The Repeat Time-On-The-Market Index

Paul E. Carrillo
George Washington University

Benjamin Williams
George Washington University

January 2015
Introduction

• Housing market is a very important one!
• Measuring housing market conditions:
  — Housing starts / Vacancy rates / Mortgage originations, etc.
  — Transaction prices
    • Median sale prices (NAR)
    • Repeat sales (Case-Shiller, OFHEO)
  — Housing distress index
    • Chauvet, Gabriel and Lutz (2013)
  — Liquidity (TOM)?
Introduction

• TOM is a key measure of housing market conditions

<table>
<thead>
<tr>
<th>Market area</th>
<th>Median Price / change</th>
<th>Median TOM</th>
<th>$\Delta$ Median TOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$200K / 5%</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>$200K / 5%</td>
<td>60</td>
<td>+40</td>
</tr>
</tbody>
</table>
Introduction

• **MLS Data:**
  – List price (list price changes)
  – Sale price
  – Number of days a home stays on the market (TOM)
    • Contract date (“under contract” / “pending”)
    • Closing date

  – Official statistics that exploit MLS Data?
This paper

• Uses MLS data to construct quality adjusted TOM indices

• Main contributions / conclusions so far:
  – New methods to measure changes in housing liquidity
    • Exploits repeat-sales ; Controls for censoring (withdrawn and expired listings) ; Straightforward to implement
  – Findings
    • Important to account for right-censoring
    • Less important to control for observed and unobserved property heterogeneity
Outline

1. **MLS Data**
   - Fairfax (MLS); CoreLogic LLC

2. Conventional TOM indicators

3. Accounting for right censoring

4. Accounting for right censoring and property heterogeneity
   - Observed heterogeneity: “hedonic index”
   - Unobserved heterogeneity: “Repeat TOM index”
     - Proportional hazard
     - Median
1. MLS Data: Fairfax, VA

- Collected from Washington DC local MLS (MRIS)

- Rich dataset:
  - 1997 - 2010
  - Detailed home’s characteristics
  - Unit’s address
    - Zipcode / Census Block Group
  - About 0.3 Million records
1. MLS Data: Fairfax, VA

• Measuring TOM

  – Difference between original (first) listing and contract date
    • We are able to identify if same listing has been posted several times consecutively (“refreshers”) (may be missing some, though)
## Descriptive Statistics

### Fairfax County

**List Prices, Transaction Prices and Marketing Time**

<table>
<thead>
<tr>
<th>Year</th>
<th>List Price ($ 000)</th>
<th>Sale Price ($ 000)</th>
<th>Time on the market (days)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>St. Dev.</td>
<td>Mean</td>
</tr>
<tr>
<td>1998</td>
<td>233.4</td>
<td>199.9</td>
<td>126.5</td>
<td>228.3</td>
</tr>
<tr>
<td>2003</td>
<td>367.2</td>
<td>322.5</td>
<td>199.1</td>
<td>365.3</td>
</tr>
<tr>
<td>2008</td>
<td>439.9</td>
<td>379.9</td>
<td>258.1</td>
<td>426.3</td>
</tr>
</tbody>
</table>

Notes: Table shows descriptive statistics of Fairfax County, VA, residential real estate listings on the MLS and sold during each of the time periods specified above.
## Descriptive Statistics

### Fairfax County

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bedrooms</td>
<td>Total number of bedrooms</td>
<td>3.327</td>
<td>1.029</td>
</tr>
<tr>
<td>Bathrooms</td>
<td>Total number of full and half bathrooms</td>
<td>3.036</td>
<td>1.079</td>
</tr>
<tr>
<td>Age</td>
<td>Age of unit in years (zero if new)</td>
<td>23.49</td>
<td>15.42</td>
</tr>
<tr>
<td>Fireplaces</td>
<td>Total number of fireplaces</td>
<td>0.913</td>
<td>0.715</td>
</tr>
<tr>
<td>Central heat</td>
<td>Indicator if unit has central heat</td>
<td>0.928</td>
<td>0.258</td>
</tr>
<tr>
<td>Detached</td>
<td>Indicator if single family unit</td>
<td>0.477</td>
<td>0.499</td>
</tr>
<tr>
<td>Townhome</td>
<td>Indicator if unit is townhome</td>
<td>0.370</td>
<td>0.483</td>
</tr>
<tr>
<td>Apartment</td>
<td>Indicator if unit is in an apartment complex</td>
<td>0.153</td>
<td>0.360</td>
</tr>
</tbody>
</table>

