Research Methodology 05
-Interpretation and Validation-

Feb 11, 2016

Chihiro Shimizu (清水千弘)
シンガポール国立大学
Professor, Institute of Real Estate Studies
National University of Singapore

0. Outline of Today's Lecture

1. Interpretation and simulation.
2. Machine Learning and Big Data.
3. Validation.

Today's Stat: Multiple Regression

Today's Case:

"Estimation of Redevelopment Probability using Panel Data-Asset Bubble Burst and Office Market in Tokyo-,"

1. Interpretation and Simulation.

- There is no clear-cut dividing line between analysis and interpretation. They very often overlap. ***Interpretation refers to the analysis of generalisations and results.***
- Through interpretation, the meanings and implications of the study become clear. Analysis is not complete without interpretation and interpretation cannot proceed without analysis. Both are thus interdependent. ***Interpretation can be conceived of as a part of analysis.***
- Analysis and interpretation occupy the last stage of the research, conceptually or in terms of thought, they occupy the first stage, since the necessary theoretical and practical knowledge of the future shape of the result is acquired much before the actual work is undertaken.

Case1: Tokyo Green Building Label

- Tokyo Metropolitan Government‘s Green Labeling System for Condominiums.
- Green Labeling System for Condominiums (2002, revised in 2005 & 2010), mandatory for new construction and major refurbishment to organize and publish information based on a) building insulation, b) energy efficiency & performance, c) lifespan extension (durability) and d) greening (plants etc.) of the building.
- The evaluation results for the respective items are expressed as a number of star symbols, max: ★ ★ ★ .

Method: Hedonic model

$$P_{(i,j,t)} = f(G_i, X_{(i,j)}, NE_k, HH_{(i,j)})$$

$P_{(i,j,t)}$: New condominium price of condominium *i* and dwelling *j* at time *t* (1: asking price, 2: transaction price)

G_i : Green label of condominium *i*

$X_{(i,j)}$: Building characteristics of condominium *i* & dwelling *j*

NE_k : Location characteristics of region *k*

$HH_{(i,j)}$: Buyer characteristics of condominium *i* and dwelling *j*

(Quasi) cross-sectional hedonic model with robust S.E., time fixed effects and buyer characteristics

Data:

Tokyo condominium prices database with property and buyer characteristics 2001-2011 (N=48,740):

- Data source: Japanese Real Estate Economic Institute's database (*asking prices* & characteristics of property) combined with large-scale questionnaire survey of *transaction prices* and household characteristics (Recruit., Co).
- Variables:

  Asking price, transaction price, name of development company, development scale, size and age of property, location characteristics (coordinates, address, nearest station, distance to nearest station), building characteristics (building area, land area, building structure).

Data:

Variables (continued):

- Buyer characteristics (age of buyer, annual income, size of family, etc.) gathered by **questionnaire survey** of the Recruit.
- Tenure type (leasehold types etc.)
- Property management type (24-hour etc.)
- First-month contract rate (i.e. time on market). Higher the first month contract rate, the more affordable prices are in relation to the condominium's features.

Estimation Models:

Model.1

$$\log P_{(i,j,t)} = a_0 + a_1 T_{(i,j)} + \boxed{a_2 G_i} + a_3 \boxed{G_i T_{(i,j)}} + \sum_m a_4^m X_{(i,j)}^m$$

$$+ \sum_n a_5^n NE_k^n + \sum_t a_6^t D_t + \varepsilon_{(i,j)}$$

Model.2

$$\log P_{(i,j,t)} = a_0 + a_1 T_{(i,j)} + \boxed{a_2 G_i} + a_3 \boxed{G_i T_{(i,j)}} + \sum_m a_3^m X_{(i,j)}^m$$

$$+ \sum_n a_4^n NE_k^n + \sum_s a_5^s HH_{(i,j)}^s + \sum_t a_6^t D_t + \varepsilon_{(i,j)} + \sum_t \boxed{a_7^t G_i D_t} + \varepsilon_{(i,j)}$$

