House Price Distribution and Price Indexes in Tokyo

Singapore October 16, 2015

Yongheng Deng (IRES, NUS)
with

Xiangyu Guo (IRES, NUS)
Daniel McMillan (Illinois University)
Chihiro Shimizu (IRES, NUS)

Institute for Real Estate Studies
National University of Singapore

1. Introduction

Residential Property Price Indexes (RPPI)

- Why RPPI is important?
  - In the wake of the release of Handbook of Residential Property Price Indices, from EuroStat with United Nations, OECD, IMF and World Bank in 2012;
  - How should different countries construct residential property price indexes?

- Main methods on constructing RPPI:
  - Methodology : Hedonic, Repeat Sales, etc
  - Data Source : Registry, Realtor, Mortgage bank, or Listing prices

Limitations for traditional RPPI: Hedonic

- Limitations for **Hedonic Model**

  1. Omitted variable bias (Case and Quigley 1991; Ekeland, Heckman and Nesheim 2004; Shimizu 2009)

  2. Structural change (Case et al. 1991; Clapp et al. 1991; Clapp and Giaccotto 1992, 1998; Shimizu and Nishimura 2006, 2007, Shimizu, Takatsuji, Ono, and Nishimura 2007; McMillen 2008)

Limitations for traditional RPPI: Repeat Sales

- Limitations for **Repeat Sales**

1. Sample selection bias (Clapp and Giaccotto 1992)

2. Age problems, characteristics changes (Case and Quigley 1991; Case and Shiller,1987, 1989; Clapp and Giaccotto, 1992, 1998, 1999; Goodman and Thibodeau,1998; Case et al. 1991)

Adjustment RPPI

- Matching approach
  - McMillen (2012), Deng, McMillen and Sing(2012, 2014)
  - Matching approach based on average treatment effect solve the non-random and aged problems of Repeat Sales

- Decomposition of Distribution Index
  - Based on the decomposition of distribution change method by Machado and Mata(2005), we construct a decomposition of distribution index
  - This approach solve the attributes change problems of hedonic

Outline

1. Introduction
2. Measures of RPPI
   - I. Traditional time dummy Hedonic index
   - II. Case-Shiller Repeat Sales index
   - III. Matching index
   - IV. Decomposition of Distributions Indexes
3. Data Description
4. Empirical Results
5. Conclusion

2. Measures of RPPI

House Price Transaction Samples

Time										
i \ t	1	2	3	4	5	6	7	8	9	10
A*	$P_{A,1}$			$p_{A,4}$					$p_{A,9}$	
B								$p_{B,8}$		
C*		$p_{C,2}$		$p_{C,4}$			$p_{C,7}$			$p_{C,10}$
D						$p_{D,6}$				
E		$p_{E,2}$								
F					$p_{F,5}$					
G*			$p_{G,3}$				$p_{G,7}$			
H				$p_{H,4}$						
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Z*								$p_{Z,8}$		$p_{Z,10}$

$P_{i,t}$: property *i*, transaction time *t*, *Repeat Sales Samples

Standard Model for Quality Adjustment

- Traditional hedonic model

$$\ln P_{it} = \sum_{k=1}^{K} \beta_k X_{ikt} + \sum_{s=1}^{\tau} \delta_s D_s + \varepsilon_{it} \quad \Rightarrow \quad \ln(\hat{P}_t / \hat{P}_{t-1}) = \hat{\delta}_t - \hat{\delta}_{t-1}$$

$\hat{\delta}_t$: Time Dummy Parameter $\Rightarrow$ House Price Index

- Traditional repeat sales model
  - Bailey, Muth and Nourse (BMN 1963)

$$\ln P_{ht_1} = \sum_{k=1}^{K} \beta_k X_{hk} + \delta_1 + \boxed{\delta_{t_1}} + \boxed{\varepsilon_{ht_1}}$$

$$\ln P_{ht_2} = \sum_{k=1}^{K} \beta_k X_{hk} + \delta_1 + \boxed{\delta_{t_2}} + \boxed{\varepsilon_{ht_2}}$$

$$\Rightarrow \quad \ln(P_{ht_2} / P_{ht_1}) = \boxed{\delta_{t_2} - \delta_{t_1}} + \boxed{\left(\varepsilon_{ht_2} - \varepsilon_{ht_1}\right)}$$

Adjustment to Repeat Sales Index

- Case-Shiller Repeat Sales index:
  - Case and Shiller (1987, 1989 AER): GLS estimation is performed taking account of heteroscedasticity.

- Hybrid Repeat Sales index:
  - Case and Quigley (1991 RES): Hybrid model consider age problems

Matching Index

- Matching model
  - McMillen(2012) and Deng, McMillen and Sing(2012, 2014) propose a matching estimator with propensity score approach.

  - The main of matching estimator is average treatment effects. For the two periods case:

$$ln\, P_{it} - ln\, P_{it-1} = (\delta_t D_{it} - \delta_{t-1} D_{it-1}) + (\beta' X_{it} - \beta' X_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1})$$

  - Average price of time t: $ln\, \bar{P}_{it} = \delta_t D_{it} + \beta' \bar{X}_{it} + \bar{\varepsilon}_{it}$

  If $\bar{X}_{it}$ is constant, $ln\, \bar{P}_{it}/\bar{P}_{it-1} = \delta_t - \delta_{t-1}$

Matching Index

- Matching model
  - **Average treatment effect (ATE):**

$$ATE_{tj} = \frac{1}{n_j}\sum_{i=1}^{n_j} D_{ij}E(lnP_{it_j} - lnP_{it_1})$$

  - Matching estimators attempt to reduce the effects of non-random sample. If sales were randomly distributed, the average sales prices is all requirement to construct the price index.

  - Matching adjustment enlarge the sample size and solve the age problem of repeat sales.

Matching Index

- Matching model

  - Step 1: Select a base period time, for example 2000Q1. Estimate logit models for each time t and 2000Q1. Defining dependent variable $I_t = 1$ if a sale occurs at time t and $I_t = 0$ if the sale is from 2000Q1. The explanatory variables for logit regressions are same as hedonic estimates.

  - Step 2: Use the estimated propensity score from each logit regressions to match n_1 observations from each t to sales from 2000Q1, where n_1 is the number of sales in 2000Q1.

Decomposition of Distributions Indexes

- Decomposition of differences
  - Oaxaca(1973) and Blinder(1973) propose a decomposition approach based on OLS

$$E(y_1 - y_0) = \underbrace{(z_1 - z_0)\beta_1}_{\text{Attributes Change}} + \underbrace{z_0(\beta_1 - \beta_0)}_{\text{Coefficient Change}}$$

  - Machado and Mata(2005) propose a new decomposition approach based on Quantile Regression(QR)
    - Allows the variability of the covariates
    - Used in house market, like McMilllen(2008), Nicodemo and Raya(2012), Fesselmeyer et al.(2013)

Decomposition of Distributions Indexes

- Decomposition of Distributions Indexes
  - Using Machado and Mata(2005) approach based on Quantile Regression(QR), we decompose a new RPPI with controlled attributes change.

$$E(\widehat{y_1} - \widehat{y_0}) = \underbrace{z_0(\widehat{\beta_1} - \widehat{\beta_0})}_{\textbf{Coefficient Change}}$$

  - Quantile regression (QR) approach

$$Q_i^{\theta}(p \mid z) = z\beta_i(\theta) \quad : \theta \in (0,1)$$

$Q_i^{\theta}(p \mid z)$: θ-*th* quantile of $F_i(p \mid z)$.

$\beta_i(\theta)$: the QR coefficient.

$\mathbf{z}$: housing attributes.

Decomposition of Distributions Indexes

- Decomposition of Distributions Indexes
  - Different from Machado and Mata (M-M 2005) and McMillen (2008) comparing two periods, we consider a series of time period 2000-2015 and construct a quality controlled RPPI.
  - Firstly, we compare each year t with the base year 2000 and get distribution of total change, coefficient change and attributes change

- Quarterly Index with Time Window
  - Secondly, besides M-M approach use t_0 and t_1 two years time period comparison, we construct a quarterly index using decomposition, the window we choose is 4 quarters as same as M-M yearly comparison.

Decomposition of Distributions Indexes

- Resampling Method by Machado and Mata (M-M 2005)
  - Step 1. Estimate QR for denoted set of $\theta \in (0,1)$. The estimates are $\hat{\beta}_0(\theta)$ and $\hat{\beta}_1(\theta)$, i.e. $\hat{\beta}_t(\theta)$ t=0,1
  - Step 2. For QR coefficients set of $\hat{\beta}_t(\theta)$, yield m estimates from QR coefficients
  - Step 3. Generate a random sample of size m with replacement from z_0 and z_1
  - Step 4. Multiple set of $\hat{\beta}_t(\theta)$ with z_0 and z_1. We get estimated samples of house prices with size m. $z_0\hat{\beta}_0(\theta)$, $z_1\hat{\beta}_1(\theta)$ and $z_0\hat{\beta}_1(\theta)$.

Decomposition of Distributions Indexes

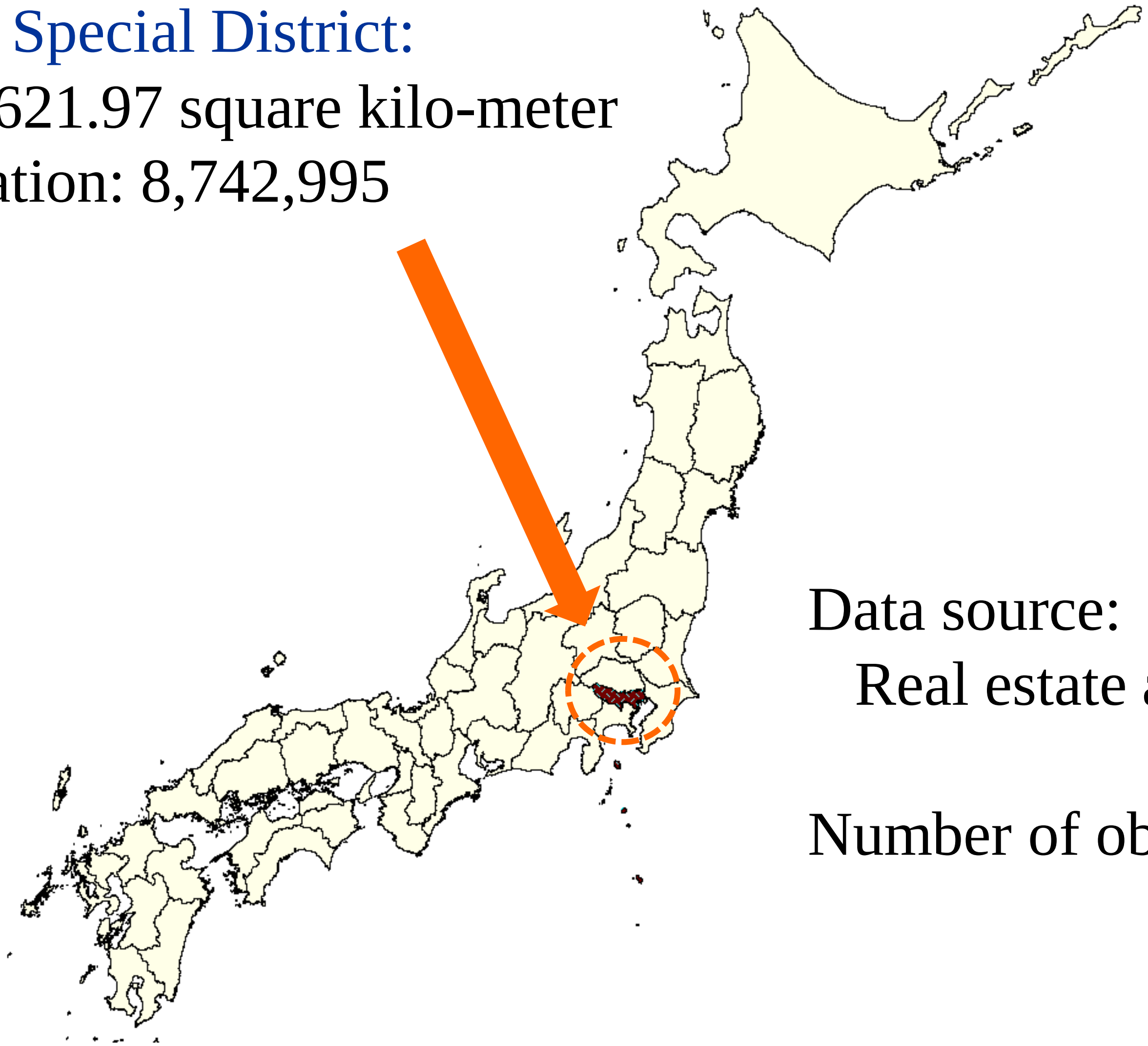
- Decomposition of Distributions Indexes
  - We set 2000 year as t_0. For each quarter q, we set $t_1 = [q-4, q]$
  - After following M-M approach, we have:
    1. $z_t\hat{\beta}_t(\theta)$: Non-quality controlled sample with size m
    2. $z_0\hat{\beta}_t(\theta)$: Quality controlled sample with size m

