2.349 | | 2.536 | | | 3.285 | 3.057 | 3.302 | 3.043 | 3.056 | 1.089 |
| 2004q3 | 1.219 | 0.825 | 1.223 | 1.989 | 2.248 | | 3.088 | | | 3.007 | 1.816 | 1.868 | 2.457 | 2.102 | 1.340 |
| 2004q4 | 1.001 | 0.845 | 0.949 | 2.445 | 2.412 | | 3.249 | | | 3.435 | 1.469 | 1.904 | 3.129 | 2.024 | 1.364 |
| 2005q1 | 1.471 | 1.308 | 2.028 | 3.600 | 3.562 | | 3.106 | 0.962 | | 3.700 | 1.908 | 2.028 | 1.918 | 2.699 | 1.490 |
| 2005q2 | 0.982 | 1.350 | 2.034 | 1.975 | 3.649 | | 3.115 | 0.689 | | 3.465 | 1.354 | 2.213 | 1.515 | 2.380 | 1.260 |
| 2005q3 | 0.591 | 1.120 | 1.329 | 0.782 | 2.608 | | 1.887 | 0.825 | | 4.360 | 0.958 | 2.284 | 1.375 | 2.072 | 1.010 |
| 2005q4 | 0.635 | 0.974 | 1.416 | 0.541 | 1.603 | | 1.432 | 1.879 | | 3.079 | 0.752 | 0.904 | 0.908 | 1.939 | 1.129 |
| 2006q1 | 0.497 | 1.182 | 1.920 | 0.527 | 1.845 | | 1.392 | 2.221 | | 3.774 | 0.672 | 1.115 | 0.903 | 2.407 | 1.228 |
| 2006q2 | 0.680 | 1.062 | 2.031 | 0.284 | 1.592 | | 0.945 | 1.905 | | 2.252 | 0.548 | 0.884 | 0.857 | 1.777 | 1.129 |
| 2006q3 | 0.346 | 0.905 | 1.535 | 0.241 | 0.946 | 1.088 | 0.852 | 1.453 | | 1.771 | 0.455 | 0.866 | 0.593 | 1.393 | 0.876 |
| 2006q4 | 0.359 | 0.981 | 1.698 | 0.302 | 0.944 | 0.787 | 0.650 | 1.478 | | 1.446 | 0.507 | 0.844 | 0.622 | 1.131 | 1.004 |
| 2007q1 | 0.381 | 1.346 | 2.092 | 0.481 | 1.690 | 0.355 | 0.882 | 1.160 | | 1.601 | 0.554 | 1.137 | 0.703 | 0.990 | 0.896 |
| 2007q2 | 0.521 | 1.137 | 1.830 | 0.277 | 1.369 | 0.231 | 1.052 | 0.702 | | 1.115 | 0.421 | 0.834 | 0.600 | 0.726 | 1.038 |
| 2007q3 | 0.529 | 1.031 | 1.164 | 0.173 | 0.835 | 0.188 | 0.577 | 0.497 | 1.709 | 0.999 | 0.326 | 0.566 | 0.453 | 0.624 | 0.690 |
| 2007q4 | 0.453 | 1.235 | 1.400 | 0.150 | 0.724 | 0.261 | 0.914 | 0.419 | 0.738 | 1.135 | 0.361 | 0.676 | 0.401 | 0.658 | 0.668 |
| 2008q1 | 0.665 | 1.291 | 1.277 | 0.222 | 0.719 | 0.448 | 0.579 | 0.665 | 0.918 | 1.119 | 0.493 | 0.595 | 0.619 | 1.022 | 0.828 |
| 2008q2 | 0.603 | 1.233 | 1.313 | 0.300 | 0.606 | 0.646 | 0.797 | 0.734 | 0.687 | 0.798 | 0.587 | 0.619 | 0.770 | 0.741 | 0.825 |
| 2008q3 | 0.736 | 1.066 | 0.960 | 0.315 | 0.463 | 0.653 | 0.711 | 0.844 | 0.619 | 1.146 | 0.746 | 0.887 | 0.698 | 0.865 | 0.883 |

