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

Market area	Median Price / change	Median TOM	Δ Median TOM
A	$200K / 5%	15	0
B	$200K / 5%	60	+40

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    - Exploits repeat-sales ; Controls for censoring (withdrawn and expired listings) ; Straightforward to implement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 heterogenity
   – 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

	List Price ($ 000)			Sale Price ($ 000)			Time on the market (days)			
Year	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.	N
1998	233.4	199.9	126.5	228.3	196.5	123.8	66.6	41.0	76.7	16,248
2003	367.2	322.5	199.1	365.3	321.4	194.9	20.7	9.0	31.8	22,846
2008	439.9	379.9	258.1	426.3	367.9	244.4	62.7	40.0	67.0	13,829

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

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

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

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

Descriptive Statistics
US MSAs

List Prices, Transaction Prices and Marketing Time: San Diego

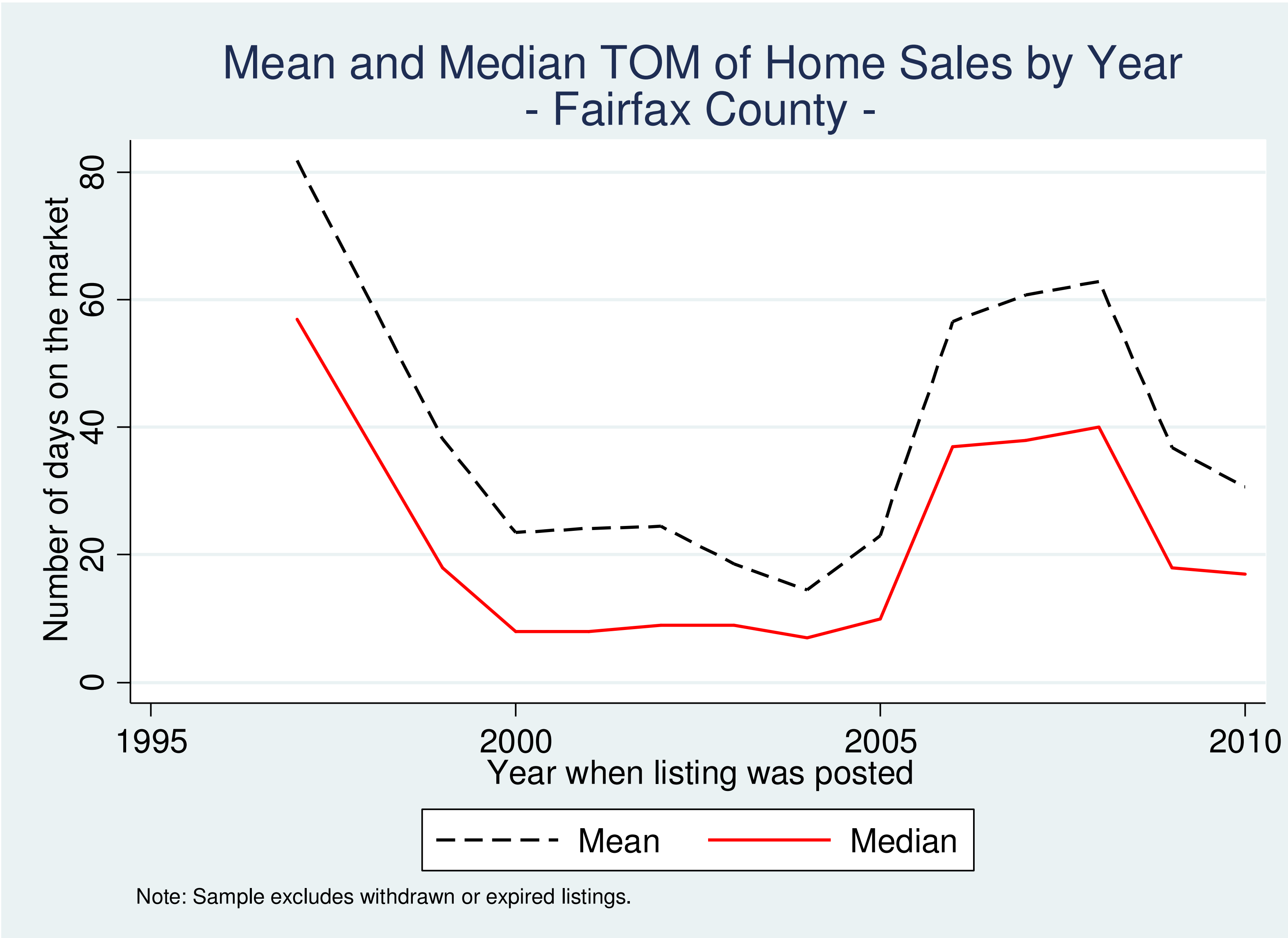
	List Price ($ 000)			Sale Price ($ 000)			Time on the market (days)			
Year	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.	N
2004	552.2	489.9	266.2	530.1	474.0	243.4	28.6	17.0	32.1	25,696
2008	438.3	365.0	297.0	397.8	337.0	260.7	69.6	45.0	71.0	19,445
2012	411.2	349.0	267.6	390.7	334.5	244.8	72.5	39.0	87.0	23,577

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 heterogenity
   - Observed heterogeneity: “hedonic index”
   - Unobserved heterogeneity: “Repeat TOM index”
     - Proportional hazard
     - Median