- Difference analysis (two periods):
  - Total Change(a): $z_1\hat{\beta}_1(\theta) - z_0\hat{\beta}_0(\theta)$
  - Coefficient Change(b): $z_0\hat{\beta}_1(\theta) - z_0\hat{\beta}_0(\theta)$
  - Attributes Change(a)-(b): $(z_1 - z_0)\hat{\beta}_1(\theta)$

3. Data Description

Data source

Tokyo Special District:
Area: 621.97 square kilo-meter
Population: 8,742,995



Data source:
Real estate advertisement magazine
(2000-2015: 15 years)
Number of observations:
87,872 samples

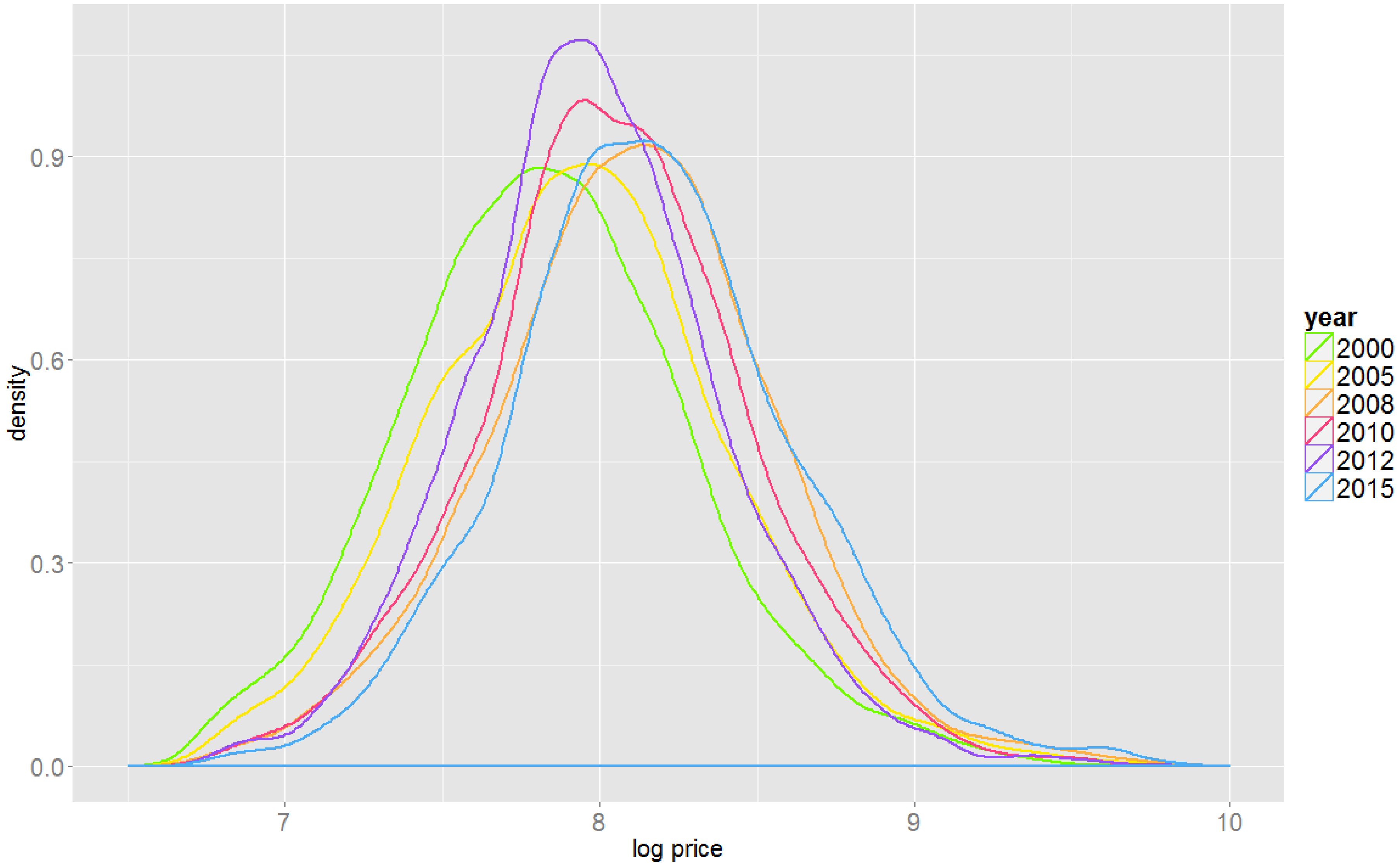
Summary statistics

Table 1. Summary Statistics

	(1) Full Sample		(2) Repeat Sales Sample		(3) Matched Sample	
	mean	sd	mean	sd	mean	sd
Transaction Price (10,000 Yen)	3304.166	1696.319	3366.593	1692.357	3270.699	1666.887
Area of Structure (m^2)	61.293	18.073	62.229	17.822	60.654	17.984
Age (month)	203.508	116.907	238.309	111.137	194.423	114.200
Time to the Nearest Station (minutes)	7.389	4.228	7.678	4.294	7.335	4.205
Time to Tokyo Central Station from the nearest station (minutes)	25.943	8.479	26.417	8.466	26.289	8.506
N	87,872		6,920		66,981	