Model.3

$$\log P_{(i,j,t)} = a_0 + a_1 T_{(i,j)} + \boxed{a_2 G_i} + a_3 \boxed{G_i T_{(i,j)}} + \sum_m a_3^m X_{(i,j)}^m$$

$$+ \sum_n a_4^n NE_k^n + \sum_s a_5^s \boxed{HH_{(i,j)}^s} + \sum_t a_6^t D_t + \varepsilon_{(i,j)} + \sum_t \boxed{a_7^t G_i D_t} + \varepsilon_{(i,j)}$$

Estimation result, *Base Model*: Model 1

	(1) baseline OLS	(2) Robust reg
	lp	lp
Transaction price discount	-0.0347***	-0.0316***
	(-27.87)	(-26.54)
Green asking price premium	0.0609***	0.0586***
	(18.66)	(18.31)
Green transaction price discount²	-0.00918*	-0.00948**

Transaction Based Premium $0.0609 - 0.00918 = 0.05172$

Estimation result considering *Time Effect* : Model 2

	Model 2: log(price)
Regressor	**Coefficient**
green2005	**0.045****
green2006	**0.0487*****
green2007	**0.0596*****
green2008	**0.0844*****
green2009	**0.096*****
green2010	**0.0438*****
tgreen2005	**-0.0486****
tgreen2006	**-0.003**
tgreen2007	**0.010**
tgreen2008	**-0.034****
tgreen2009	**-0.029****
tgreen2010	**0.008**
Property & condo attributes	Yes
Developer fixed effects	Yes
Location controls	Yes
Management fixed effects	Yes
Buyer characteristics	Yes
Time fixed effects	*Yes*
N	**48,740**
R^2	**0.814**

Robustness Test: Estimated Result considering *Household's Characteristics:* Model 3.

	(1) baseline OLS	(2) Robust reg	(3) Income Q1	(4) Income Q2	(5) Income Q3	(6) Income Q4
	lp	lp	lp	lp	lp	lp
Transaction price discount	-0.0347***	-0.0316***	-0.0359***	-0.0354***	-0.0337***	-0.0343***
	(-27.87)	(-26.54)	(-11.72)	(-15.94)	(-16.16)	(-13.37)
Green asking price premium	0.0609***	0.0586***	0.0408***	0.0398***	0.0702***	0.0777***
	(18.66)	(18.31)	(3.63)	(6.74)	(13.12)	(12.45)
Green transaction price discount^2	-0.00918*	-0.00948**	-0.0158	-0.00692	-0.00936	-0.00975
			0.025	0.03288	0.06084	0.06795

Conclusions: **Interpretation**

- Compared to non-labelled properties, labelled buildings commanded a premium of 6.09% for the base asking price and 5.19% for the base transaction price (6.09% - 0.9%).
- Premium appears to *rise over time* (exception: 2010)
- Green asking price premia are found to *progress with increasing incomes of buyers* (from 4% to nearly 8%).

→

- The average price premium observed in recorded transaction prices is mainly driven by households with above-average incomes paid for green-labelled properties.

Case2: Aging and Housing Market

Empirical method: ECM & Estimation Models.

Nishimura (2011), Nishimura and Takáts (2012) , Takáts (2012)

$$\Delta \ln P_{it} = \alpha + \beta \Delta \ln GDPPC_{it} + \gamma \Delta \ln OLDDEP_{it} + \delta \Delta \ln TPOP_{it} + \varepsilon_{it} \quad (1)$$

GDPPC is per capita GDP,

OLDDEP is the old age dependency ratio, which is defined by the ratio of population aged 65+ to the working population (i.e. population aged 20-64),

TPOP is total population.

The disturbance term is represented by ϵ_{it} .