Notes: Table shows descriptive statistics of Fairfax County, VA, residential real estate units that were listed on the MLS and sold between 1997 and 1999. The total number of observations is 243,182.
1. MLS Data: US MSAs

• Gathering MLS data for all regions in the U.S. involves a substantial effort
  – Listings data are collected and owned by Real Estate Agent’s associations
  – Legal agreements and a careful data validation process are needed

• Private data provider
  – Collects / validates MLS data from over 100 MSAs (Real Estate Suite)
  – Our data: All listings available in 13 US MSAs
    • 1.4 Million listings
    • 2004 - 2013
## MLS Data: US MSAs

<table>
<thead>
<tr>
<th>MSA Name</th>
<th># Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann Arbor, MI MSA</td>
<td>21,096</td>
</tr>
<tr>
<td>Durham, NC MSA</td>
<td>965</td>
</tr>
<tr>
<td>Gainesville, GA MSA</td>
<td>6,552</td>
</tr>
<tr>
<td>Honolulu, HI MSA</td>
<td>51,560</td>
</tr>
<tr>
<td>Las Vegas-Paradise, NV MSA</td>
<td>227,571</td>
</tr>
<tr>
<td>Medford, OR MSA</td>
<td>16,305</td>
</tr>
<tr>
<td>Miami-Miami Beach-Kendall, FL MD</td>
<td>112,144</td>
</tr>
<tr>
<td>New Orleans-Metairie-Kenner, LA MSA</td>
<td>75,188</td>
</tr>
<tr>
<td>Olympia, WA MSA</td>
<td>21,732</td>
</tr>
<tr>
<td>San Diego-Carlsbad-San Marcos, CA MSA</td>
<td>190,592</td>
</tr>
<tr>
<td>San Luis Obispo-Paso Robles, CA MSA</td>
<td>17,908</td>
</tr>
<tr>
<td>Santa Barbara-Santa Maria, CA MSA</td>
<td>19,670</td>
</tr>
<tr>
<td>Youngstown-Warren-Boardman, OH-PA MSA</td>
<td>27,066</td>
</tr>
</tbody>
</table>
## Descriptive Statistics

**US MSAs**

### List Prices, Transaction Prices and Marketing Time: San Diego

<table>
<thead>
<tr>
<th>Year</th>
<th>List Price ($ 000)</th>
<th>Sale Price ($ 000)</th>
<th>Time on the market (days)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>St. Dev.</td>
<td>Mean</td>
</tr>
<tr>
<td>2004</td>
<td>552.2</td>
<td>489.9</td>
<td>266.2</td>
<td>530.1</td>
</tr>
<tr>
<td>2008</td>
<td>438.3</td>
<td>365.0</td>
<td>297.0</td>
<td>397.8</td>
</tr>
<tr>
<td>2012</td>
<td>411.2</td>
<td>349.0</td>
<td>267.6</td>
<td>390.7</td>
</tr>
</tbody>
</table>

Notes: Table shows descriptive statistics of San Diego residential real estate listings on the MLS and sold during each of the time periods specified above.
Outline

1. MLS Data
2. Conventional TOM indicators
3. Accounting for right censoring
4. Accounting for right censoring and property heterogeneity
   - Observed heterogeneity: “hedonic index”
   - Unobserved heterogeneity: “Repeat TOM index”
     • Proportional hazard
     • Median
Mean and Median TOM of Home Sales by Year
- Fairfax County -

Note: Sample excludes withdrawn or expired listings.
Conventional indicators

- Exclude listings that are withdrawn from the market
- Exclude listings that expired
Fairfax County