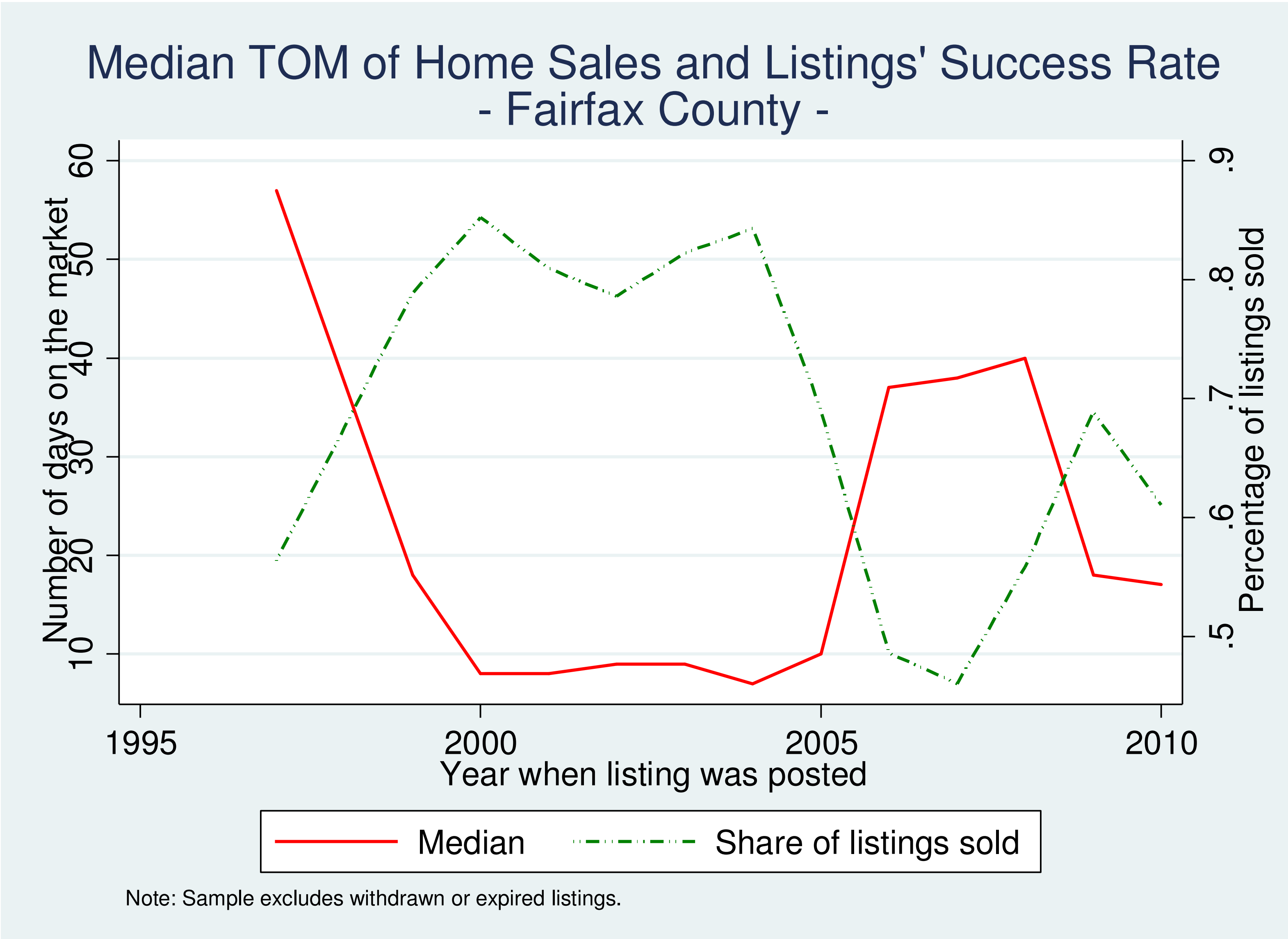
Fairfax County



Conventional indicators

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

Fairfax County



Conventional indicators

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

- Does accounting for censoring matter?

Outline

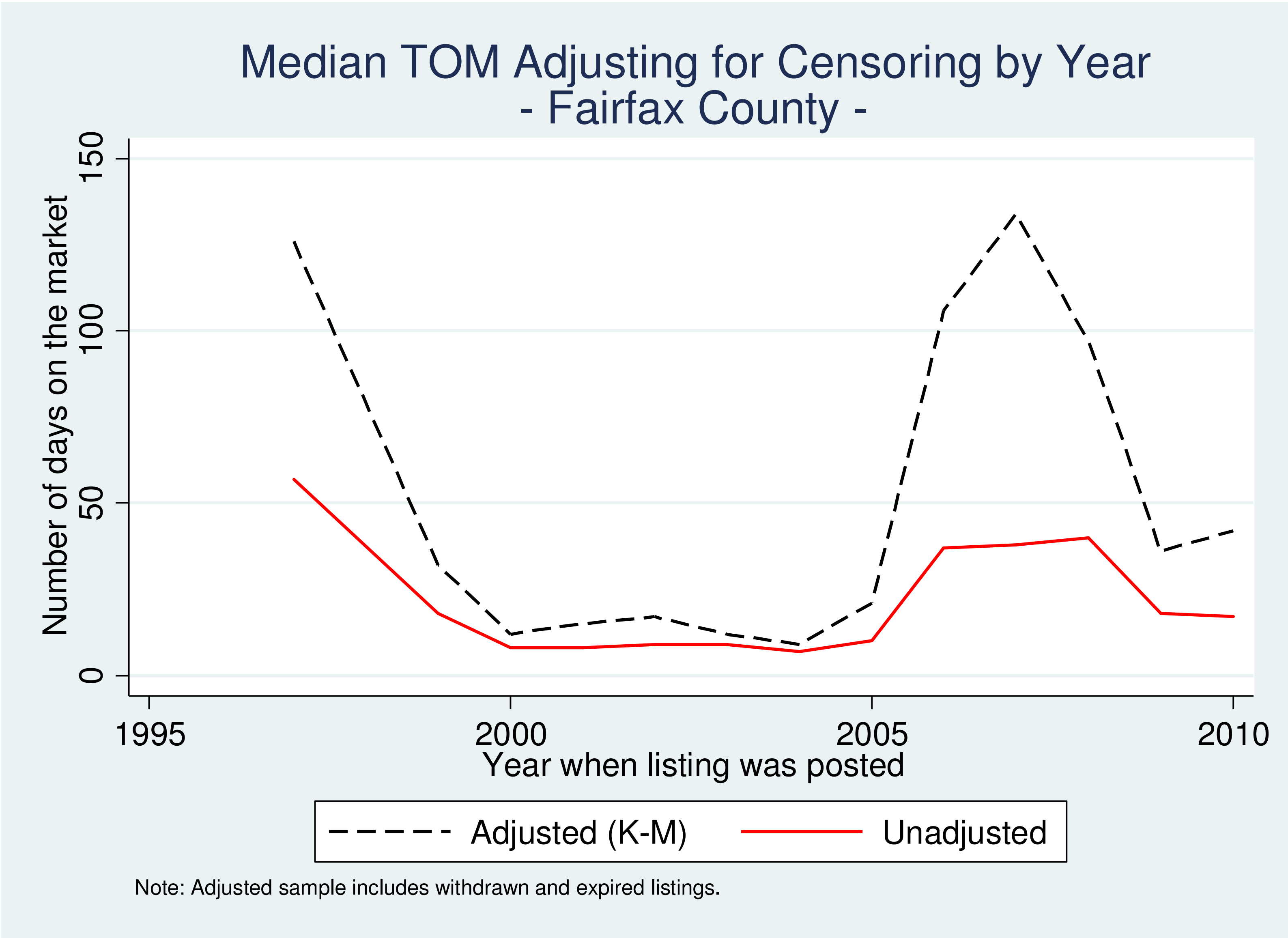
1. MLS Data
2. ConventionalTOM indicators
3. **Accounting for right censoring**
4. Accounting for right censoring and property heterogenity
   - Observed heterogeneity: “hedonic index”
   - Unobserved heterogeneity: “Repeat TOM index”
     - Proportional hazard
     - Median