Data:

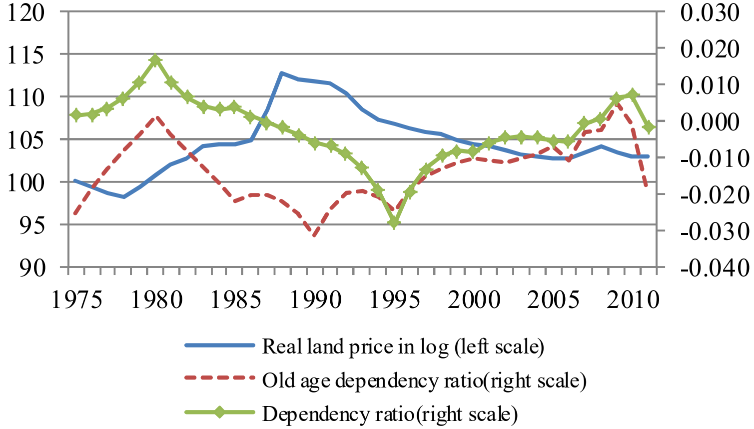
	Japan	U.S.
Data format	Prefectural panel data	State panel data
Data period	1976 to 2010	1975 to 2011
Housing price	Estimated hedonic function based on Ministry of Land, Infrastructure, Transport and Tourism "Land Price Data (Residential Land)," we set data for representative locations by prefecture and estimated quality-adjusted public land prices (amount base) using estimated hedonic function	We estimated state-by-state quality-adjusted housing prices (amount base) using Federal Housing Finance Agency "All-Transactions Indexes" rates of change and "Summary Statistics for House Prices" price median values
Income	Prefectural income based on the Cabinet Office's "Prefectural Economic Accounts" (linked using price comparisons at base points in time)	U.S. Department of Commerce, "Bureau of Economic Analysis" GDP by state (Chained by price ratio between base periods)
Interest rate	National value from Bank of Japan's "Average Contractual Interest Rate on Bank Loans" (synthesized rate for all Japanese banks)	Federal Reserve Board, "Contract Rate on 30-Year, Fixed-Rate Conventional Home Mortgage Commitments" (National)
Consumer price index	Consumer price index by prefectural capital (synthesized) from Statistics Japan's "Consumer Price Index"	United States Department of Labor, "Bureau of Labor Statistics" CPI (All Items) by state
New housing supply	New housing starts (total number for owned homes, rental homes, issued housing, and condominiums) from the Ministry of Land, Infrastructure, Transport and Tourism's "Statistical Survey of Construction Starts"	U.S. Census, "Building Permits Survey," New Privately-Owned Housing Units Authorized by Building Permits by state
Population by age group	Based on 'national census(population ratios by five-year age groups),Ministry of Internal Affairs and Communications calculated population figures by multiplying these ratios by the population, demographics, and household data based on the Basic Resident Register	U.S. Census, "State Population Estimates" Population by age and state

Demographic Changes: Dependency Ratio and Old age dependency ratio

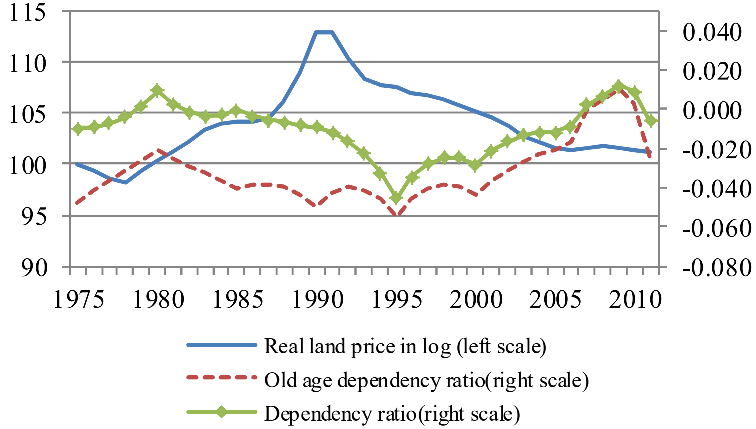
- Nishimura (2011)
- **Dependency Ratio** $= \dfrac{aged\ 0\text{–}19\ and\ 65+,}{population\ aged\ 20\text{–}64}$
-
- Takáts (2012)
- **Old age dependency Ratio** $= \dfrac{aged\ 65+,}{population\ aged\ 20\text{–}64}$