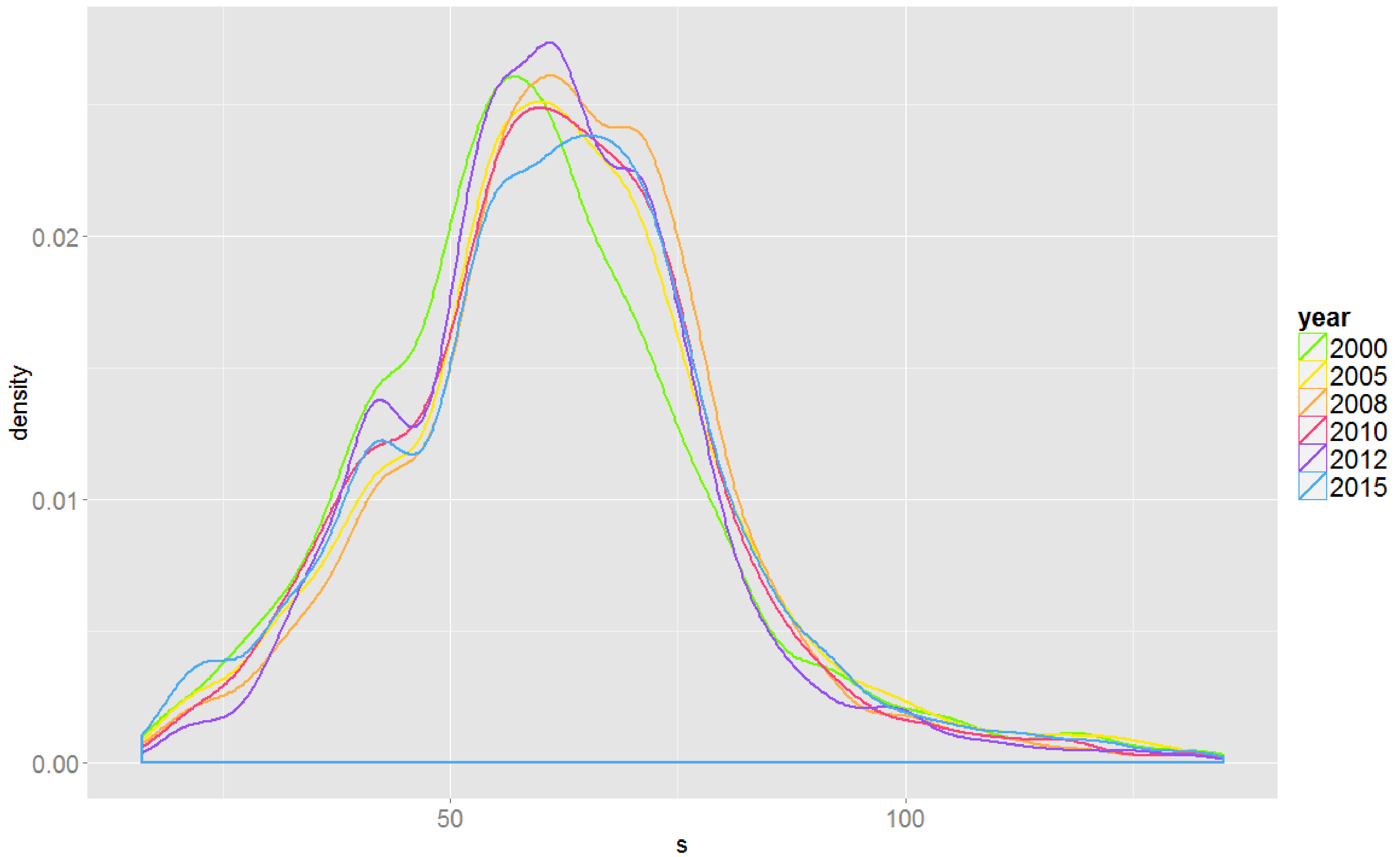
Price distribution

Figure 1. Price distribution: Full Sample



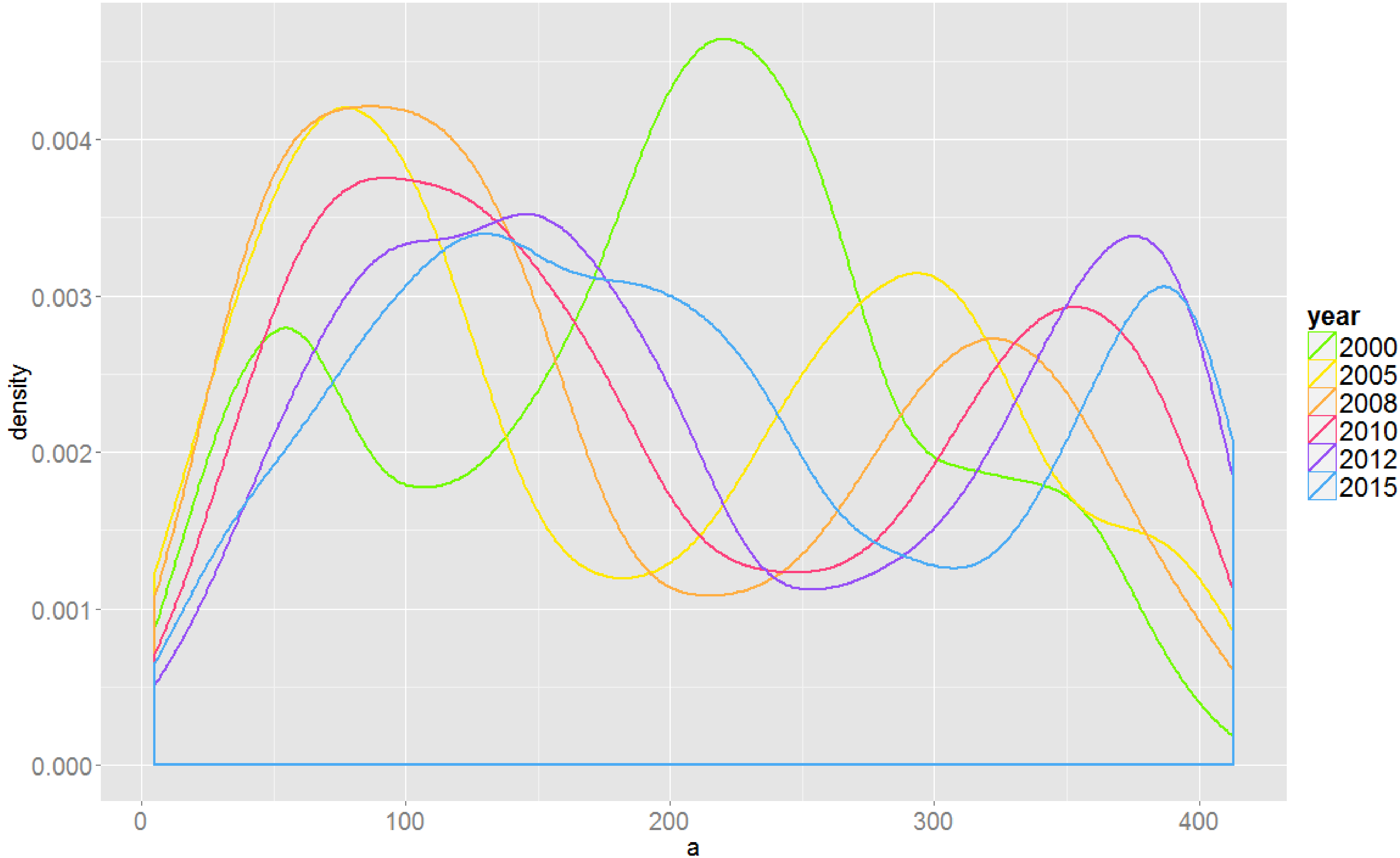
Attributes distribution: floor space

Figure 2a. Attributes distribution: Floor Space



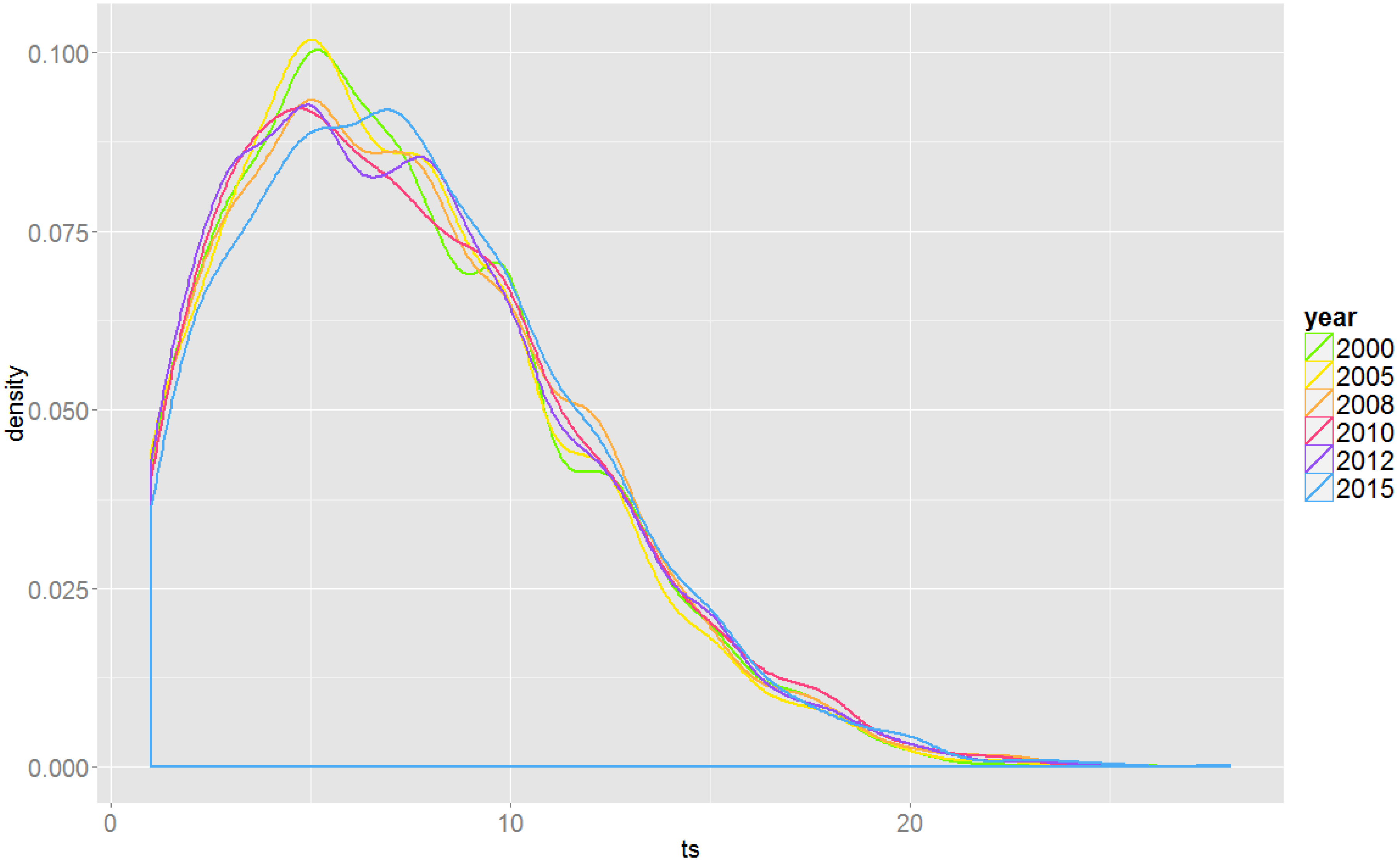
Attributes distribution: age of building

Figure 2b. Attributes distribution: Age



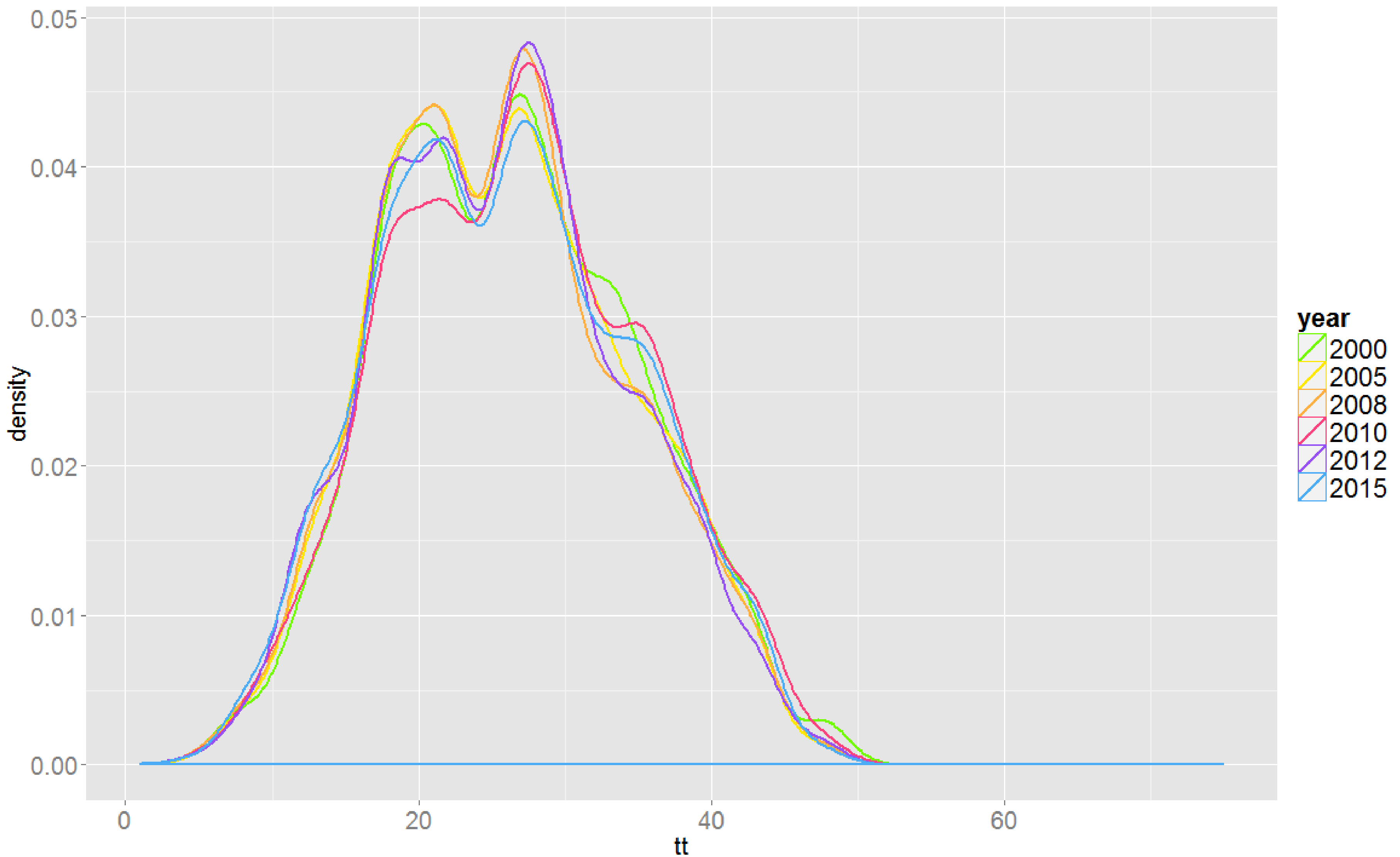
Attributes distribution: time to nearest station