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!



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

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

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

Counterfactual density:

"Density of Y | T1 if X were the same as in T0"

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

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

$$f(y)_{x_1 \to x_o} = \int \underline{f(y \mid x, T = t_1) h(x \mid T = t_1)} \tau_{x_1 \to x_0}(x) dx$$

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 | T1 if X were the same as in T0"

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

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

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

Counterfactual TOM density

- Use logit model to estimate

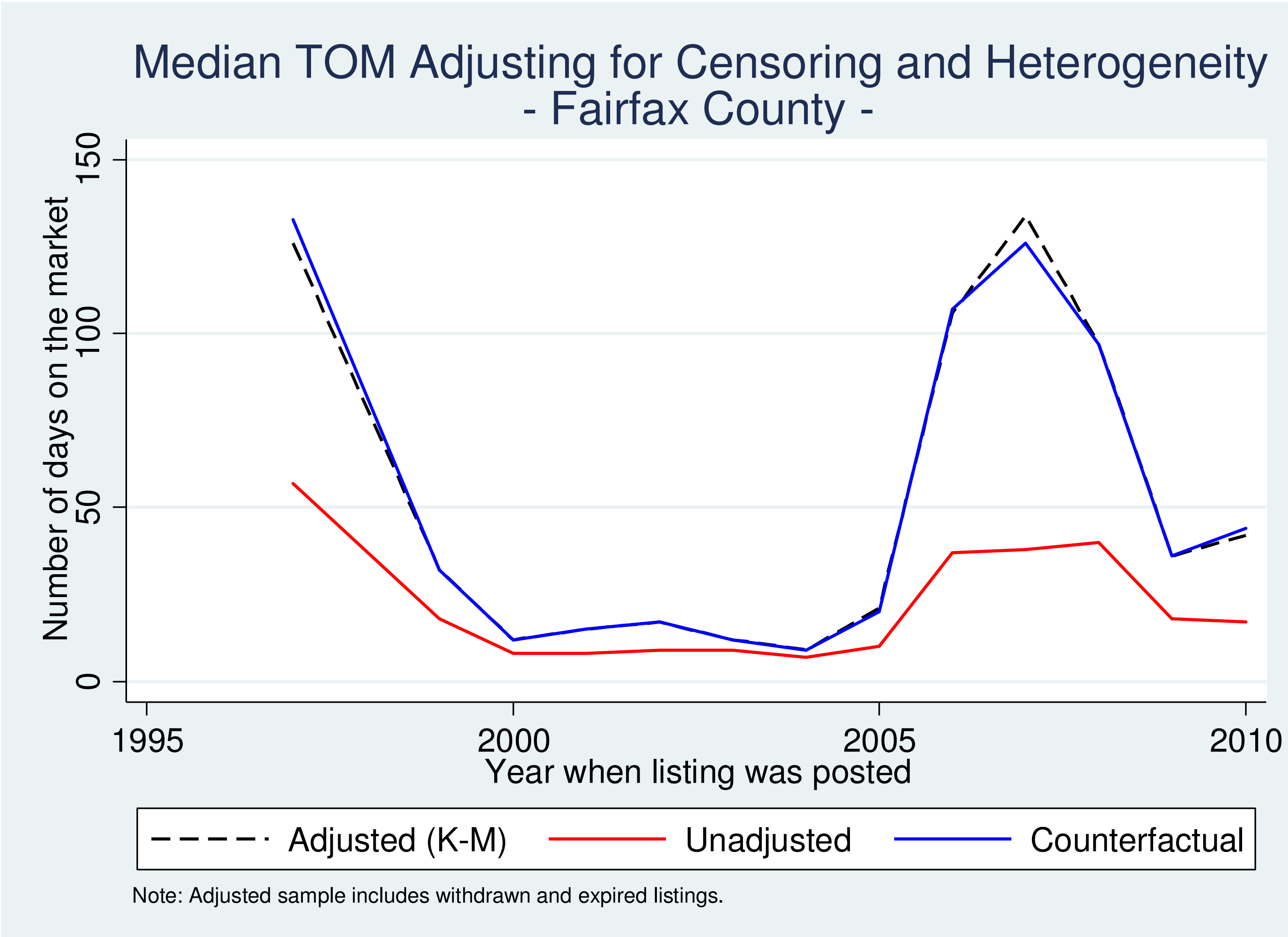
$$\hat{P}(T = t_0 \mid X = x)$$

- Estimate weights for T= t1