Describing Data: Graphical

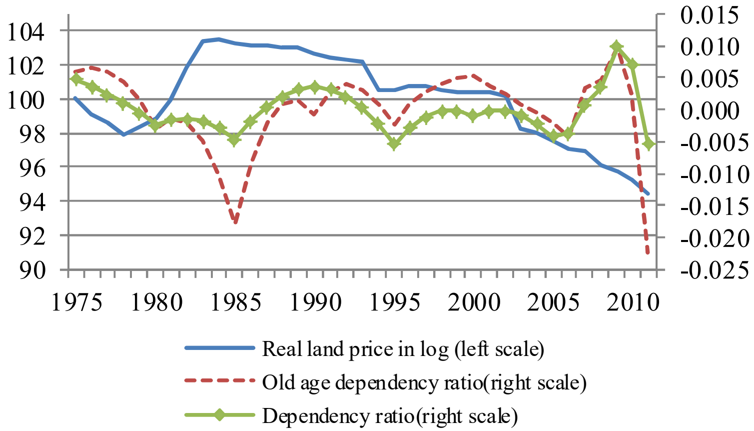
(Index: 1970=100) Tokyo



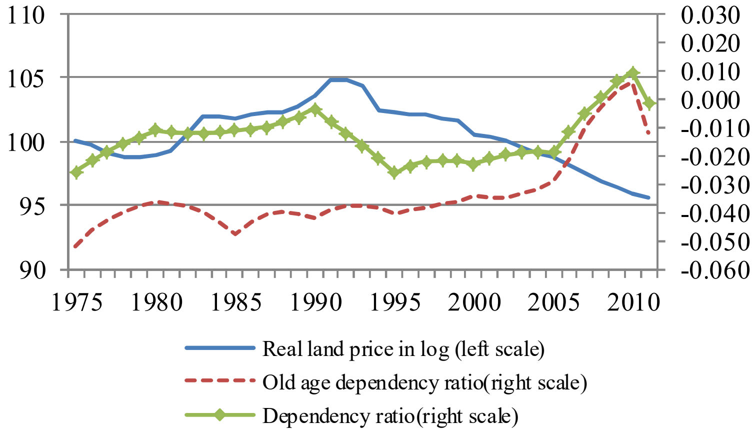
(Index: 1970=100) Osaka



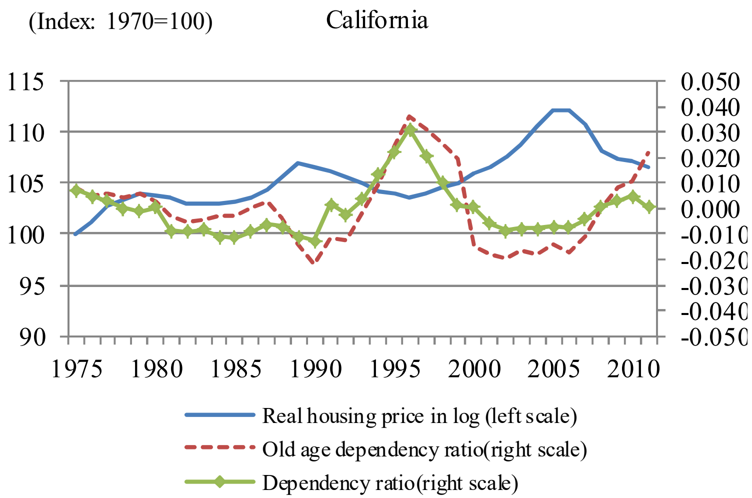
(Index: 1970=100) Aomori

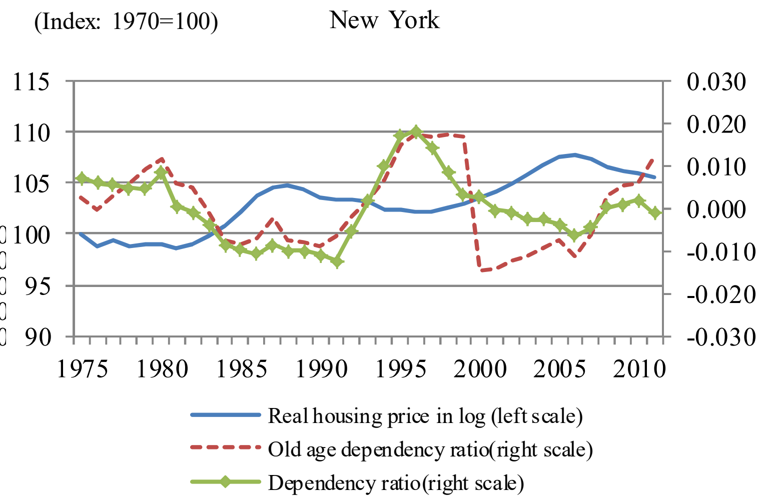


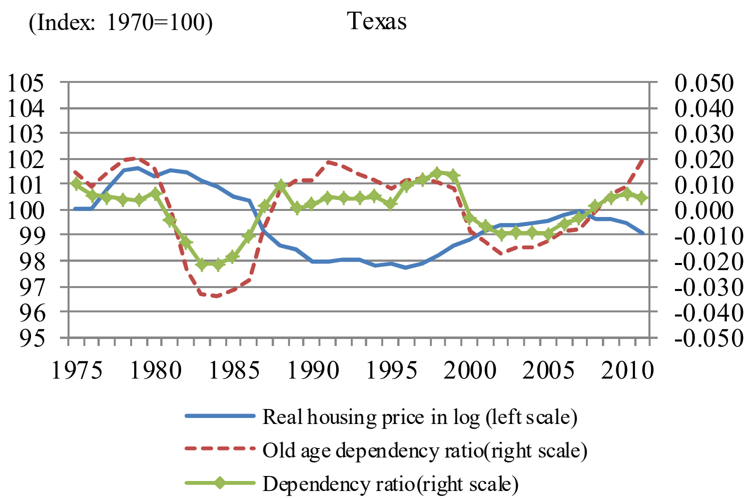
(Index: 1970=100) Kagawa

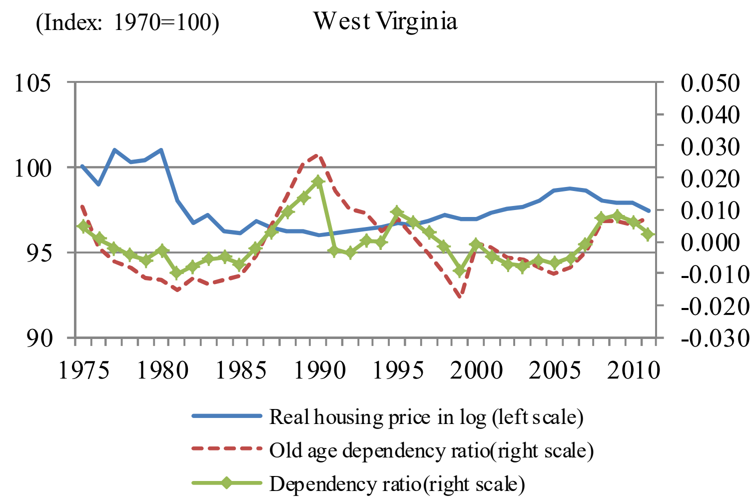


Describing Data: Graphical









Data Analyzing: Regression Model.

	No. of observations	Adj. R2	GDP per capita			Old dependency ratio			Total population			EC term		
Japan	1,645	0.629	**0.2188**	0.0000		**-1.3167**	0.0000		**0.9177**	0.00		-0.1033	0.00	
Standard error/t value			0.058	/	3.76	0.186	/	-7.06	0.290	/	3.17	0.009	/	-11.33
U.S.	1,836	0.439	**0.4515**	0.0000		**-0.9067**	0.0000		**0.7514**	0.00		-0.1272	0.00	
Standard error/t value			0.042	/	10.66	0.116	/	-7.79	0.116	/	6.46	0.010	/	-12.29

The coefficient on per capita GDP :

Japan 0.2188, US 0.4515, Takáts:0.8842.