Figure 2c. Attributes distribution: time to nearest station



Attributes distribution: time to Tokyo station

Figure 2d. Attributes distribution: time to Tokyo station



4. Empirical Results

Hedonic Regressions

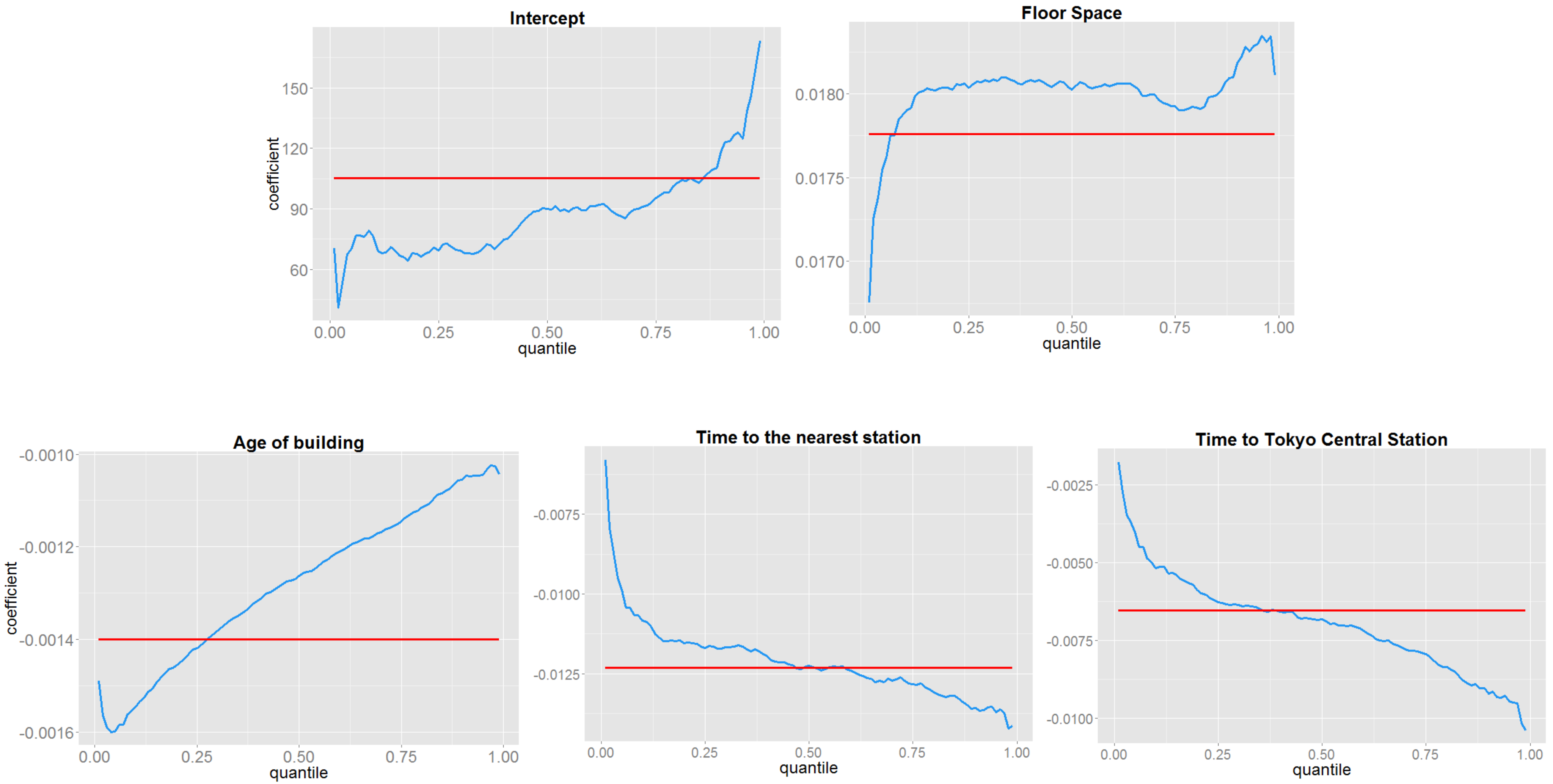
Table 2. Results of hedonic regressions

	(1) Full Sample Hedonic		(2) Matched Sample Hedonic	
	Coefficient	t-value	Coefficient	t-value
Floor Area (m^2)	0.0178***	(358.21)	0.0182***	(325.38)
Building Age (month)	-0.00140***	(-255.99)	-0.00142***	(-223.01)
Time to the Nearest Station (minutes)	-0.0123***	(-83.61)	-0.0125***	(-74.51)
Time to Tokyo Central Station from the Nearest Station (minutes)	-0.00653***	(-51.01)	-0.00643***	(-43.71)
Structure: SRC (dummy)	0.0118***	(8.93)	0.0109***	(7.30)
Facing South (dummy)	0.00430***	(3.52)	0.00569***	(4.16)
Longitude (x)	-0.399***	(-12.18)	-0.395***	(-10.67)
Latitude (y)	-1.180***	(-29.80)	-1.025***	(-22.86)
Time Dummy	Yes		Yes	
District Dummy	Yes		Yes	
_cons	105.4***	(21.99)	99.30***	(18.25)
N	87872		66981	
R^2	0.861		0.866	

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

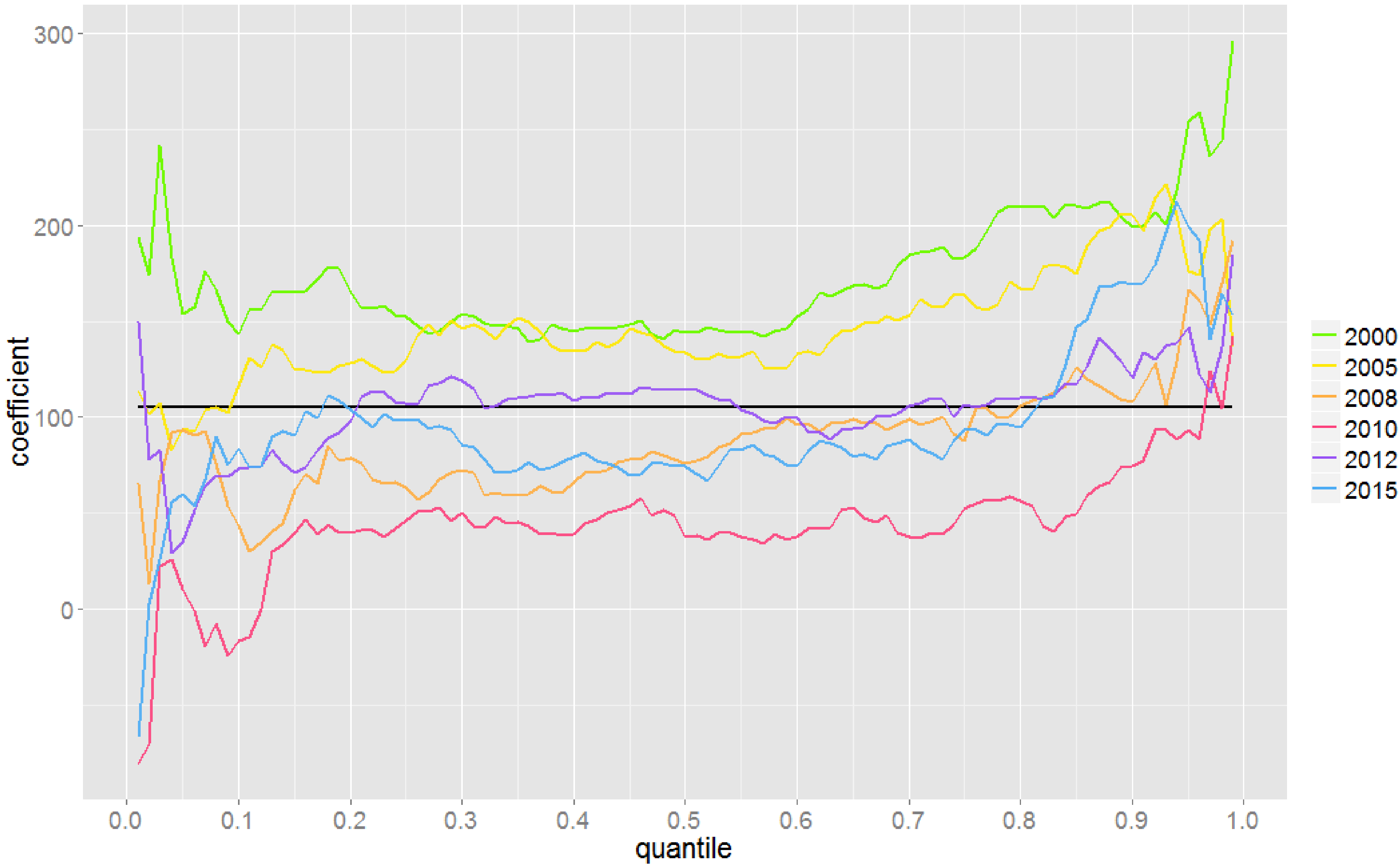
Quantile Regressions Approach

Figure 3. Quantile regressions



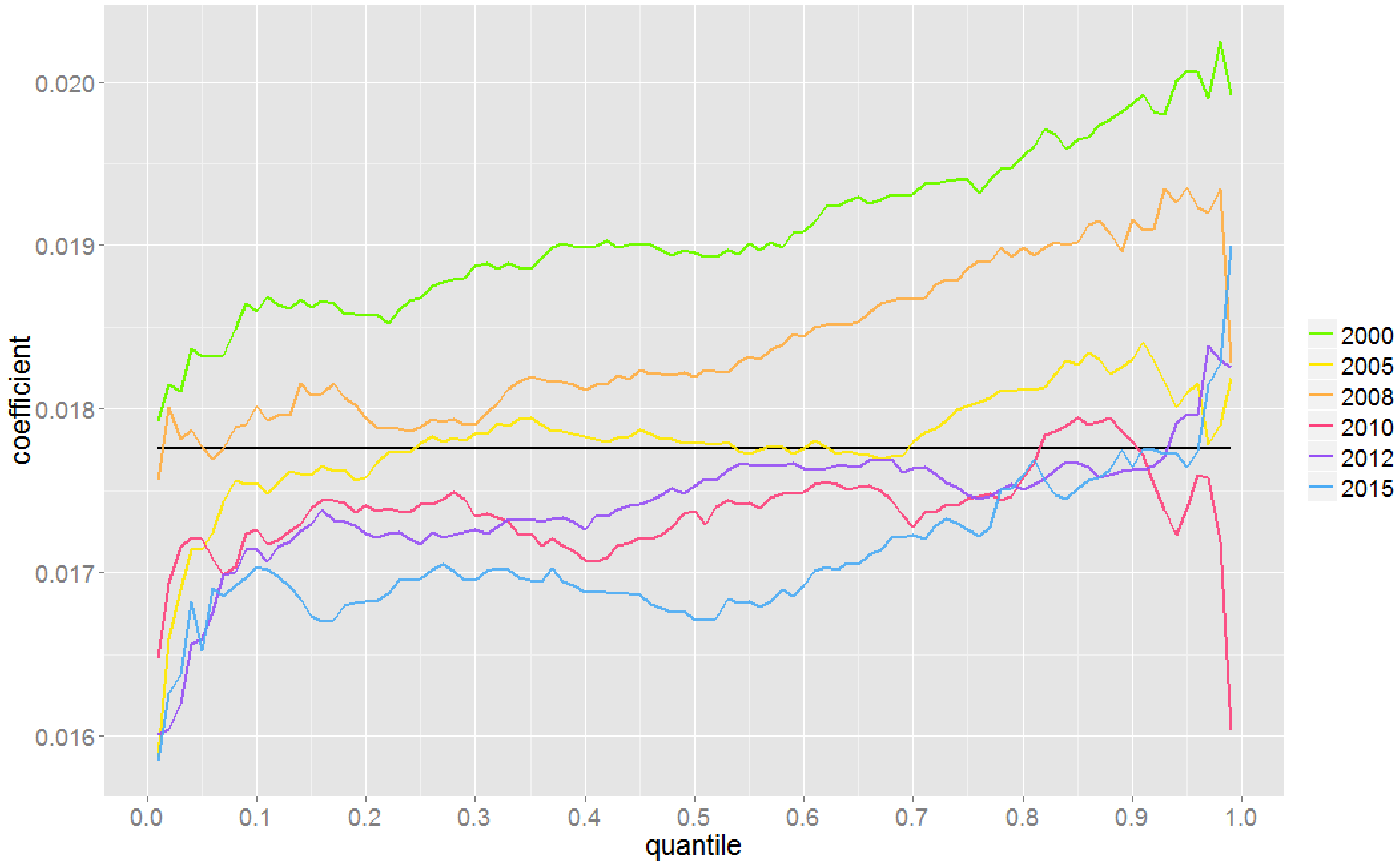
Quantile Regressions Approach: Intercept