Median TOM of Home Sales and Listings' Success Rate - Fairfax County -

Number of days on the market

Year when listing was posted

Percentage of listings sold

Note: Sample excludes withdrawn or expired listings.
Conventional indicators

• Exclude listings that are withdrawn from the market
• Exclude listings that expired

• Does accounting for censoring matter?
Outline

1. MLS Data
2. Conventional TOM indicators
3. Accounting for right censoring
4. Accounting for right censoring and property heterogeneity
   - Observed heterogeneity: “hedonic index”
   - Unobserved heterogeneity: “Repeat TOM index”
     • Proportional hazard
     • Median
Accounting for censoring

• Assumption:
  1. Time on the market is subject to random censoring

• Treat withdrawn and expired listings as right-censored observations
  – Kaplan – Meier estimator
    • Non-parametric estimate of the unconditional distribution of TOM
Accounting for censoring matters!

Median TOM Adjusting for Censoring by Year
- Fairfax County -

Note: Adjusted sample includes withdrawn and expired listings.
Outline

1. MLS Data
2. Conventional TOM indicators
3. Accounting for right censoring
4. Accounting for right censoring and property heterogeneity
   - **Observed heterogeneity:** “hedonic index”
   - **Unobserved heterogeneity:** “Repeat TOM index”
     - Proportional hazard
     - Median
Controlling for Observed Heterogeneity

• Select a base period (say year, 2000)
• Simulate distribution of TOM on other periods (say 2004) assuming that housing characteristics remain constant as in base year

• Methods:
  – Based on “sample reweighting:” Extension of Dinardo, Fortin and Lemieux (DFL), EMA 1996
  
  – Intuition: If a property listed in period t was more likely to appear in the base period, this observation receives a higher sampling weight
Dinardo, Fortin and Lemieux (DFL)

Random vector: $[y, X, T]$

Marginal density: $f(y \mid T = t_j); j = 0, 1$

$$f(y \mid T = t_0) = \int f(y \mid x, T = t_0) h(x \mid T = t_0) \, dx$$

$$f(y \mid T = t_1) = \int f(y \mid x, T = t_1) h(x \mid T = t_1) \, dx$$

Counterfactual density:

“Density of $Y \mid T_1$ if $X$ were the same as in $T_0$”

$$f(y)_{x_1 \rightarrow x_0} = \int f(y \mid x, T = t_1) h(x \mid T = t_0) \, dx$$

$$f(y)_{x_1 \rightarrow x_0} = \int f(y \mid x, T = t_1) h(x \mid T = t_1) \frac{h(x \mid T = t_0)}{h(x \mid T = t_1)} \, dx$$

$$f(y)_{x_1 \rightarrow x_0} = \int f(y \mid x, T = t_1) h(x \mid T = t_1) \tau_{x_1 \rightarrow x_0}(x) \, dx$$
Counterfactual TOM density

Random vector: \([y, X, T]\)

Marginal density:
\[ f(y \mid T = t_j); j = 0, 1 \]
\[ f(y \mid T = t_0) = \int f(y \mid x, T = t_0) h(x \mid T = t_0) dx \]
\[ f(y \mid T = t_1) = \int f(y \mid x, T = t_1) h(x \mid T = t_1) dx \]

Use Kaplan - Meier method to estimate counterfactual density:

“Density of \( Y \mid T1 \) if \( X \) were the same as in \( T0 \)”
\[ f(y)_{x_1 \rightarrow x_0} = \int f(y \mid x, T = t_1) h(x \mid T = t_0) dx \]
\[ f(y)_{x_1 \rightarrow x_0} = \int f(y \mid x, T = t_1) h(x \mid T = t_1) \frac{h(x \mid T = t_0)}{h(x \mid T = t_1)} dx \]
\[ f(y)_{x_1 \rightarrow x_0} = \int f(y \mid x, T = t_1) h(x \mid T = t_1) \tau_{x_1 \rightarrow x_0}(x) dx \]
Counterfactual TOM density

• Use logit model to estimate

\[ \hat{P}(T = t_0 \mid X = x) \]

