House Price Dispersion in Boom-Bust Cycles: Evidence from Tokyo

Takaaki Ohnishi
Graduate School of Information Science and Technology, Univ of Tokyo

Takayuki Mizuno
National Institute of Informatics

Tsutomu Watanabe
Graduate School of Economics, Univ of Tokyo

June 8, 2019
# Heatmap (Financial Activity Indexes)

**Financial System Report, BOJ, April 2019**

**Chart III-4-1: Heat map**

| CY | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Financial institutions | DI of lending attitudes of financial institutions | Growth rate of M2 |
| Financial markets | Equity weighting in institutional investors’ portfolios | Stock purchases on margin to sales on margin ratio |
| Private sector | Private investment to GDP ratio | Total credit to GDP ratio |
| Household | Household investment to disposable income ratio | Household loans to GDP ratio |
| Corporate | Business fixed investment to GDP ratio | Corporate credit to GDP ratio |
| Real estate | Real estate firms’ investment to GDP ratio | Real estate loans to GDP ratio |
| Asset prices | Stock prices | Land prices to GDP ratio |

Note: The latest data for the DI of lending attitudes of financial institutions and stock prices are as at the January-March quarter of 2019. The latest data for the land prices to GDP ratio are as at the July-September quarter of 2018. The latest data for the other indicators are as at the October-December quarter of 2018.

Application of anomaly detection techniques to real estate markets

1. Anomalies in real estate prices
   - Especially anomalies in the cross-sectional distribution of real estate prices
2. Anomalies in real estate transactions
3. Anomalies in real estate credit
Anomaly Detection

- Supervised vs. Unsupervised
- What is “Normal”? 

Price

Attribute

Anomalies

Normal
Heterogeneity in house prices

- A new look at house price dynamics
  - **Not time series but cross-section:** Case and Mayer (1996); Guerrieri et al (2013)
  - **Not mean but higher moments:** McMillen (2008); Sinai (2012); Van Nieuwerburgh and Weill (2010); Gyourko et al (2013); Deng et al (2012); Villar (2015); Zhang (2018); Andersson et al (2003); Blackwell (2018)
  - **Not across cities but within-city:** Case and Mayer (1996); Ferreia and Gyourko (2012, 2017); Glaeser et al (2012); Lyons (2015); Waltl (2016); Zhu (2018); Bogin et al (2018)
McMillen (J of Urban Economics, 2008)

**Fig. 1.** Kernel density estimates for log of real sales price.

**Fig. 2.** Estimated cumulative density function for log of real sales price.
Heterogeneity in house prices

- Heterogeneity in house prices in boom-bust cycles
  - The recent US housing boom cannot be interpreted as a single, national event, as different markets began to boom across a decade-long period from the mid-1990s to the mid-2000s (Ferreira and Gyourko 2012, 2017).

Figure 3: Individual Metropolitan Area Hedonic House Price Indexes by Quarter

San Francisco-Oakland-Fremont, CA

Ferreira and Gyourko (2017)
Assignment models of housing markets

• Previous papers in macroeconomics typically study an economy with a homogeneous housing capital good, where households choose different quantities of that good at a common per-unit price. Those studies do not tell anything about the cross-section of house prices.

• Assignment models of housing markets (Landvoigt et al 2015, Piazzesi and Schneider 2016, Määttänen and Terviö 2014, Rios-Rull and Sanchez-Marcos 2008)

  – A house is a bundle of units. Different houses consist of different sets of units (Rosen 1974). Each household buys one house and obtains a flow of housing services from it.

  – Equilibrium prices are determined so that households with high (low) demand for housing services are assigned to high (low) quality houses. The distribution of equilibrium prices across houses is determined by the distribution of qualities across houses and the distribution of characteristics across households.

What we do in this paper

- We propose a methodology to detect anomalies in real estate transactions
- We apply this methodology to real estate transaction data from Greater Tokyo Area
- We show that the cross-sectional distribution of house prices was close to a log-normal in most of the sample period, but deviated from it during the bubble years
- We ask why the cross-sectional distribution deviated from a log-normal during the bubble years
Outline

• Methodology
  – Hedonic model
  – CLT
• Data
• Size-adjustment to house prices
• Submarket hypothesis
• Conclusion
Anomaly Detection

- Supervised vs. Unsupervised
- What is “Normal”?
A famous textbook example of the central limit theorem is the distribution of persons’ height. The height distribution of, say, mature men of a certain age can be considered normal, because height can be seen as the sum of many small and independent effects. Similarly, the log house prices will be normally distributed if house prices are determined as the sum of many small and independent effects.
Housing Attributes

- We empirically show that the heterogeneity in house prices stems mainly from the size and the location. The variances of these two factors are very large relative to the variances of other factors, so that the Lindeberg’s condition is violated.

- **The size of a house**
  - We construct size-adjusted prices, thereby eliminating the effect of house size on prices.

- **The location of a house**
  - We restrict price comparison to the neighborhood area, thereby eliminating the effect of location on prices.

\[
\ln P_i = \sum_{k=1}^{K} x_{ik}
\]
Outline

• Methodology
  – Hedonic model
  – CLT

• Data
• Size-adjustment to house prices
• Submarket hypothesis
• Conclusion
Greater Tokyo Area
Weekly data from 1986 to 2009 including the period of housing bubble in the late 1980s and its collapse in the first half of the 1990s
Condominiums: 724,416 listings
Single family houses: 1,602,918 listings
PDF and CDF of original prices in 2008

Prob Density Function

Counter Cumulative Distribution Fn

PDF

CDF

House price [Ten thousand yen]

House price [Ten thousand yen]
CDFs of house prices in 1986-2008
Outline

• Methodology
  – Hedonic model
  – CLT
• Data
• **Size-adjustment to house prices**
• Submarket hypothesis
• Conclusion
Joint distribution of log price and house size

House size [square meters]

House price [10 thousand yen]
Size-adjustment to house prices

We empirically show:

\[ P_{it} \sim PL(\zeta_t) \]  \hspace{2cm} (1)

\[ S_i \sim Exp(\lambda_t) \]  \hspace{2cm} (2)

This implies:

\[ \ln P_{it} = \left( \frac{\lambda_t}{\zeta_t} \right) S_i + \epsilon_{it} \]  \hspace{2cm} (3)

where \( \epsilon_{it} \) is a Gaussian disturbance term.