Figure 4a. Quantile regressions for structure change: intercept



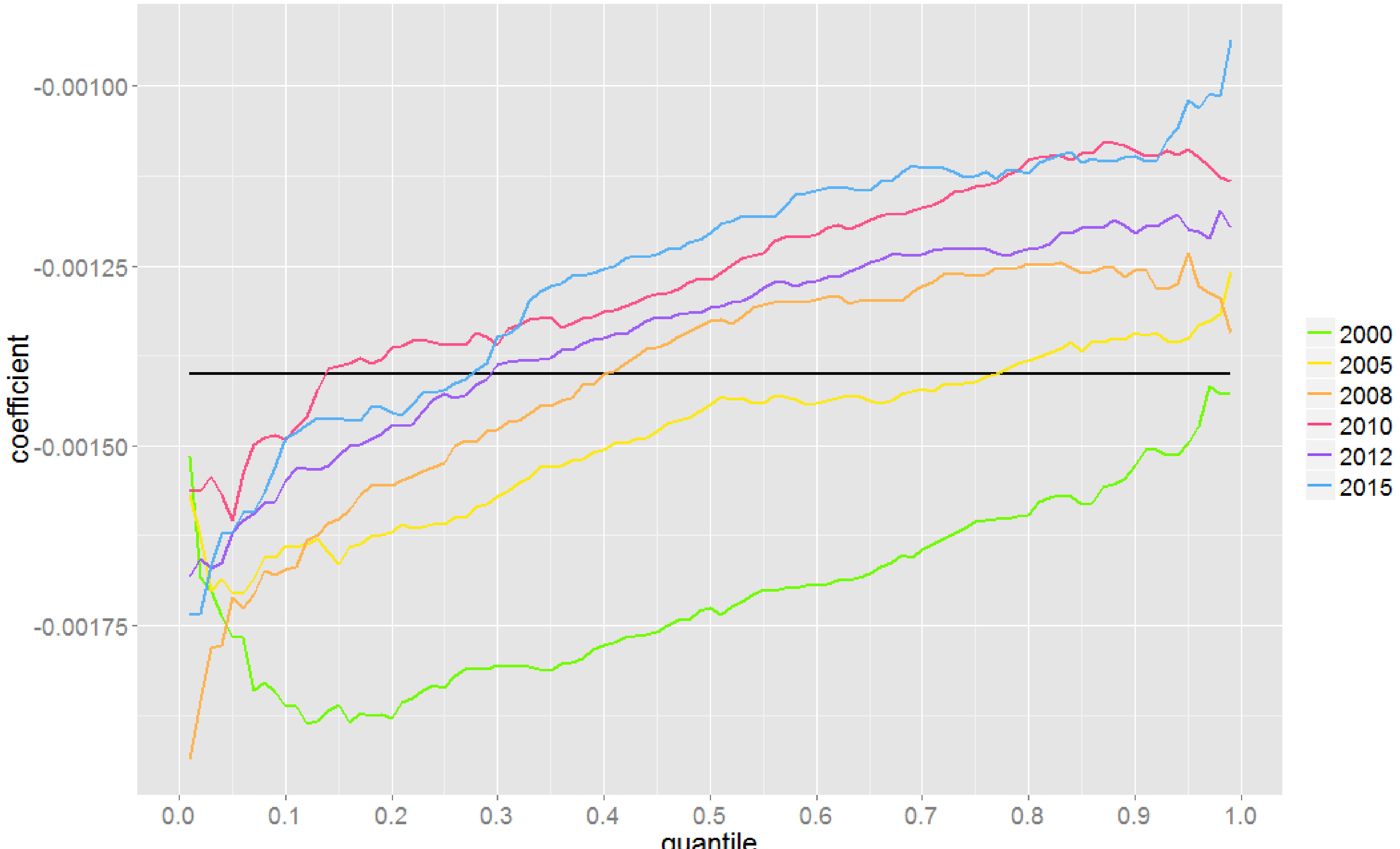
Quantile Regressions Approach: Floor Space

Figure 4b. Quantile regressions for structure change: floor space



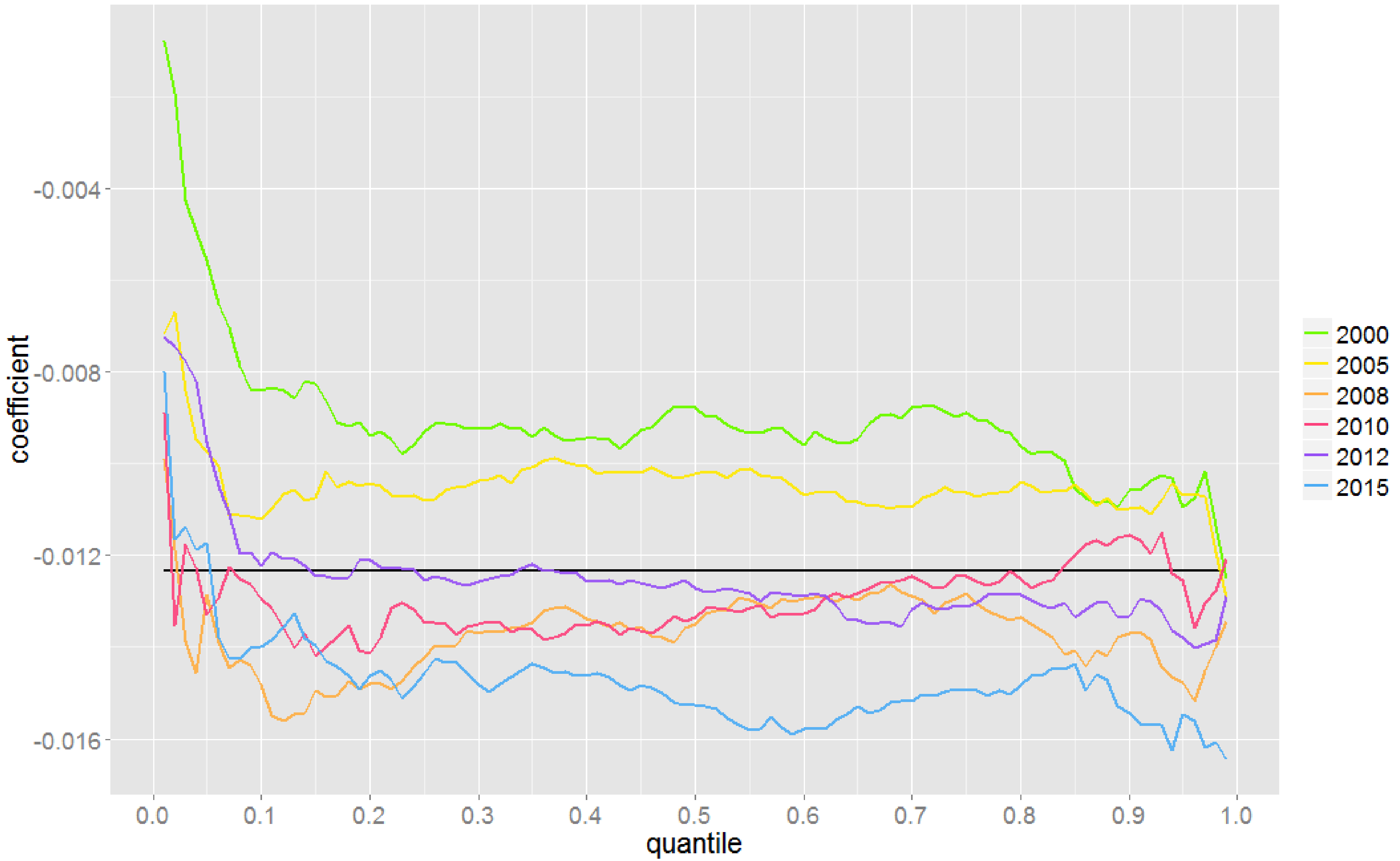
Quantile Regressions Approach: Age of Building