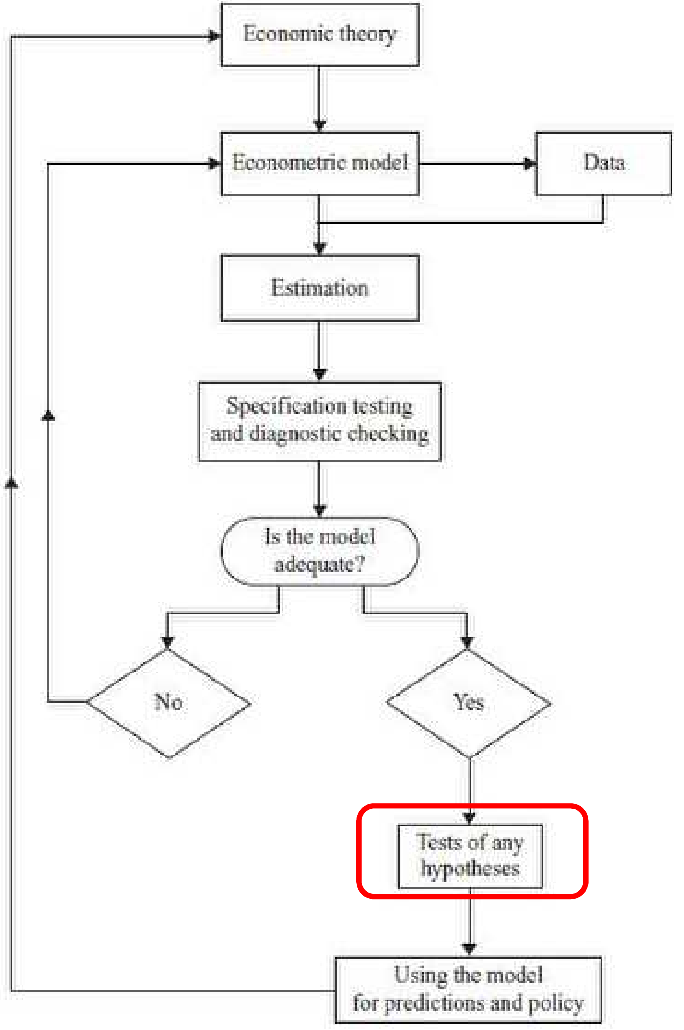
Old age dependency ratio:

Japan -1.3167, US -0.9067, Takáts:-0.6818.

Total population:

Japan 0.9177 , U.S. 0.7514, Takáts: 1.0547.

A revised schematic description of the steps involved in an econometric analysis of economic models.



G.S.Maddara and K. Lahiri (2009), “Introduction to Econometrics” 4th edition.