$$\hat{\tau}_{x_1 \to x_0}(x) = \frac{\hat{P}(T = t_0 \mid X = x)}{1 - \hat{P}(T = t_0 \mid X = x)} \Bigg/ \frac{\hat{P}(T = t_0)}{1 - \hat{P}(T = t_0)}$$

- Estimate TOM density for T=t1 using KM estimator and DFL weights

Counterfactual Median TOM
Constant Home Characteristics (as in 2000)



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

- β_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:

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

$$-\varepsilon_i^s \mid t_i^s \sim EV1(0,1)$$

  – 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" α (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 *first* sale was made in period *j*; Let $X_{ij} = +1$ if the *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: $Y_{SOLD} > Y^*_{\ WITHDRAWN}$

Home	Period 0 (Y0*)	Period 1 (Y1*)	Y1* > Y0*		Y1 > Y0?
A	Sold	Sold	YES	Y1 > Y0	YES
B	Withdrawn	Sold	YES		?
C	Sold	Withdrawn	YES	Y1>Y1*>Y0	YES
D	Withdrawn	Withdrawn	YES		?

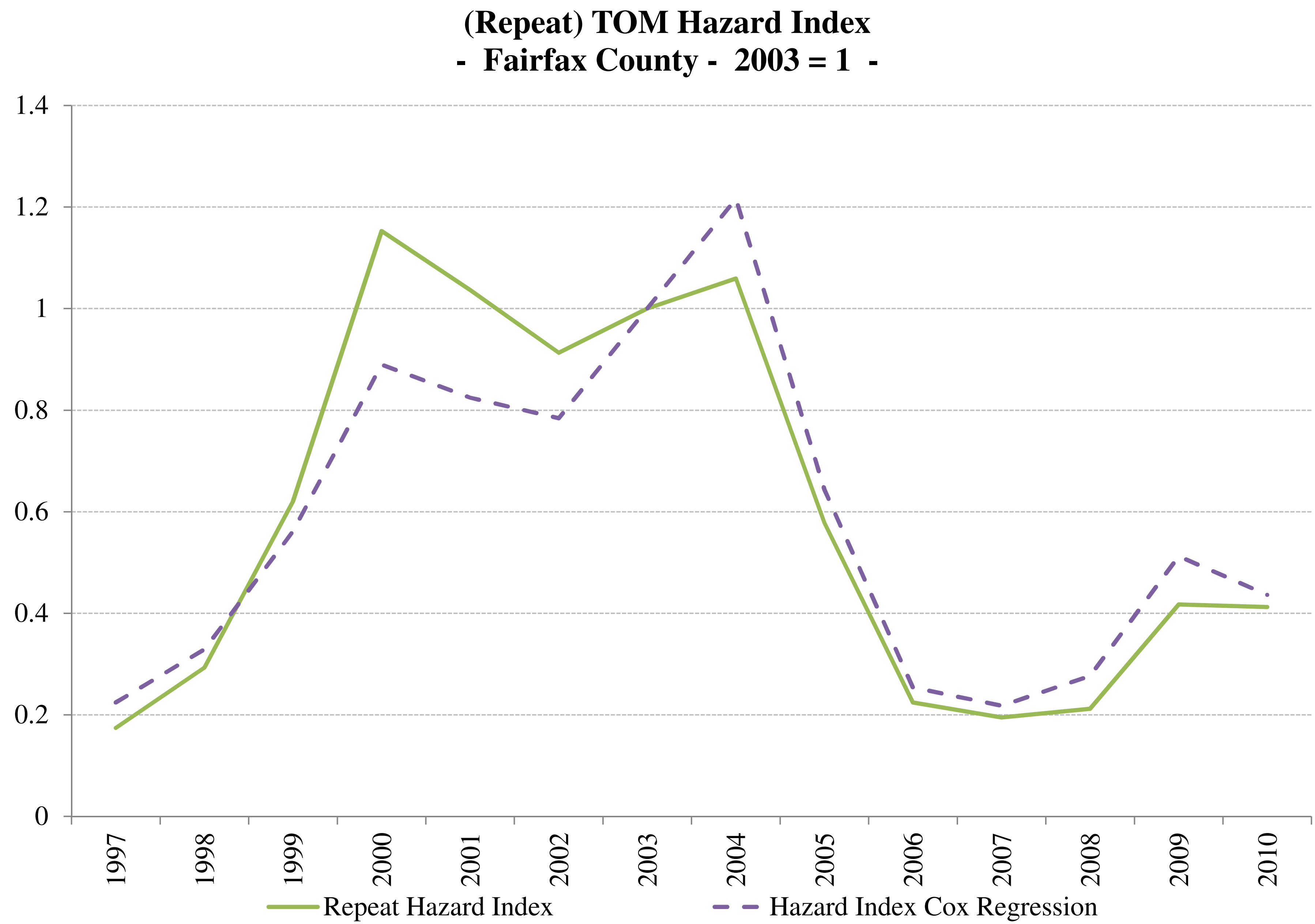
Use A and C for estimation! (We show estimator is consistent)