Figure 4c. Quantile regressions for structure change: age



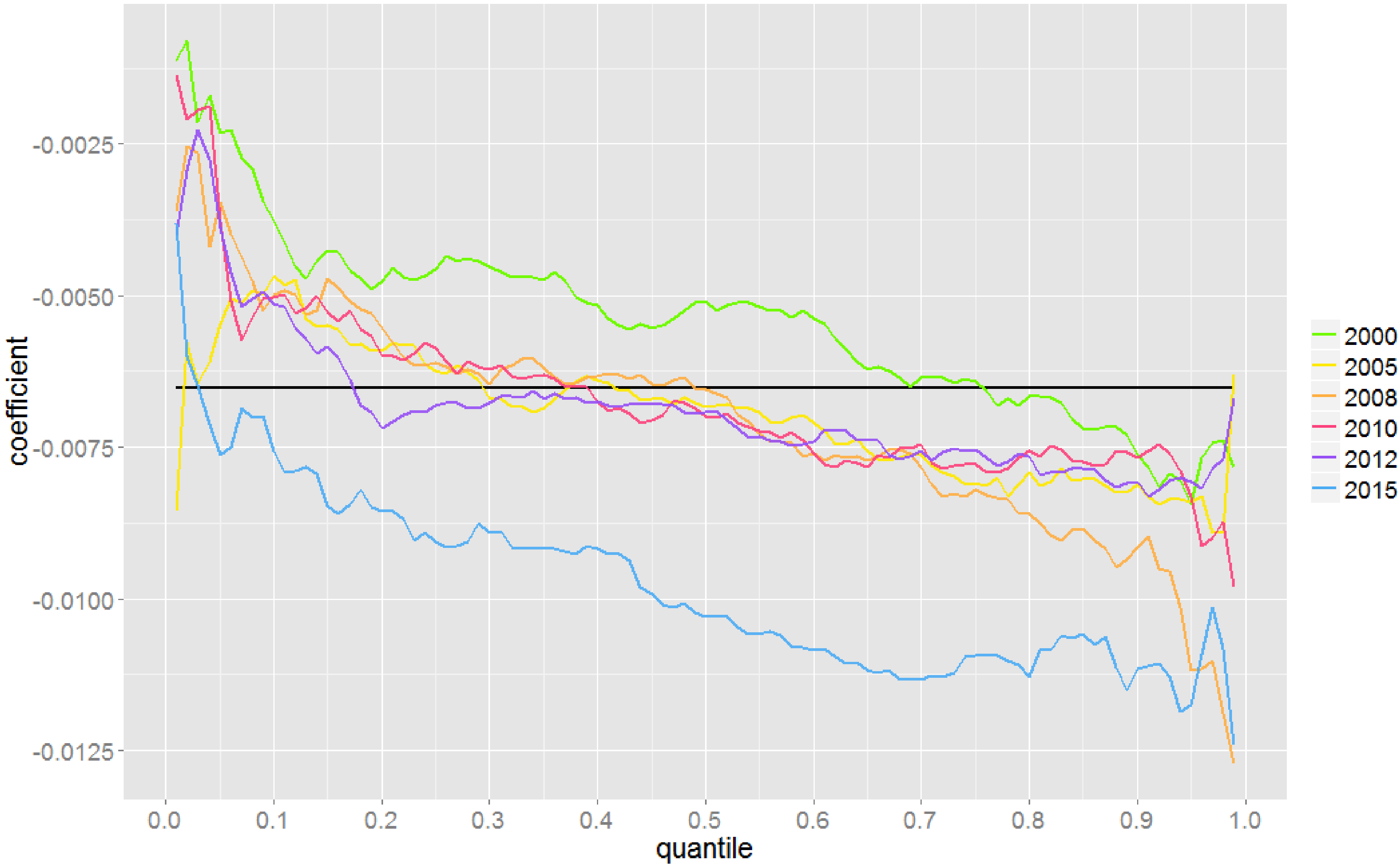
Quantile Regressions Approach: Time to Station

Figure 4d. Quantile regressions for structure change: time to station



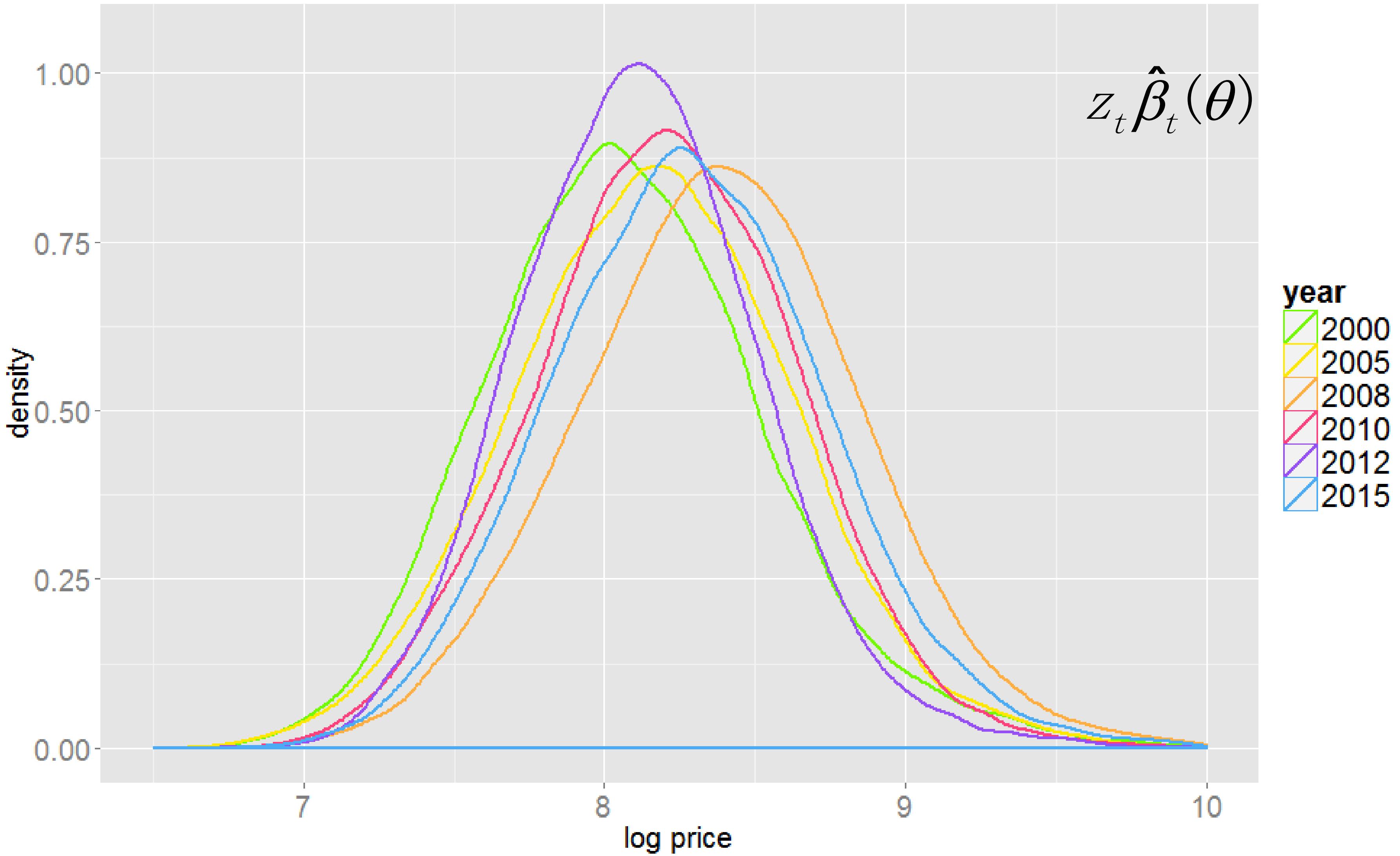
Quantile Regressions Approach: Time to Tokyo Station