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

| | Geographic Area | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Period | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 |
| 2008q4 | 0.933 | 0.917 | 0.665 | 0.352 | 0.466 | 0.640 | 0.593 | 1.151 | 0.565 | 1.043 | 0.881 | 1.049 | 1.084 | 0.904 | 1.092 |
| 2009q1 | 0.957 | 0.910 | 1.047 | 0.520 | 0.522 | 0.766 | 0.645 | 1.059 | 0.960 | 1.045 | 0.635 | 0.863 | 0.773 | 1.085 | 0.949 |
| 2009q2 | 1.164 | 1.191 | 1.295 | 0.897 | 0.748 | 1.067 | 0.886 | 1.062 | 0.949 | 0.869 | 1.201 | 0.991 | 0.850 | 1.021 | 0.923 |
| 2009q3 | 1.146 | 1.398 | 1.197 | 0.956 | 1.037 | 1.292 | 1.131 | 0.976 | 0.732 | 1.255 | 1.197 | 1.282 | 1.169 | 1.104 | 0.897 |
| 2009q4 | 1.134 | 1.152 | 1.178 | 0.930 | 1.059 | 1.186 | 1.319 | 1.025 | 0.896 | 1.337 | 1.037 | 1.102 | 0.625 | 0.876 | 1.061 |
| 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 | 0.770 | 1.157 | 0.806 | 0.717 | 1.009 | 0.952 | 0.727 | 1.200 | 0.947 | 0.980 | 0.852 | 0.986 | 0.771 | 0.908 | 0.874 |
| 2010q3 | 0.761 | 0.736 | 0.709 | 0.524 | 0.892 | 0.884 | 0.845 | 1.438 | 0.908 | 1.329 | 0.681 | 1.125 | 0.738 | 0.855 | 0.809 |
| 2010q4 | 1.424 | 0.949 | 1.006 | 0.448 | 1.063 | 0.824 | 1.471 | 1.826 | 1.263 | 0.852 | 0.775 | 1.017 | 0.737 | 1.604 | 1.227 |
| 2011q1 | 1.108 | 1.166 | 0.808 | | 1.047 | 1.025 | 1.077 | 2.142 | 1.411 | 1.209 | 0.872 | 1.006 | 0.737 | 1.448 | 1.035 |
| 2011q2 | 1.235 | 1.186 | 1.027 | | 1.070 | 1.278 | 1.363 | 1.976 | 1.703 | 0.980 | 0.946 | 1.358 | 0.987 | 1.362 | 1.114 |
| 2011q3 | 1.196 | 0.987 | 1.240 | | 1.149 | 1.381 | 1.109 | 2.000 | 1.536 | 1.052 | 1.032 | 1.264 | 0.978 | 1.209 | 0.920 |
| 2011q4 | 1.371 | 1.015 | 1.692 | | 1.397 | 1.541 | 2.178 | 2.199 | 1.705 | 1.533 | 1.001 | 1.207 | 0.947 | 1.210 | 1.479 |
| 2012q1 | 2.279 | 2.452 | 1.277 | | 1.320 | 2.182 | 2.287 | 2.875 | 2.152 | 1.282 | 1.508 | 2.050 | 1.087 | 1.988 | 2.004 |
| 2012q2 | 1.863 | 3.202 | 1.421 | | 1.333 | 2.196 | 2.285 | 2.871 | 2.080 | 1.873 | 1.619 | 1.962 | 1.628 | 1.713 | 1.708 |
| 2012q3 | 1.739 | 2.244 | 1.399 | | 1.864 | 1.993 | 1.730 | 2.774 | 1.744 | 1.083 | 1.539 | 1.930 | 1.863 | 1.549 | 1.317 |
| 2012q4 | 1.734 | 3.762 | 1.556 | | 1.644 | 2.000 | 2.056 | 2.817 | 2.325 | 1.260 | 2.274 | 1.800 | 2.205 | 1.893 | 1.688 |
| 2013q1 | 1.809 | 3.725 | 1.741 | | 2.930 | 2.452 | 2.644 | 2.321 | 1.919 | 2.412 | 2.472 | 2.585 | 2.553 | 2.360 | 1.298 |

*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 CBSAs.