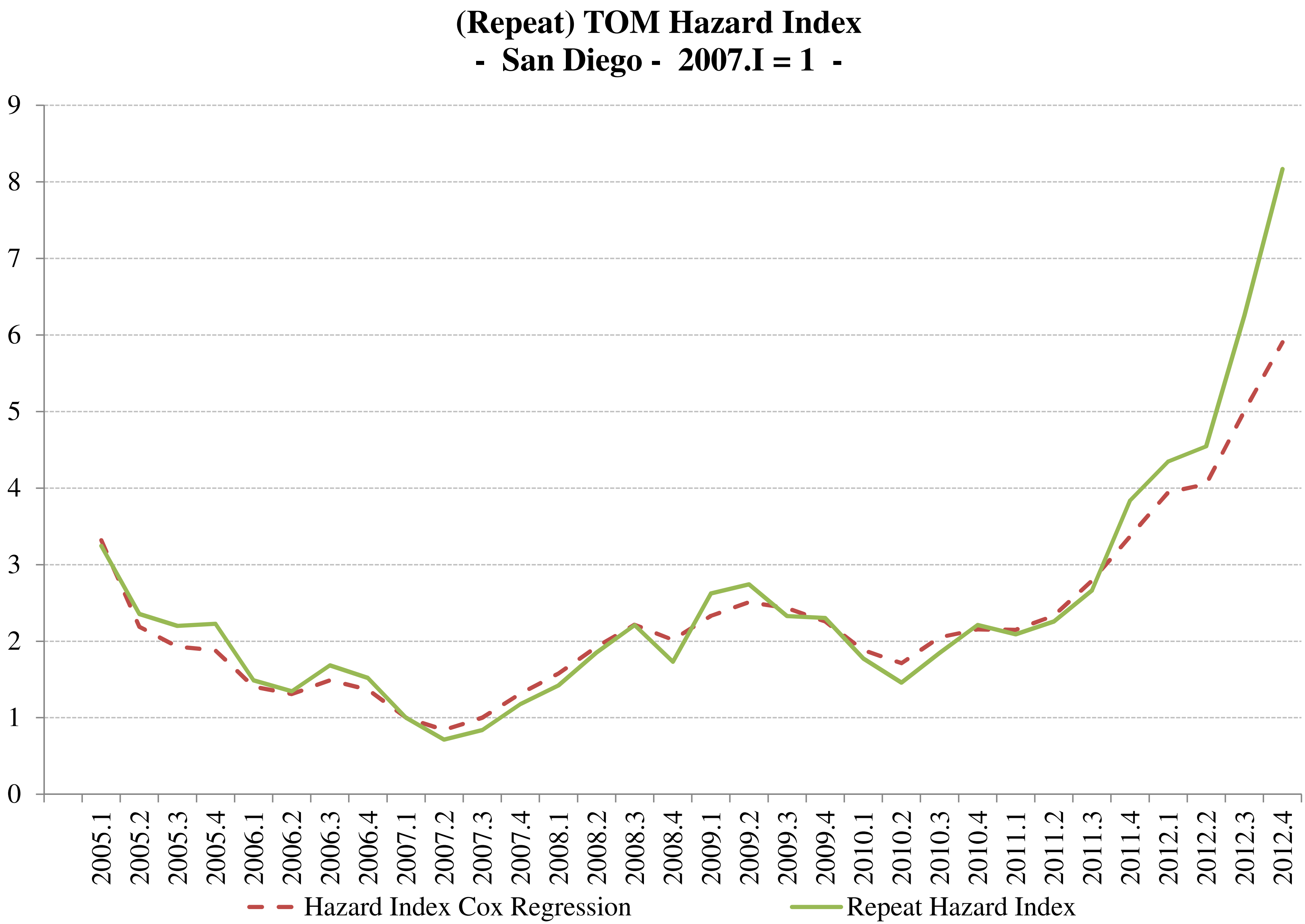
A. Proportional Hazard

- Repeat Proportional Hazard Index

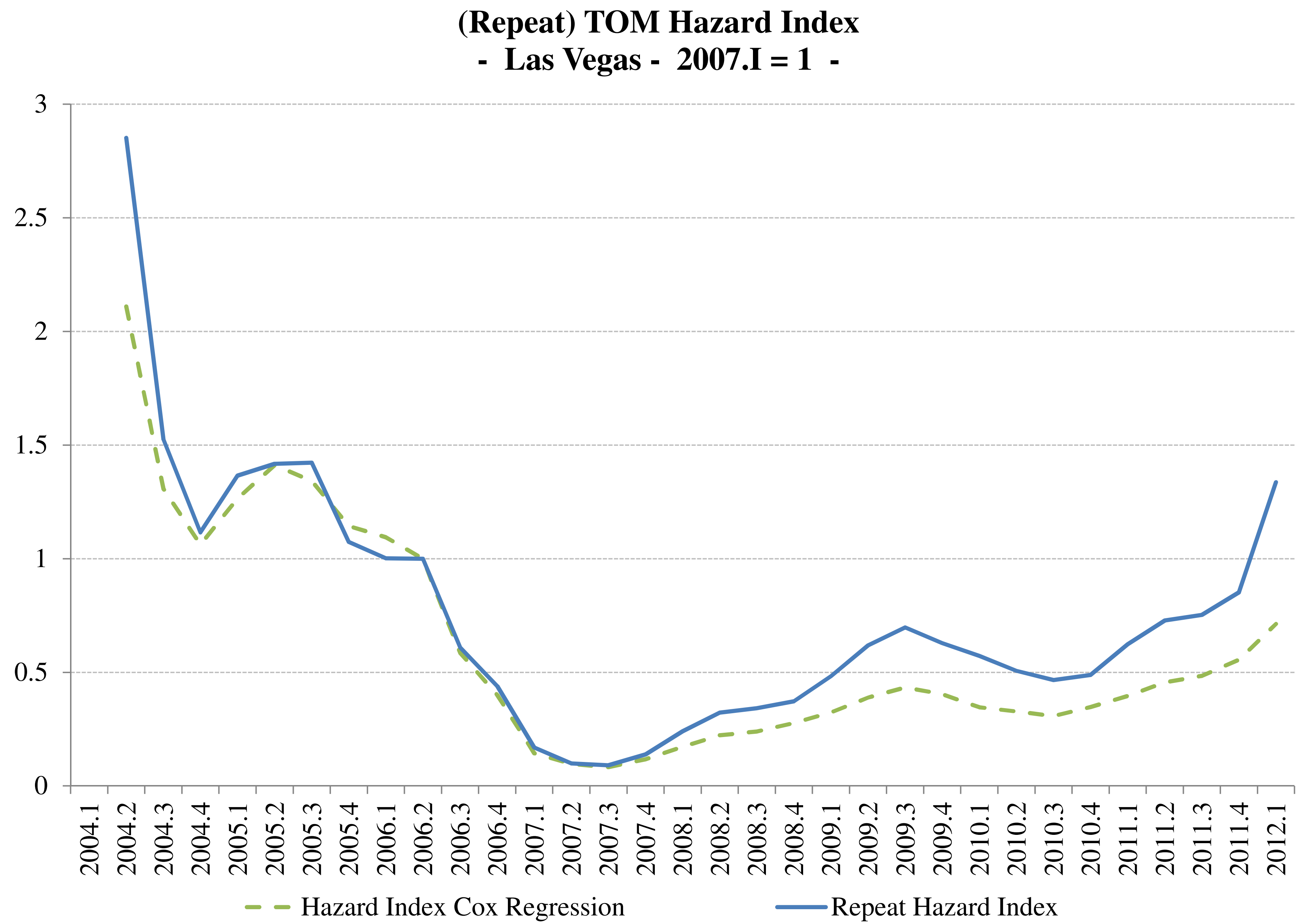
$$\mu_t = \exp(\beta_t)$$

- $(\mu_t - 1)*100$: Gross percentage increase (decrease) in the hazard rate compared to the base period


(Repeat) TOM Hazard Index
- Fairfax County - 2003 = 1 -
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Repeat Hazard Index
Hazard Index Cox Regression



(Repeat) TOM Hazard Index
- San Diego - 2007.I = 1 -
9
8
7
6
5
4
3
2
1
0
2005.1
2005.2
2005.3
2005.4
2006.1
2006.2
2006.3
2006.4
2007.1
2007.2
2007.3
2007.4
2008.1
2008.2
2008.3
2008.4
2009.1
2009.2
2009.3
2009.4
2010.1
2010.2
2010.3
2010.4
2011.1
2011.2
2011.3
2011.4
2012.1
2012.2
2012.3
2012.4
Hazard Index Cox Regression
Repeat Hazard Index


(Repeat) TOM Hazard Index
- Las Vegas - 2007.I = 1 -
3
2.5
2
1.5
1
0.5
0
2004.1
2004.2
2004.3
2004.4
2005.1
2005.2
2005.3
2005.4
2006.1
2006.2
2006.3
2006.4
2007.1
2007.2
2007.3
2007.4
2008.1
2008.2
2008.3
2008.4
2009.1
2009.2
2009.3
2009.4
2010.1
2010.2
2010.3
2010.4
2011.1
2011.2
2011.3
2011.4
2012.1
Hazard Index Cox Regression
Repeat Hazard Index

B. Accelerated failure time model

- Linear model:

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

- 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 β

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:

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

$$(ii)\ Y_i^s \ \ and\ C_i^s\ \ are\ \ independent\ \ conditional\ \ on\ t_i^1, t_i^2$$

- Proposed method:
  - Use sample of repeat listings to estimate ΔMed(Y^s)
  - Use OLS to estimate β

B. Accelerated failure time model

- Details:
- Step 1:
  - Select a repeat **listing** sample such that $t_i^1, t_i^2 = t_1, t_2$
  - For each pair of t_1, t_2, such that $t_1 < t_2$, use Kaplan Meier estimator to compute $\widehat{\delta M_i}$:

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

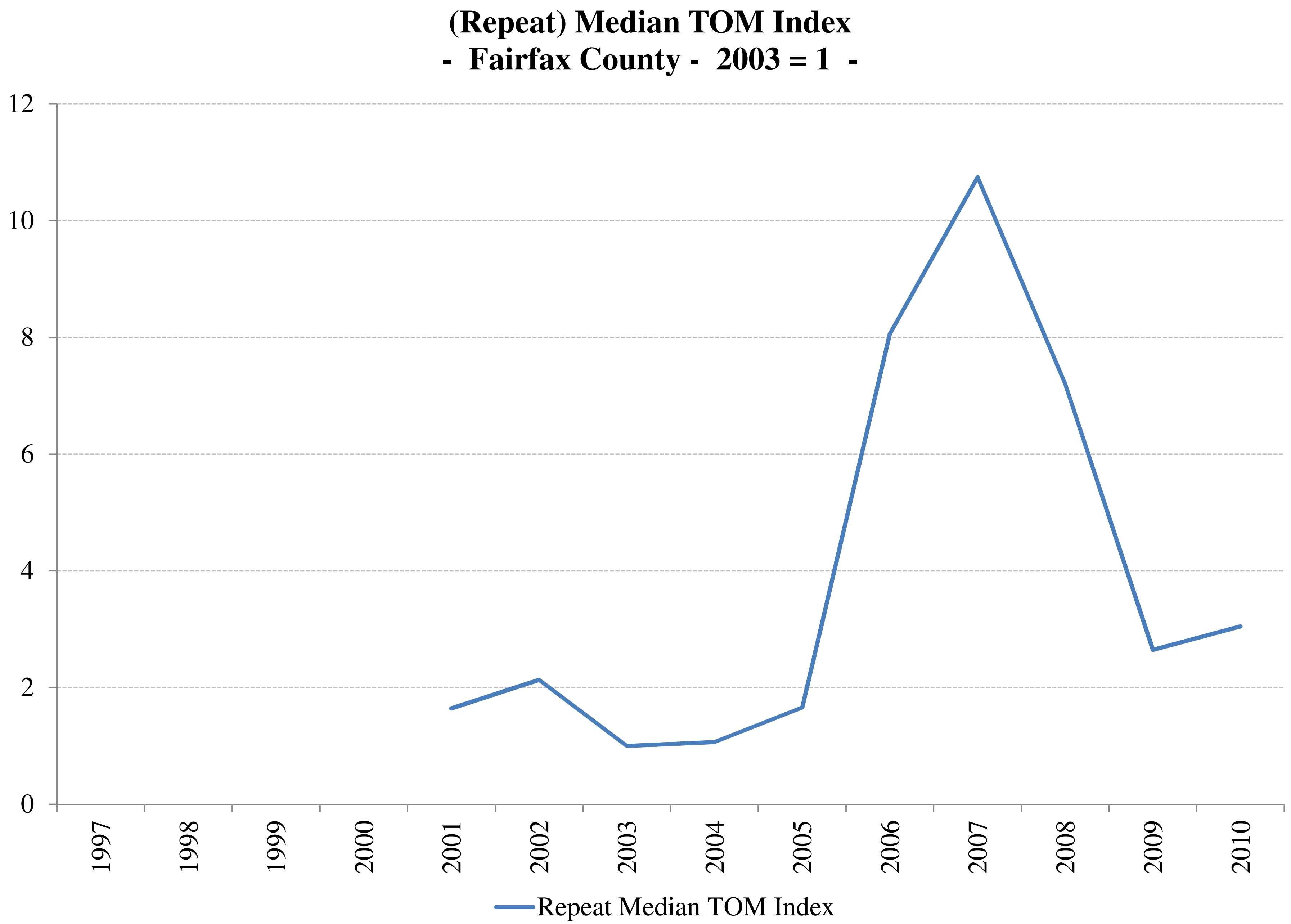
- Step 2:
  - Use OLS to estimate regression of $\widehat{\delta M_i}$ on X_i.