Figure 4e. Quantile regressions for structure change: time to Tokyo station



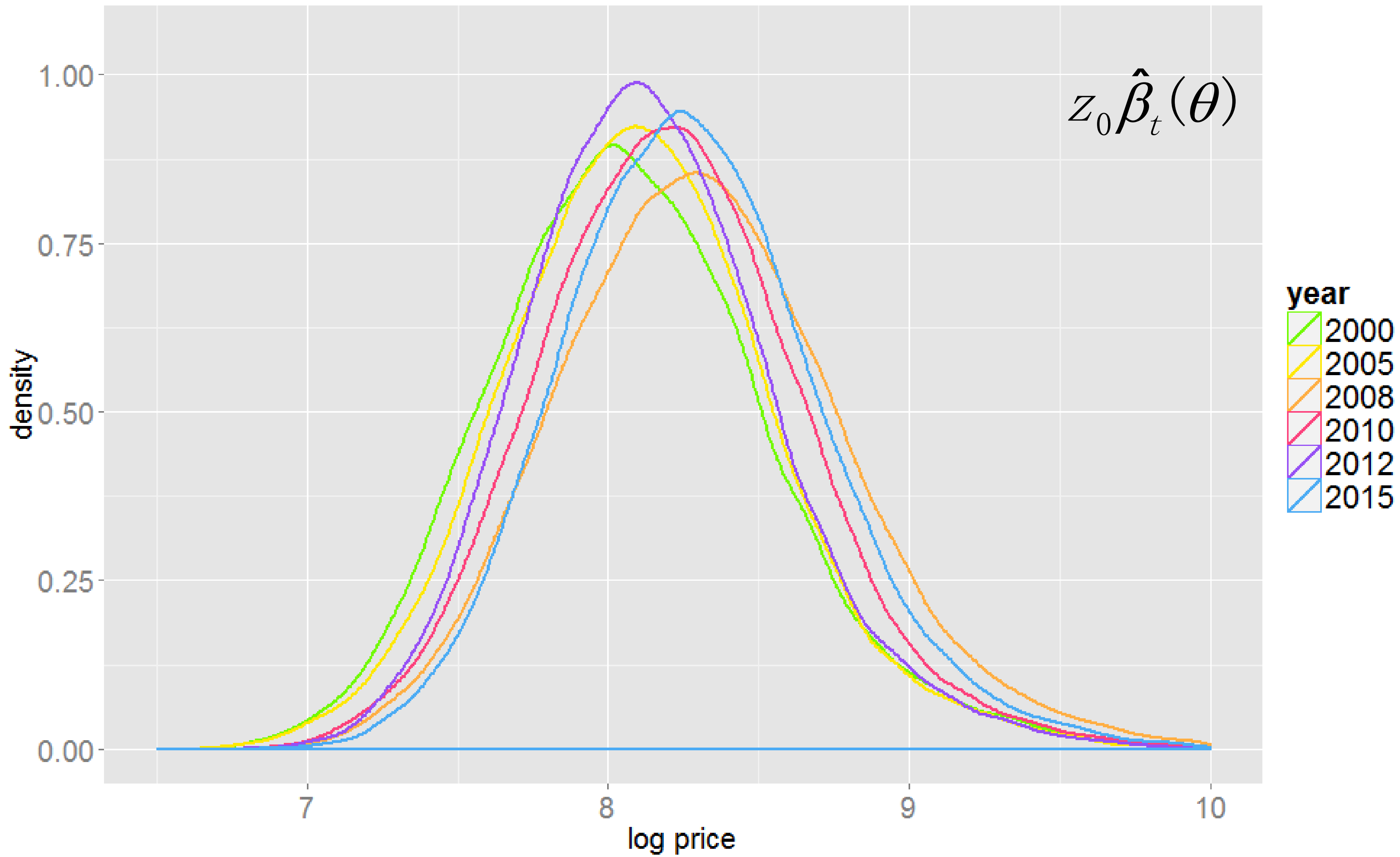
Decomposition: Total Distribution

Figure 5a. Decoimposition distribution: Total Distribution



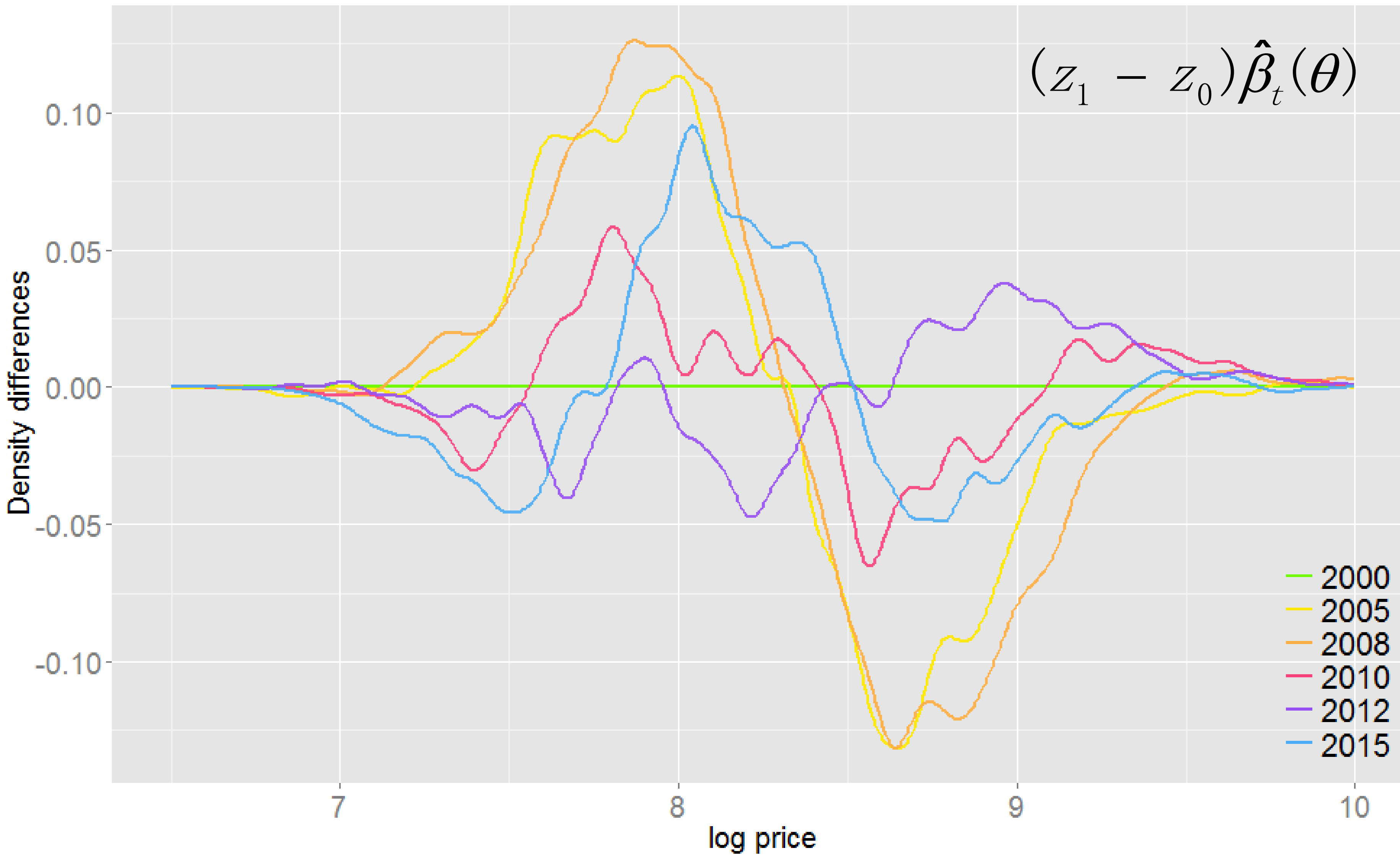
Decomposition : Coefficient Distribution