| | Geographic Area | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Period | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 |
| 2002q1 | | | | 0.279 | | | | | | | | | | | |
| 2002q2 | | | | 0.344 | | | | | | | | | | | |
| 2002q3 | | | | 0.725 | | | | | | | | | | | |
| 2002q4 | | | | 0.916 | | | | | | | | | | | |
| 2003q1 | | | | 0.453 | | | | | | | | | | | |
| 2003q2 | | | | 0.392 | | | | | | | | | | | |
| 2003q3 | | | | 0.464 | | | | | | | | | | | |
| 2003q4 | | | | 0.426 | | | | | | | | | | | |
| 2004q1 | 0.681 | 0.882 | 0.749 | 0.258 | 0.559 | | 0.506 | | | 0.595 | 0.105 | 0.264 | 0.266 | 0.551 | 0.819 |
| 2004q2 | 0.700 | 0.805 | 0.664 | 0.326 | 0.618 | | 0.511 | | | 0.459 | 0.226 | 0.230 | 0.247 | 0.449 | 0.918 |
| 2004q3 | 0.961 | 1.076 | 0.837 | 0.466 | 0.631 | | 0.461 | | | 0.514 | 0.497 | 0.350 | 0.373 | 0.755 | 1.042 |
| 2004q4 | 1.145 | 1.108 | 0.829 | 0.360 | 0.565 | | 0.551 | | | 0.471 | 0.655 | 0.413 | 0.319 | 0.755 | 0.867 |
| 2005q1 | 0.980 | 0.854 | 0.597 | 0.225 | 0.473 | | 0.416 | 1.304 | | 0.395 | 0.450 | 0.344 | 0.452 | 0.620 | 0.804 |
| 2005q2 | 1.292 | 0.842 | 0.637 | 0.384 | 0.424 | | 0.464 | 1.261 | | 0.377 | 0.687 | 0.360 | 0.505 | 0.564 | 0.894 |
| 2005q3 | 2.017 | 1.102 | 0.899 | 1.408 | 0.533 | | 0.623 | 0.704 | | 0.404 | 1.105 | 0.479 | 0.665 | 0.673 | 0.976 |
| 2005q4 | 1.821 | 0.976 | 0.745 | 2.290 | 0.799 | | 0.882 | 0.379 | | 0.461 | 1.315 | 0.778 | 0.860 | 0.608 | 1.023 |
| 2006q1 | 1.176 | 0.819 | 0.605 | 2.305 | 0.624 | | 0.916 | 0.342 | | 0.439 | 1.384 | 0.800 | 1.070 | 0.665 | 0.785 |
| 2006q2 | 1.876 | 0.874 | 0.604 | 4.694 | 0.736 | | 1.001 | 0.403 | | 0.539 | 1.718 | 1.060 | 1.240 | 0.739 | 0.879 |
| 2006q3 | 2.175 | 0.896 | 0.760 | 5.066 | 1.231 | 0.777 | 1.174 | 0.566 | | 0.638 | 1.903 | 1.144 | 1.359 | 0.804 | 1.051 |
| 2006q4 | 2.015 | 0.930 | 0.717 | 4.361 | 1.129 | 1.285 | 1.157 | 0.619 | | 0.727 | 1.870 | 0.889 | 1.483 | 1.285 | 0.993 |
| 2007q1 | 2.075 | 0.727 | 0.549 | 3.025 | 0.834 | 3.770 | 1.072 | 0.795 | | 0.716 | 1.951 | 0.601 | 1.444 | 1.395 | 0.862 |
| 2007q2 | 2.157 | 0.902 | 0.613 | 5.298 | 1.025 | 3.658 | 0.873 | 1.214 | | 0.949 | 1.856 | 0.955 | 1.762 | 1.485 | 1.046 |
| 2007q3 | 1.729 | 1.078 | 1.016 | 6.837 | 1.348 | 3.459 | 1.319 | 1.270 | 0.608 | 1.049 | 2.624 | 1.651 | 2.288 | 1.886 | 1.299 |
| 2007q4 | 2.133 | 0.813 | 0.712 | 6.616 | 1.320 | 2.798 | 1.199 | 1.622 | 1.181 | 0.955 | 2.054 | 1.001 | 2.354 | 1.699 | 1.127 |
| 2008q1 | 1.324 | 0.893 | 0.848 | 5.355 | 1.456 | 2.436 | 1.226 | 1.222 | 1.161 | 0.932 | 1.797 | 1.017 | 1.606 | 1.221 | 1.137 |
| 2008q2 | 1.785 | 0.633 | 0.780 | 4.809 | 1.698 | 1.890 | 1.692 | 1.020 | 1.030 | 1.084 | 2.013 | 1.593 | 1.182 | 1.797 | 1.334 |
| 2008q3 | 1.282 | 1.198 | 0.996 | 5.018 | 2.021 | 1.594 | 1.198 | 1.053 | 1.254 | 0.992 | 1.424 | 0.991 | 1.214 | 1.384 | 1.381 |