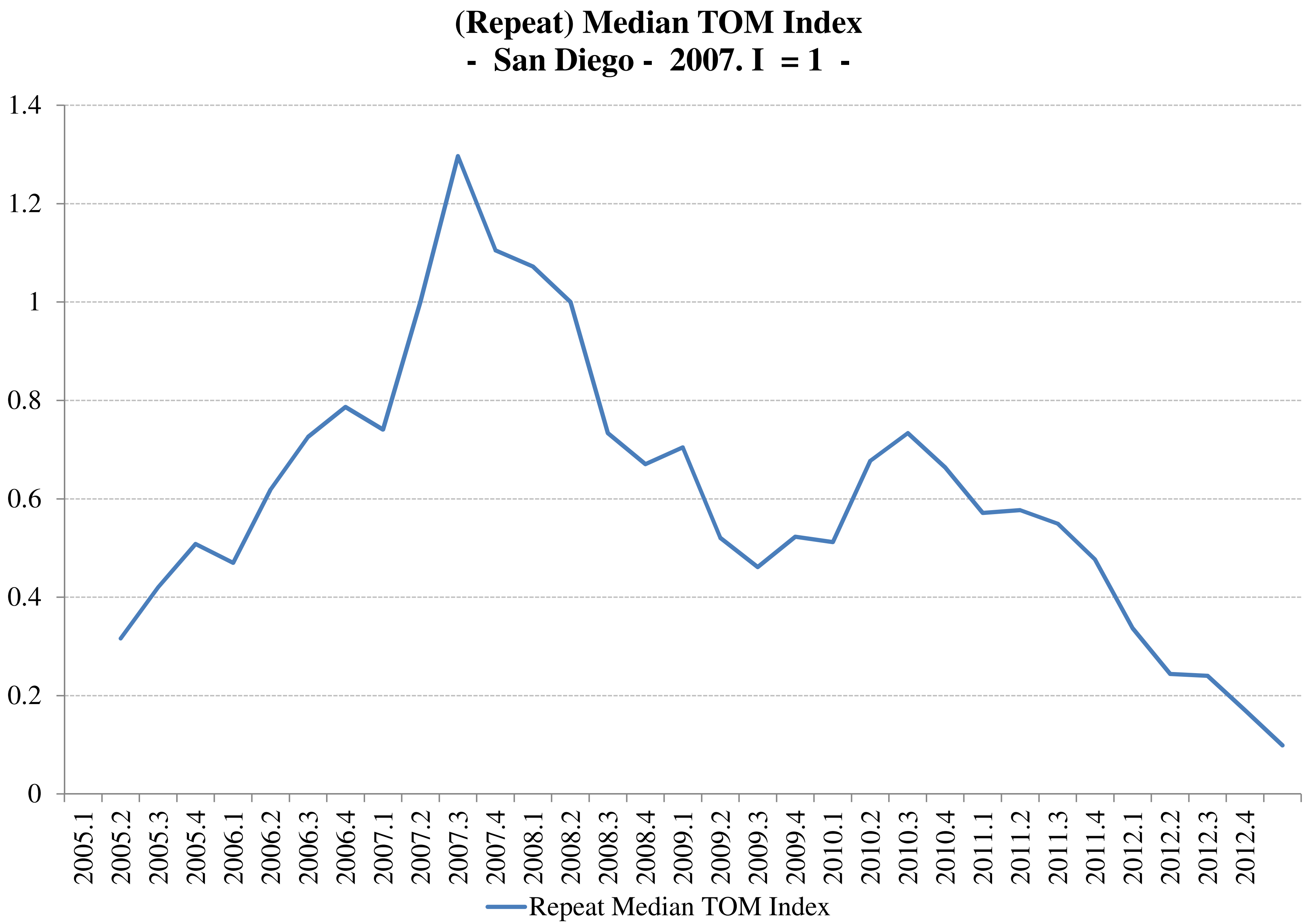
B. Accelerated failure time model

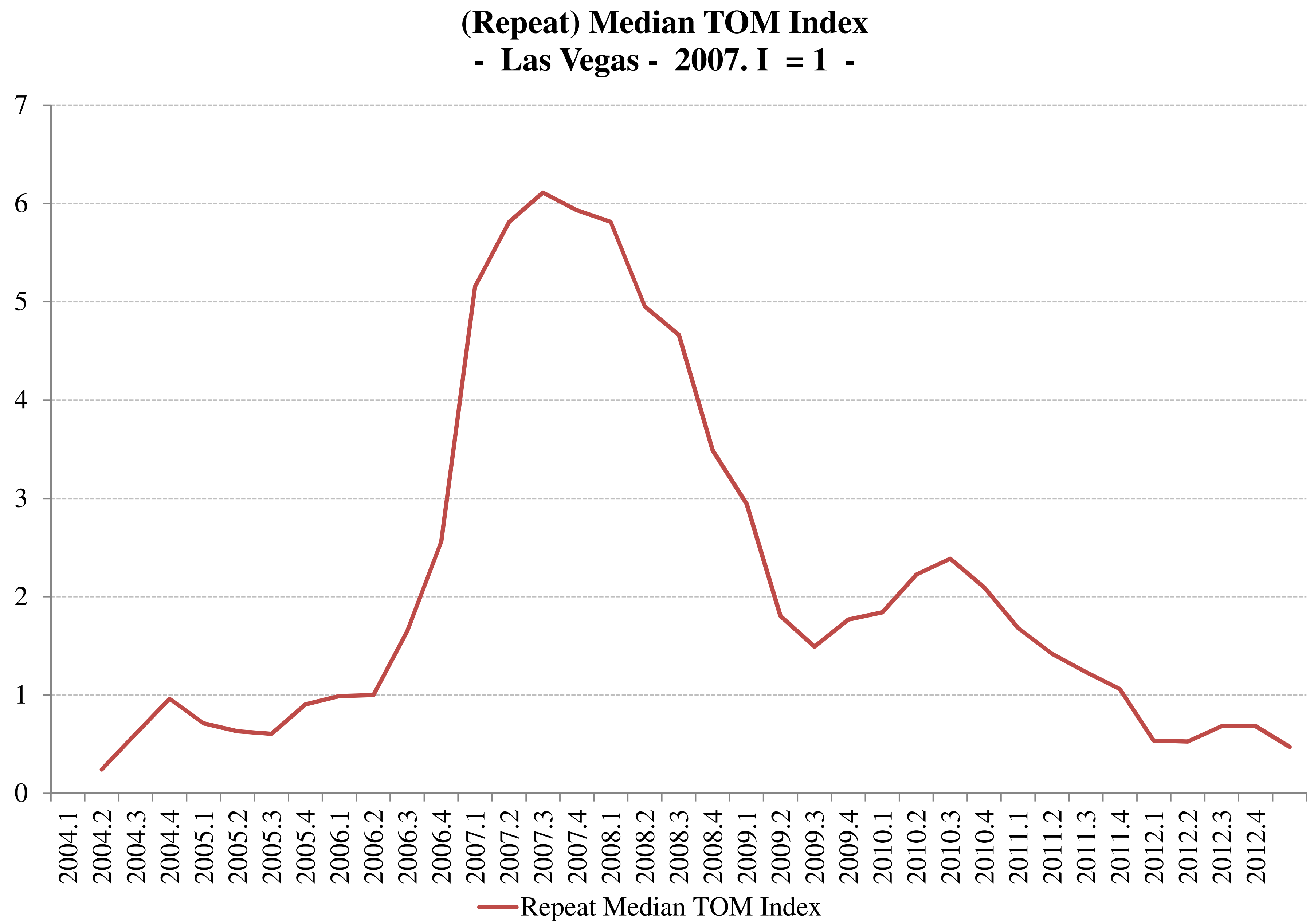
- Repeat Median TOM Index

$$\mu_t = \exp(\beta_t)$$

- $(\mu_t - 1)*100$: Gross percentage increase (decrease) in the median TOM compared to the base period


(Repeat) Median TOM Index
- Fairfax County - 2003 = 1 -
0
2
4
6
8
10
12
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Repeat Median TOM Index


(Repeat) Median TOM Index
- San Diego - 2007. I = 1 -
1.4
1.2
1
0.8
0.6
0.4
0.2
0
2005.1
2005.2
2005.3
2005.4
2006.1
2006.2
2006.3
2006.4
2007.1
2007.2
2007.3
2007.4
2008.1
2008.2
2008.3
2008.4
2009.1
2009.2
2009.3
2009.4
2010.1
2010.2
2010.3
2010.4
2011.1
2011.2
2011.3
2011.4
2012.1
2012.2
2012.3
2012.4
Repeat Median TOM Index


(Repeat) Median TOM Index
- Las Vegas - 2007. I = 1 -
0
1
2
3
4
5
6
7
2004.1
2004.2
2004.3
2004.4
2005.1
2005.2
2005.3
2005.4
2006.1
2006.2
2006.3
2006.4
2007.1
2007.2
2007.3
2007.4
2008.1
2008.2
2008.3
2008.4
2009.1
2009.2
2009.3
2009.4
2010.1
2010.2
2010.3
2010.4
2011.1
2011.2
2011.3
2011.4
2012.1
2012.2
2012.3
2012.4
Repeat Median TOM Index

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