• Estimate weights for \( T = t1 \)

\[ \hat{t}_{x_1 \rightarrow x_0}(x) = \frac{\hat{P}(T = t_0 \mid X = x)}{1 - \hat{P}(T = t_0 \mid X = x)} \bigg/ \frac{\hat{P}(T = t_0)}{1 - \hat{P}(T = t_0)} \]

• Estimate TOM density for \( T = t1 \) using KM estimator and DFL weights
Counterfactual Median TOM
Constant Home Characteristics (as in 2000)

Median TOM Adjusting for Censoring and Heterogeneity
- Fairfax County -

Note: Adjusted sample includes withdrawn and expired listings.
Controlling for Observed Heterogeneity

• Controlling for observed home heterogeneity does not seem to substantially affect estimated of TOM
  – Consistent with low R2 in TOM regressions

• Unobserved heterogeneity could be important!
  – Repeat TOM index
Outline

1. MLS Data
2. Conventional TOM indicators
3. Accounting for right censoring
4. Accounting for right censoring and property heterogeneity
   - Observed heterogeneity: “hedonic index”
   - Unobserved heterogeneity: “Repeat TOM index”
     • A. Proportional hazard
     • B. Accelerated Failure Time Model
A. Proportional Hazard

- Proportional hazard model

\[ \lambda_{it}(y) = \exp(\beta_t)\lambda_{0i}(y) \]

- Home \( i \), period \( t \)
- \( \lambda_{0i}(\cdot) \) **individual** home baseline hazard

- \( \beta_t \) are period specific shifters of baseline hazard (basic input for TOM index)
A. Proportional Hazard

- Proportional hazard model

\[ \lambda_{it}(y) = \exp(\beta_t) \lambda_{0i}(y) \]

- Standard result:

\[ -\ln(\Lambda_{0i}(Y_{is})) = \beta_{t_i} + \varepsilon_{i}^{s} \]

\[ -\varepsilon_{i}^{s} \mid t_{i}^{s} \sim EV1(0, 1) \]

- Log of integrated baseline hazard has an extreme value distribution
A. Proportional Hazard

- Assume:
  \[ \lambda_{0i}(y) = \exp(\alpha_i)\lambda_0(y) \]
- Baseline hazard has a “fixed effect”
  \[ -\ln(\Lambda_0(Y_{i}^s)) = \beta_{t_{i}^s} + \alpha_i + \varepsilon_{i}^s \]
- We cannot simply “difference out” \( \alpha \) (why?)
- But we can compute
  \[ Pr(\ln(\Lambda_0(Y_{i}^2)) \geq \ln(\Lambda_0(Y_{i}^1))) \]
A. Proportional Hazard

• Formally:

Assumption 3.1. Once we control for each property baseline hazard, \( Y \) is i.i.d.

(i) \( Y_{i}^1, Y_{i}^2 \) are independent conditional on \( t_{i}^1, t_{i}^2, \lambda_{0i}(\cdot) \)

(ii) \( Y_{i}^s \mid t_{i}^1, t_{i}^2, \lambda_{0i}(\cdot) = d \ Y_{it_{i}^s} \mid \lambda_{0i}(\cdot) \)

• Then

\( -\varepsilon_{i}^1 \) and \( -\varepsilon_{i}^2 \) are independent and both distributed \( EV1(0, 1) \)
A. Proportional Hazard

• Formally:

\[
Pr(Y_i^2 \geq Y_i^1 \mid t_i^1 = t_1, t_i^2 = t_2) = Pr(\ln(\Lambda_0(Y_i^2)) \geq \ln(\Lambda_0(Y_i^1)) \mid t_i^1 = t_1, t_i^2 = t_2)
\]

\[
= Pr(\varepsilon_i^2 - \varepsilon_i^1 \geq \beta_{t_2} - \beta_{t_1} \mid t_i^1 = t_1, t_i^2 = t_2)
\]

\[
= \frac{\exp(\beta_{t_1})}{\exp(\beta_{t_1}) + \exp(\beta_{t_2})}
\]

• We can estimate the coefficients of interest using a simple logit regression!
  — Approach is based on Lancaster (2000) and Chamberlain (1985). Couple papers in labor economics have used a somewhat similar approach
A. Proportional Hazard