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

| | Geographic Area | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Period | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 |
| 2008q4 | 1.173 | 1.004 | 0.961 | 3.821 | 1.696 | 1.552 | 0.970 | 0.904 | 1.231 | 0.736 | 1.213 | 0.776 | 0.678 | 1.362 | 0.963 |
| 2009q1 | 1.251 | 0.858 | 0.874 | 2.008 | 1.459 | 1.228 | 1.490 | 0.843 | 1.140 | 0.905 | 1.675 | 1.138 | 1.016 | 1.041 | 1.158 |
| 2009q2 | 1.412 | 0.837 | 0.889 | 1.040 | 1.366 | 0.722 | 1.135 | 0.901 | 0.975 | 1.099 | 0.875 | 0.858 | 1.085 | 1.495 | 1.084 |
| 2009q3 | 0.898 | 0.773 | 0.800 | 0.980 | 0.986 | 0.671 | 0.963 | 0.967 | 1.348 | 0.999 | 0.774 | 0.947 | 0.671 | 1.354 | 1.255 |
| 2009q4 | 0.924 | 0.824 | 0.935 | 1.196 | 0.991 | 0.965 | 1.459 | 1.055 | 1.344 | 0.714 | 1.221 | 0.821 | 1.137 | 1.122 | 1.106 |
| 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 | 1.200 | 0.823 | 1.540 | 1.773 | 1.140 | 1.123 | 1.495 | 0.885 | 1.026 | 1.241 | 1.394 | 1.058 | 1.123 | 1.193 | 1.291 |
| 2010q3 | 1.212 | 1.012 | 1.160 | 2.436 | 1.176 | 1.491 | 1.320 | 0.762 | 0.957 | 0.793 | 1.324 | 0.933 | 1.545 | 1.304 | 0.933 |
| 2010q4 | 0.742 | 1.168 | 0.942 | 2.202 | 1.025 | 1.014 | 0.773 | 0.540 | 0.881 | 1.391 | 1.440 | 0.826 | 1.127 | 1.001 | 0.847 |
| 2011q1 | 1.205 | 0.972 | 1.279 | | 0.964 | 1.098 | 1.187 | 0.630 | 0.623 | 0.815 | 1.219 | 0.984 | 1.300 | 1.083 | 1.115 |
| 2011q2 | 0.951 | 0.800 | 1.197 | | 1.096 | 1.002 | 0.999 | 0.576 | 0.909 | 0.947 | 1.082 | 0.568 | 1.150 | 1.343 | 1.164 |
| 2011q3 | 1.242 | 0.887 | 0.864 | | 1.034 | 0.906 | 0.972 | 0.523 | 0.781 | 1.175 | 0.953 | 0.754 | 1.013 | 1.276 | 1.017 |
| 2011q4 | 0.938 | 0.973 | 0.973 | | 0.843 | 0.690 | 0.663 | 0.448 | 0.659 | 0.754 | 0.794 | 0.966 | 1.325 | 1.153 | 0.807 |
| 2012q1 | 0.976 | 0.544 | 0.694 | | 0.879 | 0.356 | 0.668 | 0.345 | 0.593 | 0.866 | 0.608 | 0.556 | 0.902 | 0.898 | 0.910 |
| 2012q2 | 0.736 | 0.472 | 0.785 | | 1.001 | 0.247 | 0.571 | 0.336 | 0.556 | 0.780 | 0.504 | 0.500 | 0.487 | 0.853 | 0.699 |
| 2012q3 | 0.792 | 0.502 | 0.825 | | 0.630 | 0.339 | 0.826 | 0.324 | 0.549 | 0.810 | 0.541 | 0.466 | 0.579 | 0.909 | 0.726 |
| 2012q4 | 0.503 | 0.474 | 0.890 | | 0.711 | 0.363 | 0.557 | 0.304 | 0.370 | 0.653 | 0.342 | 0.583 | 0.469 | 0.754 | 0.747 |
| 2013q1 | 0.439 | 0.363 | 0.581 | | 0.388 | 0.287 | 0.389 | 0.491 | 0.212 | 0.433 | 0.256 | 0.238 | 0.183 | 0.700 | 0.491 |