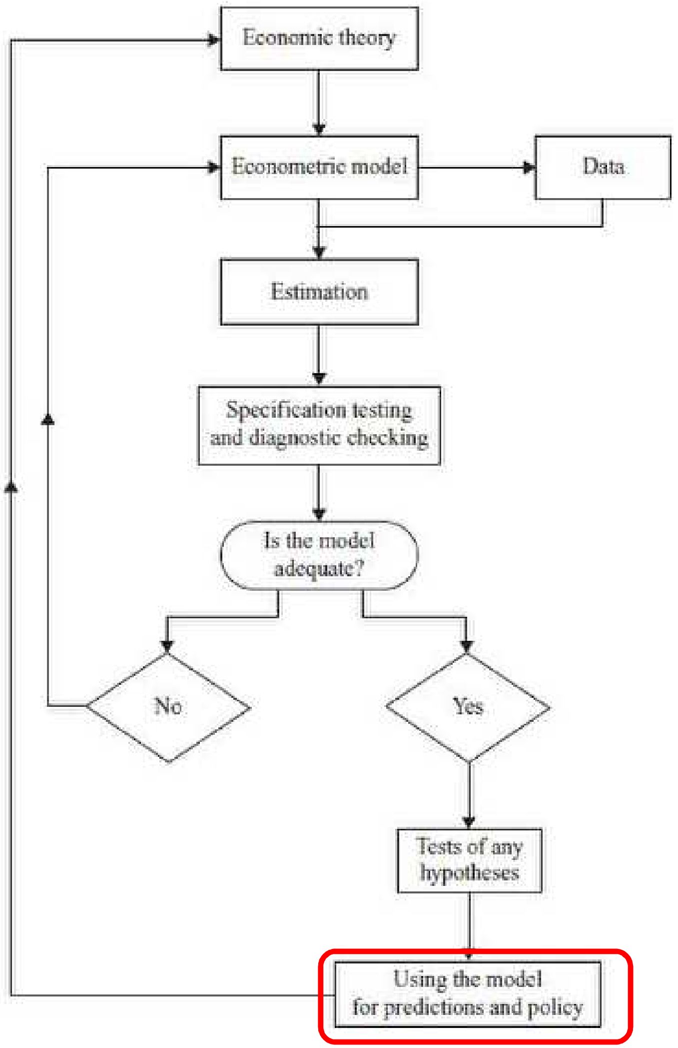
Robustness Check Japan

Model	No. of observations	Adj. R2	GDP per capita		Old age dependency ratio		Total population		Time fixed effect	Local fixed effect
Japan										
Base model: BM	1,645	0.629	**0.2188**	***	**-1.3167**	***	**0.9177**	***	Yes	None
without time fixed effect	1,645	0.159	**0.4401**	***	**-1.9702**	***	**2.5376**	***	None	None
with local fixed effect	1,645	0.621	**0.2302**	***	**-1.7280**	***	**2.0220**	***	Yes	Yes
with local fixed effect and without time fixed effect	1,645	0.182	**0.3891**	***	**-2.2071**	***	**4.0806**	***	None	Yes
without EC term	1,645	0.602	**0.1468**	**	**-1.0790**	***	**0.8333**	***	Yes	None
BM+ Interest rate	1,598	0.629	**0.1433**	**	**-1.4071**	***	**1.0508**	***	Yes	None
BM + New housing supply	1,645	0.627	**0.2297**	***	**-1.2701**	***	**1.1372**	***	Yes	None
BM + interest rate + new housing supply	1,598	0.629	**0.1664**	***	**-1.3675**	***	**1.2517**	***	Yes	None
BM + interest rate + new housing supply (1 period lag)	1,598	0.628	**0.0890**		**-1.3569**	***	**1.1941**	***	Yes	None

Robustness Check U.S.

Model	No. of observations	Adj. R2	GDP per capita		Old age dependency ratio		Total population		Time fixed effect	Local fixed effect
U.S.										
Base model: BM	1,836	0.439	**0.4515**	***	**-0.9067**	***	**0.7514**	***	Yes	None
without time fixed effect	1,836	0.247	**0.5874**	***	**-1.1576**	***	**0.6163**	***	None	None
with local fixed effect	1,836	0.454	**0.4525**	***	**-0.5363**	***	**1.8079**	***	Yes	Yes
with local fixed effect and without time fixed effect	1,836	0.263	**0.5847**	***	**-1.2666**	***	**0.8503**	***	None	Yes
without EC term	1,836	0.394	**0.4714**	***	**-0.7821**	***	**0.8222**	***	Yes	None
BM+ Interest rate	1,783	0.449	**0.4415**	***	**-0.9375**	***	**0.7385**	***	Yes	None
BM + New housing supply	1,834	0.459	**0.3819**	***	**-0.7824**	***	**0.6308**	***	Yes	None
BM + interest rate + new housing supply	1,783	0.468	**0.3725**	***	**-0.8128**	***	**0.6139**	***	Yes	None
BM + interest rate + new housing supply (1 period lag)	1,783	0.469	**0.4555**	***	**-0.6489**	***	**0.4272**	***	Yes	None

A revised schematic description of the steps involved in an econometric analysis of economic models.



G.S.Maddara and K. Lahiri (2009), “Introduction to Econometrics” 4th edition.