• Estimation procedure:
  - For each pair of repeat-sales

  - Let \( W_i = 1 \) if \( Y_{2i} > Y_{1i} \)
  - Let \( X_{ij} = -1 \) if the \textit{first} sale was made in period \( j \); Let \( X_{ij} = +1 \) if the \textit{second} sale was made in period \( j \); and \( X_{ij} = 0 \) otherwise

  - Estimate \( \Pr\{ W=1 \} = \text{Logit} (X^*\beta) \)
A. Proportional Hazard

- What about censoring?
  - Random censoring: \( Y_{\text{SOLD}} > Y^* \) with withdrawn

<table>
<thead>
<tr>
<th>Home</th>
<th>Period 0 (( Y_0^* ))</th>
<th>Period 1 (( Y_1^* ))</th>
<th>( Y_1^* &gt; Y_0^* )</th>
<th>( Y_1 &gt; Y_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Sold</td>
<td>Sold</td>
<td>YES</td>
<td>( Y_1 &gt; Y_0 )</td>
</tr>
<tr>
<td>B</td>
<td>Withdrawn</td>
<td>Sold</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Sold</td>
<td>Withdrawn</td>
<td>YES</td>
<td>( Y_1 &gt; Y_1^* &gt; Y_0 )</td>
</tr>
<tr>
<td>D</td>
<td>Withdrawn</td>
<td>Withdrawn</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>

Use A and C for estimation! (We show estimator is consistent)
A. Proportional Hazard

• Repeat Proportional Hazard Index

\[ \mu_t = \exp(\beta_t) \]

• \((\mu_t - 1) \times 100\) : Gross percentage increase (decrease) in the hazard rate compared to the base period
B. Accelerated failure time model

• Linear model:

\[
\log(Y_i^s) = \beta_i + u_i^s
\]

\[
u_i^s = \alpha_i + \varepsilon_i^s
\]

• Take differences:

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

• If no-censoring, OLS could be used to estimate \( \beta \)
B. Accelerated failure time model

- With censoring:
  \[ \log(Y_i^s) = \beta_{t_i^s} + u_i^s \]
  \[ u_i^s = \alpha_i + \varepsilon_i^s \]

- Assumption:
  \[(i) \ Med(\alpha_i + \varepsilon_i^1 | t_i^1, t_i^2) = Med(\alpha_i + \varepsilon_i^2 | t_i^1, t_i^2) \]
  \[(ii) Y_i^s \ and \ C_i^s \ are \ independent \ conditional \ on \ t_i^1, t_i^2 \]

- Proposed method:
  - Use sample of repeat listings to estimate \( \Delta Med(Y_i^s) \)
  - Use OLS to estimate \( \beta \)
B. Accelerated failure time model

- Details:
- Step 1:
  - Select a repeat **listing** sample such that \( t_{i1}, t_{i2} = t_1, t_2 \)
  - For each pair of \( t_1, t_2 \), such that \( t_1 < t_2 \), use Kaplan Meier estimator to compute \( \hat{\delta M_i} \):
    \[
    Med(\log(Y_i^2) \mid X_i) - Med(\log(Y_i^1) \mid X_i)
    \]
- Step 2:
  - Use OLS to estimate regression of \( \hat{\delta M_i} \) on \( X_i \).
B. Accelerated failure time model

- Repeat Median TOM Index

\[ \mu_t = \exp(\beta_t) \]

- \((\mu_t - 1) \times 100\) : Gross percentage increase (decrease) in the median TOM compared to the base period
(Repeat) Median TOM Index
- Fairfax County - 2003 = 1 -
(Repeat) Median TOM Index
- Las Vegas - 2007. I = 1 -
Summary

• Proposed two methods to measure changes in housing marketing time
  – A. Repeat proportional hazard index
  – B. Repeat median TOM index

• Methods control for censoring and (unobserved) individual heterogeneity

• Methods are straightforward to implement (particularly A)
Next steps

• Comprehensive literature review

• Standard errors

• Robustness of results (How to treat subsequent repeat listings?)

• Use listings data to adjust hedonic home price index?
Thank you