*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:
Time-on-the-market measures and sales volumes

| | conventional TOM | | | RPHI | | | RMTI | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) |
| log of transactions | -0.499 | -0.514 | -0.514 | -0.749 | -0.838 | -0.838 | -0.768 | -0.854 | -0.854 |
| | (0.136) | (0.146) | (0.207) | (0.153) | (0.159) | (0.255) | (0.180) | (0.188) | (0.303) |
| MSA dummies? | yes | yes | yes | yes | yes | yes | yes | yes | yes |
| quarter dummies? | no | yes | yes | no | yes | yes | no | yes | yes |
| newey west std. errors? | no | no | yes | no | no | yes | no | no | yes |

*Notes:* Each column represents a different regression of a log TOM measure on the log of the number of transactions. Data is pooled across the three MSAs: Las Vegas, Miami, and San Diego. In columns (1) and (2), White heteroskedasticity-robust standard errors are reported. In column (3), we report Newey West standard errors using a maximum lag length of 8 quarters. Transaction numbers were obtained from Standard and Poor's.

Table 5:
Seasonality in time on the market

| | conventional TOM | | | RPHI | | | RMTI | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) |
| 2nd quarter | -0.120 | -0.142 | -0.117 | 0.094 | 0.045 | 0.030 | 0.116 | 0.047 | 0.034 |
| | (0.034) | (0.033) | (0.022) | (0.015) | (0.020) | (0.018) | (0.016) | (0.021) | (0.020) |
| 3rd quarter | -0.027 | -0.049 | -0.043 | 0.224 | 0.176 | 0.152 | 0.226 | 0.157 | 0.130 |
| | (0.027) | (0.024) | (0.028) | (0.024) | (0.030) | (0.020) | (0.026) | (0.033) | (0.016) |
| 4th quarter | 0.103 | 0.081 | 0.083 | 0.196 | 0.148 | 0.128 | 0.191 | 0.122 | 0.095 |
| | (0.028) | (0.024) | (0.026) | (0.027) | (0.033) | (0.025) | (0.032) | (0.038) | (0.024) |
| MSA dummies? | yes | yes | yes | yes | yes | yes | yes | yes | yes |
| year dummies? | no | yes | yes | no | no | yes | no | no | yes |
| balanced sample? | no | no | yes | no | no | yes | no | no | yes |

*Notes:* Each column represents a different regression of a log TOM measure on quarter of year dummies. Data is pooled across all 15 MSAs, except in column (3) where Fairfax County, VA and Miami, FL are excluded.

Table 6:
Time on the market and determinants of demand

| | Conventional TOM | RPHI | RMTI |
|---|---|---|---|
| **a. 2009-2013** | | | |
| log income | -1.673 | -2.386 | -3.181 |
| | (1.087) | (1.599) | (1.597) |
| log population | -3.873 | -5.639 | -6.154 |
| | (1.702) | (3.115) | (3.459) |
| log unemployment | 0.422 | 0.419 | 0.305 |
| | (0.068) | (0.200) | (0.173) |
| **b. 2004-2008** | | | |
| log income | 0.243 | 1.042 | -0.171 |
| | (1.875) | (3.330) | (3.146) |
| log population | 1.206 | 4.404 | 5.025 |
| | (2.063) | (3.482) | (4.335) |
| log unemployment | 0.820 | 1.308 | 1.326 |
| | (0.193) | (0.510) | (0.444) |

*Notes:* Each column represents a different regression of a log TOM measure on log income, log population, and log unemployment rate. Data is pooled across 13 of the 15 MSAs, excluding only Fairfax County, VA and New Orleans, LA. The regressions in panel (a) used the 17 quarters of data from 2009q1-2013q1 and the regressions in panel (b) used the 20 quarters of data from 2004q1-2008q4. Per capita personal income, population, and unemployment rates for each metropolitan area were obtained from the U.S. Bureau of Economic Analysis.

Table 7:
Time on the market and interest rates

| | Conventional TOM | RPHI | RMTI |
|---|---|---|---|
| real federal funds rate | 0.029 | 0.074 | 0.057 |
| | (0.023) | (0.035) | (0.032) |

*Notes:* Each column represents a different regression of a log TOM measure on the real federal funds rate. The TOM measure is average across 13 of the 15 MSAs, excluding only Fairfax County, VA and New Orleans, LA. We report Newey West standard errors using a maximum lag length of 8 quarters. The regressions are computed from 37 quarterly observations from 2004$q$1 to 2013$q$1.