Figure 5b. Decomposition distribution: Coefficient Distribution



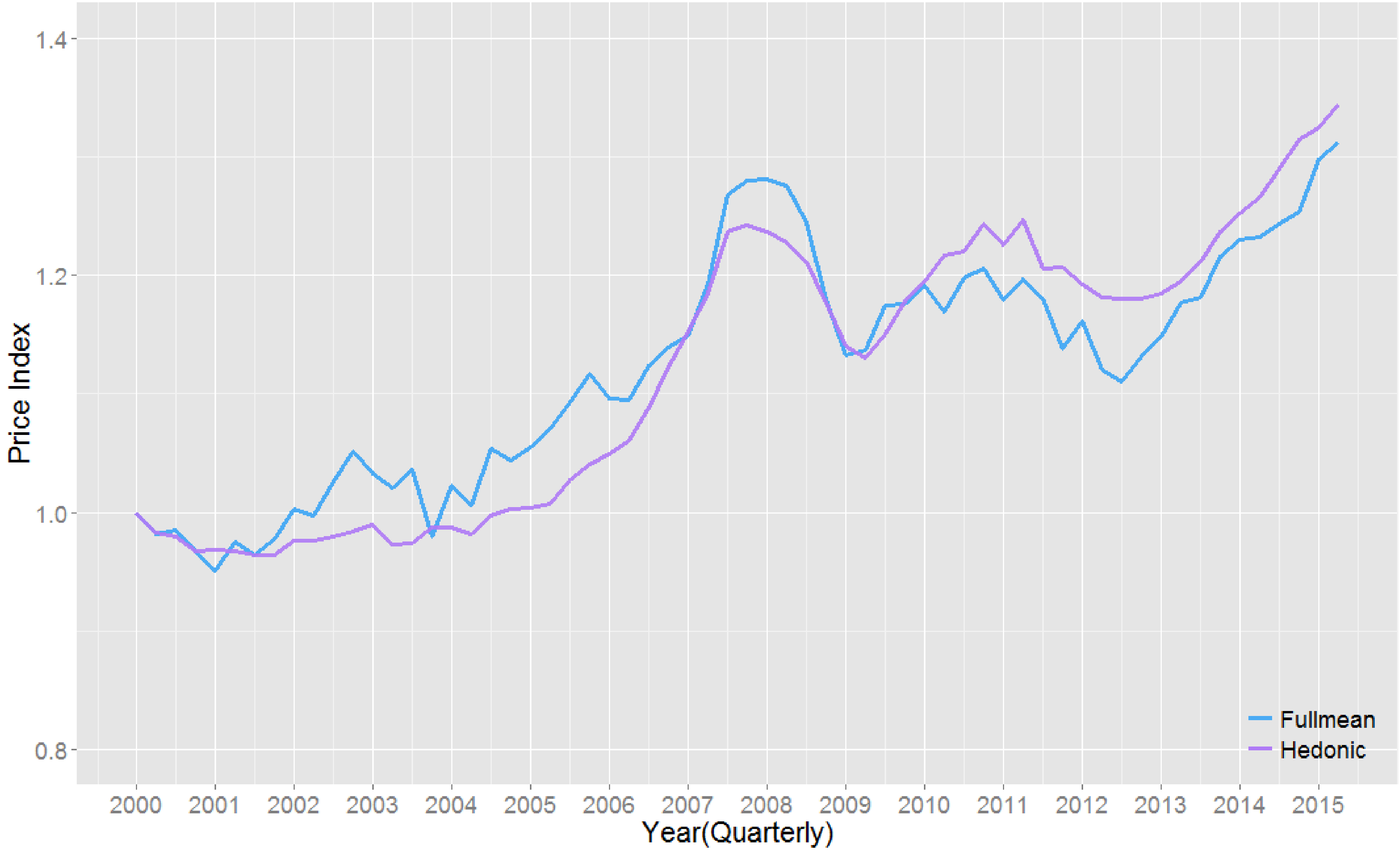
Decomposition : Attributes Change

Figure 5c. Decomposition change: Attributes Change



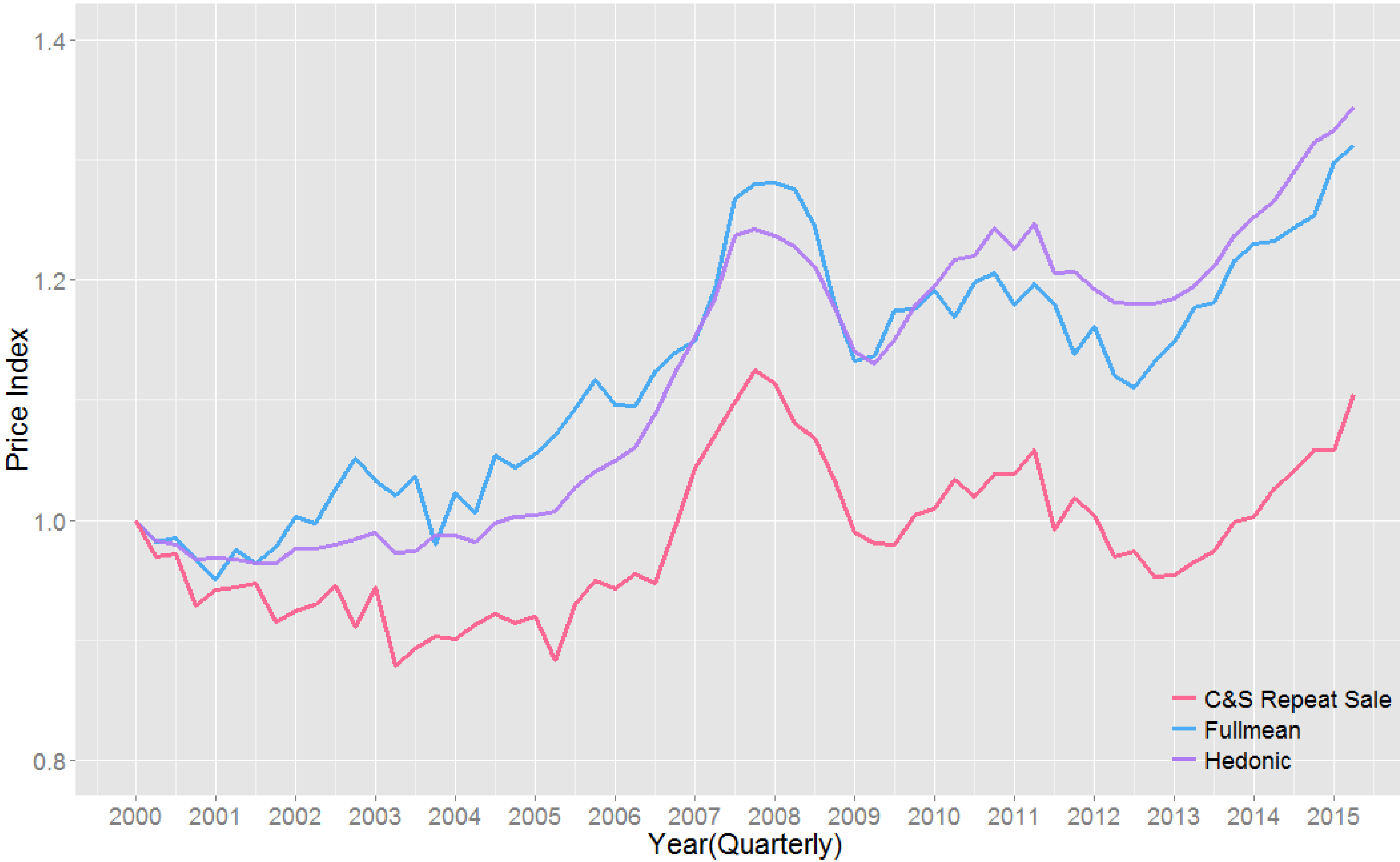
Indexes Comparison: Hedonic

Figure 6a. Compare Indexes: Hedonic Index



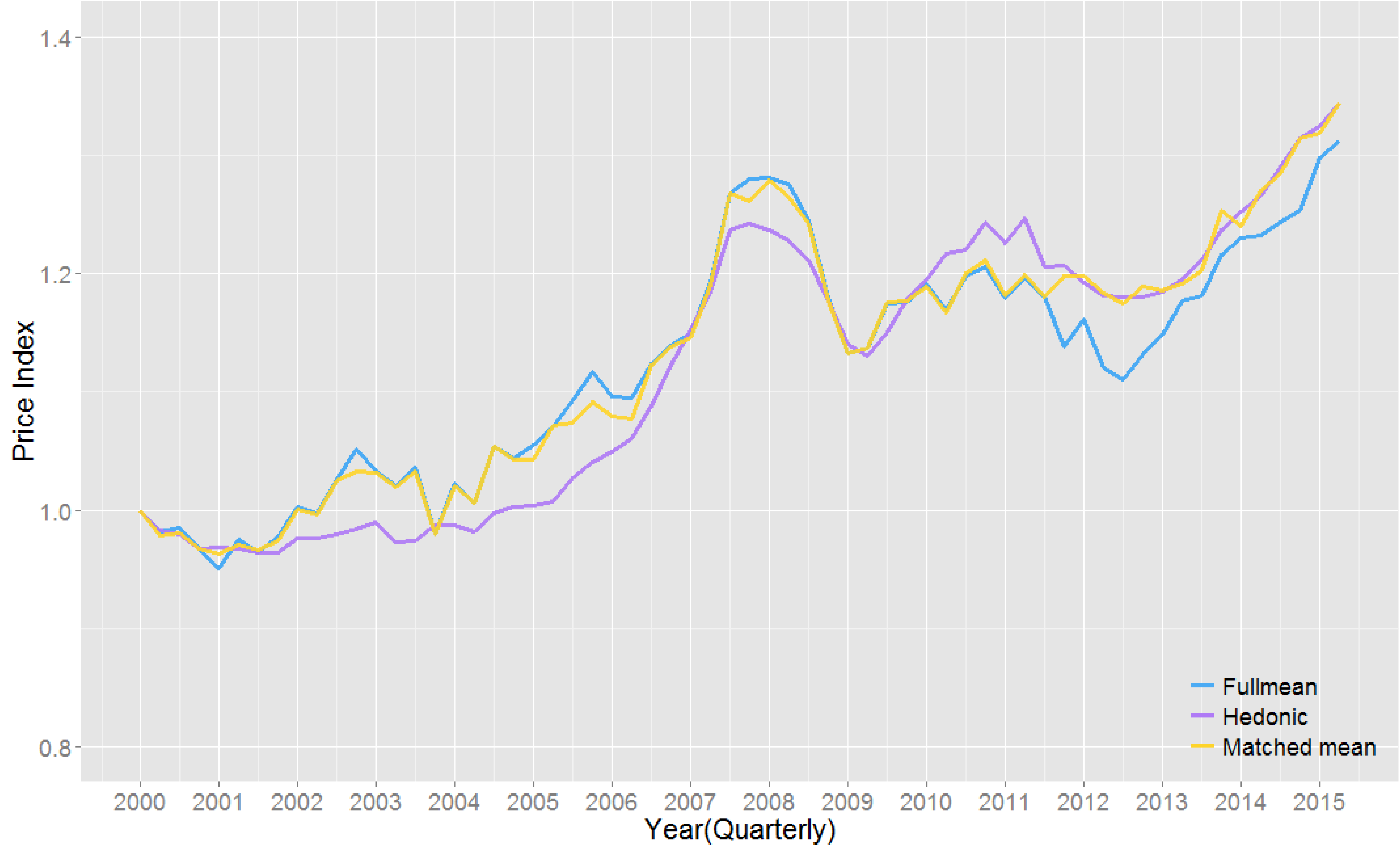
Indexes Comparison: Repeat Sale

Figure 6b. Compare Indexes: Case-Shiller Repeat Sale



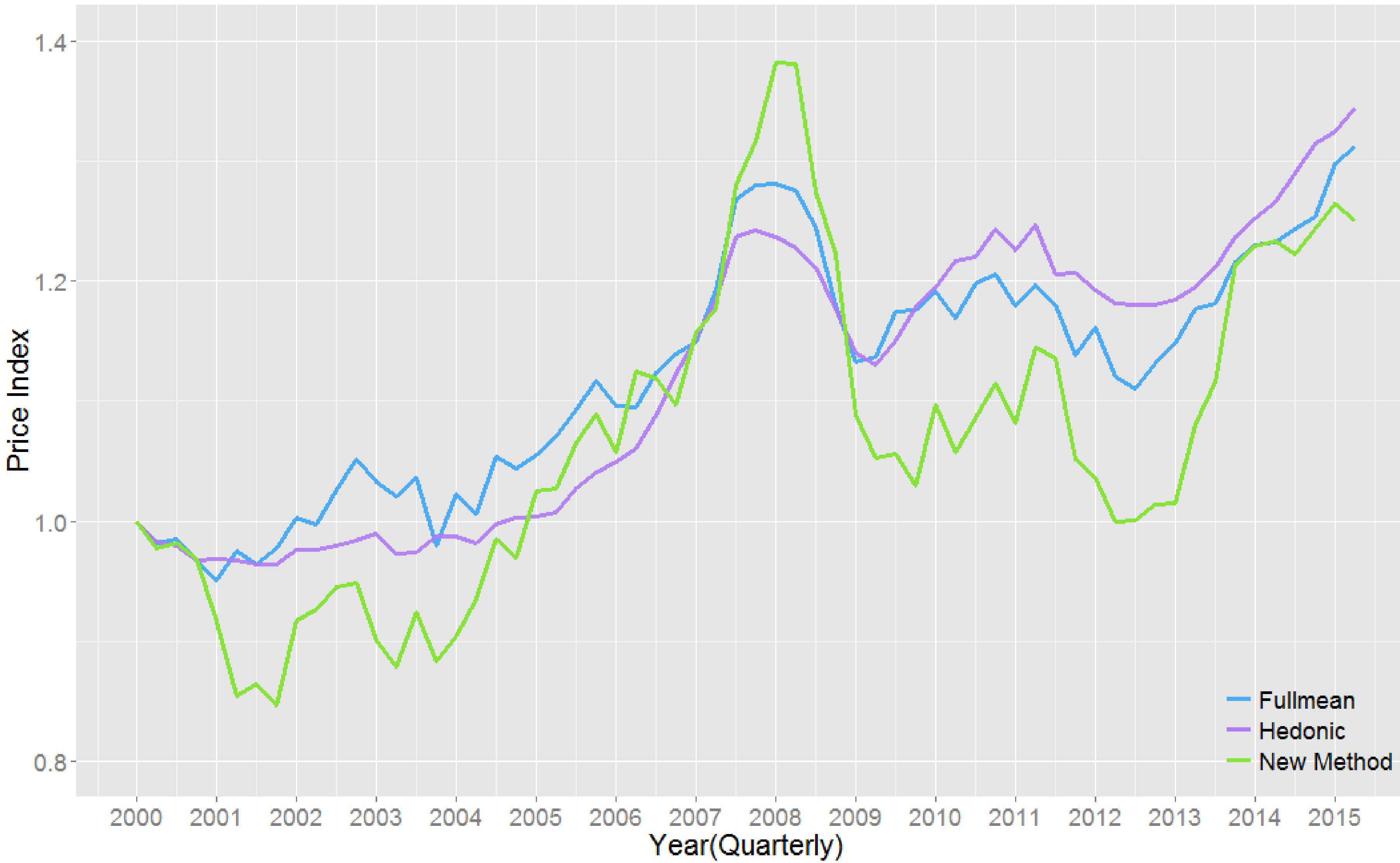
Indexes Comparison: Matched

Figure 6c. Compare Indexes: Matched Mean



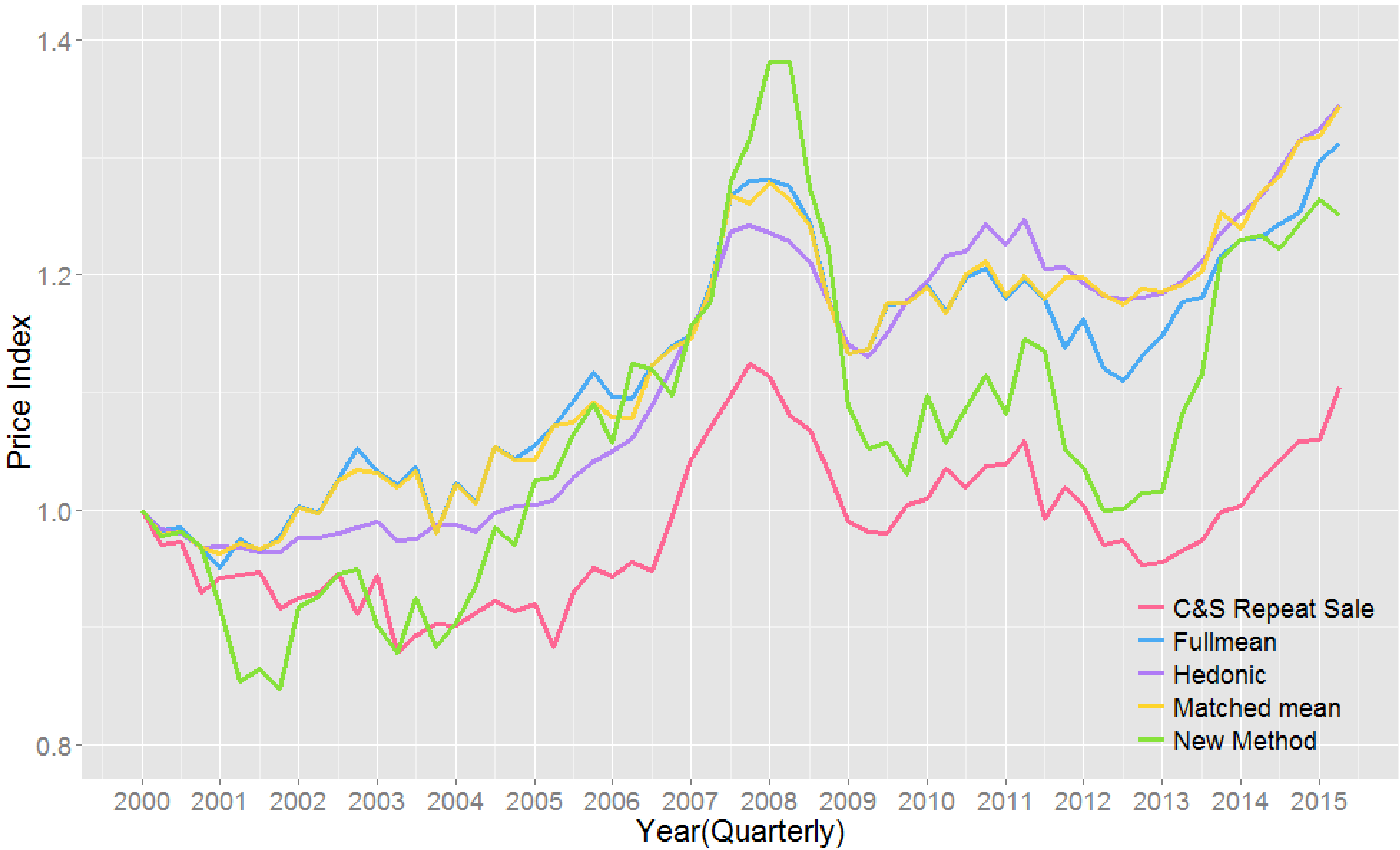
Indexes Comparison: Decomposition

Figure 6d. Compare Indexes: Decomposition



Indexes Comparison: All

Figure 6e. Compare Indexes: All



5. Conclusion

Conclusion

- Conclusion
  - It is hard to construct an unbiased residential property price index, even if considering various adjustments.
  - Hedonic model has the limitations of structure change and omitted variable problems. Decomposition approach identify coefficient change with controlled structure change.
  - Repeat sales index has the limitations of non-random sampling and age problem. Matched sample enlarge the sample size and avoid non-random sampling.
- Further work
  - We plan to use decomposition approach constructing a globe property price index. That will be easily compare attributes and coefficient change across countries.