Size-adjusted prices

\[ \tilde{P}_{it} \equiv \frac{P_{it}}{\exp \left[ (\lambda_t/\zeta_t) S_i \right]} \sim LN \]
CDF of size-adjusted prices in 2008

Normalized price

Size-adjusted price in 2008
Original price in 2008
Lognormal
Outline

• Methodology
  – Hedonic model
  – CLT
• Data
• Size-adjustment to house prices
• Submarket hypothesis
• Conclusion
Lognormal Distribution as Benchmark

Lindeberg-Feller Central Limit Theorem

Consider a hedonic model:

\[ \ln P_i = \sum_{k=1}^{K} x_{ik} \]

The price follows a lognormal distribution if \( x_{ik} \) are independent from each other, and the following condition is satisfied.

\[ \lim_{K \to \infty} \frac{\max_{k \leq K} \{ s_k \}}{K \bar{s}_K} = 0 \]

Lindeberg's condition

where \( s_k^2 \) is the variance of \( x_{ik} \) and \( \bar{s}_K \equiv \frac{1}{K} (s_1^2 + s_2^2 + \cdots + s_K^2) \)

A famous textbook example of the central limit theorem is the distribution of persons’ height. The height distribution of, say, mature men of a certain age can be considered normal, because height can be seen as the sum of many small and independent effects. Similarly, the log house prices will be normally distributed if house prices are determined as the sum of many small and independent effects.
Submarket hypothesis

Bubble years

The size-adjusted prices follow a distribution with a heavier tail than a log-normal distribution

Normal years

The size-adjusted prices follow a log-normal distribution
# Submarket hypothesis

### Bubble years

| HHHH | MMMM | MMMM | LLLL |

### Randomly distributed

<table>
<thead>
<tr>
<th></th>
<th>HM</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>ML</td>
<td>ML</td>
</tr>
</tbody>
</table>

Heterogeneous within each submarket but homogeneous across different submarkets

### Clustered

<table>
<thead>
<tr>
<th>HH</th>
<th>MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>MM</td>
</tr>
</tbody>
</table>

Homogeneous within each submarket but heterogeneous across different submarkets
Size-adjusted prices that are normalized at the pixel level

2008

Pixel size:
- 3.3 x 3.3 km
- 26.2 x 26.2 km
- 52.4 x 52.4 km
- 419 x 419 km

CDF

Normalized log(Price) – aS
Size adjusted prices that are normalized at the pixel level

1990

Pixel size:
- 3.3 x 3.3 km
- 26.2 x 26.2 km
- 52.4 x 52.4 km
- 419 x 419 km

Normalized log P – aS

CDF

10^{-5} 10^{-4} 10^{-3} 10^{-2} 10^{-1} 10^{0}
Distributions of size-adjusted prices for different pixel sizes
Deviation from the standard normal distribution: Official Land Price data

Kolmogorov-Smirnov statistic

$D\sqrt{N}$

1% significance
5% significance


$D\sqrt{N}$

1% significance
5% significance


All n=1024 n=512 n=256
Law of One Price during the tech bubble

- Lamont and Thaler (2001); Ofek and Richardson (2001, 2002)
- Cochrane (2002): The “bubble” was concentrated. ... [I]f there was a “bubble,” or some behavioral overenthusiasm for stocks, it was concentrated on Nasdaq stocks, and Nasdaq tech and internet stocks in particular.
Procedure to estimate the size of a submarket within which house prices follow a log-normal distribution

1. For a house $i$, we collect $n-1$ nearest houses.

2. We calculate a size-adjusted price $Q$ for each of the $n$ houses (i.e. house $i$ and the $n-1$ nearest houses) and conduct Geary’s test to see whether the log of $Q$ follows a normal distribution.

3. We start this from a small value for $n$ and repeat this exercise until we find $n_i^*$ such that the null is rejected for values greater than $n_i^*$.

4. We define $\Theta_{it}$ for house $i$ transacted in month $t$ as:

$$\Theta_{it} \equiv -\log \frac{n_i^*}{N_t}$$

where $N_t$ is the number of houses transacted in month $t$.

5. We calculate $\Theta_{it}$ for each house transacted in each month of the sample period.

$n_i^* = 8$
Estimates of $\Theta$ for individual property listings in 1989, 1992, and 1995
Mean of $\Theta_i$ over individual houses at the monthly frequency
Conclusion

• We show that
  
  – House price distribution in Tokyo had a power law tail during the bubble period in the late 1980s, while it was very close to a lognormal after the bubble period.
  
  – House price distribution within a submarket was close to a lognormal even during the bubble period, but the mean and variance of the lognormal distribution differed substantially across submarkets.

• This spatial heterogeneity is the source of the power law tail observed during the bubble period, and can be regarded as evidence for the lack of price arbitrage across submarkets.

• We argue that the shape of the cross-sectional distribution of prices, especially the tail part of the distribution, may contain information useful for the detection of housing bubbles.
References on heterogeneity in house prices

References on heterogeneity in house prices


References on assignment models of housing markets