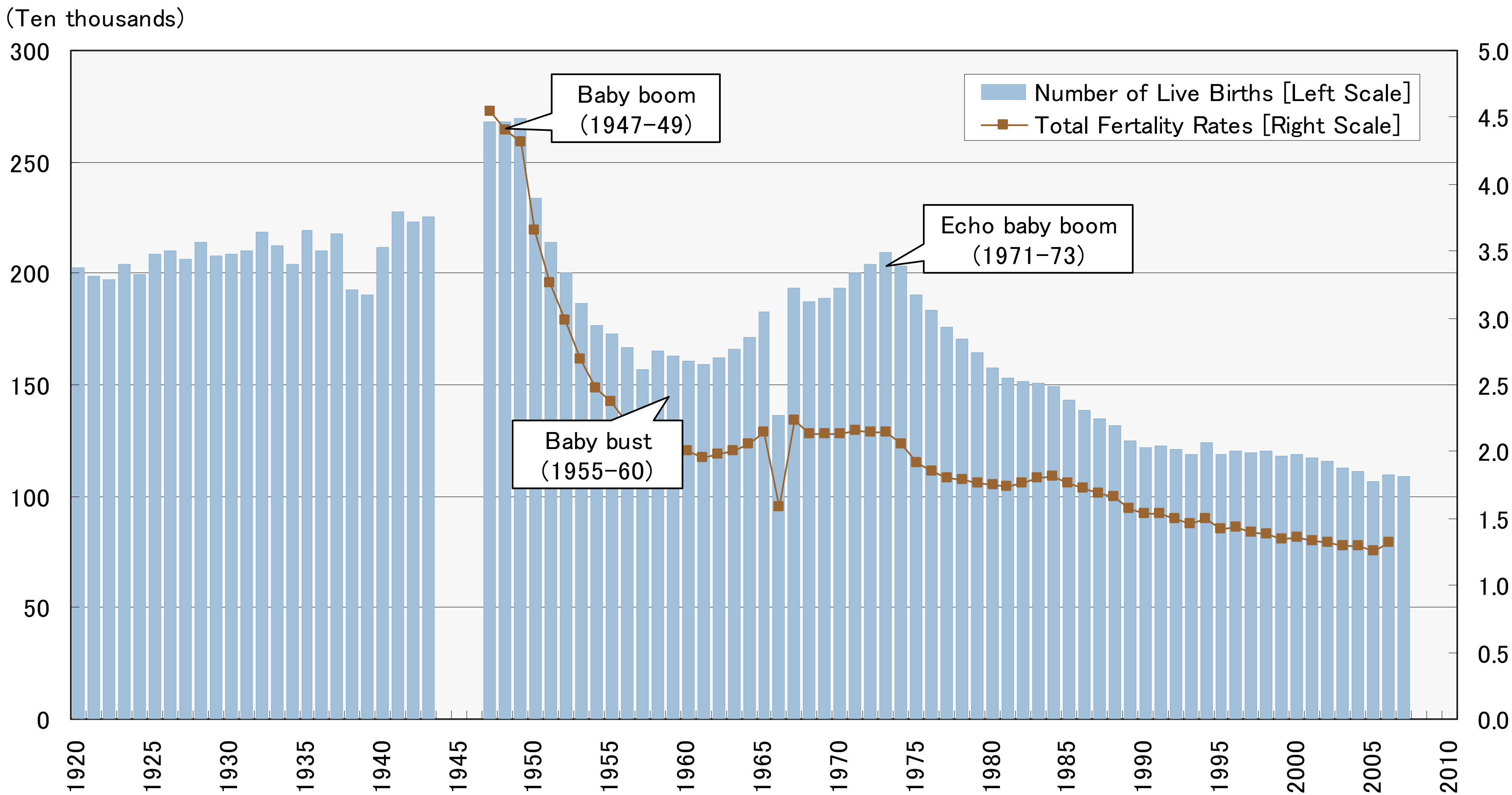
Forecast:

- Forecast the real land prices in Japan using the regression,
- The projection on demographic changes released by the **IPSS(National Institute of Population and Social Security Research).**

- Based on natural increases/decreases calculated from the survival probability and the number of births by cohort and social increases/decreases due to movement between regions.

- Population projections : **the medium variant projection**, which is based on the assumption of medium fertility, unless otherwise mentioned.

Demand: Number of live births (JPN)

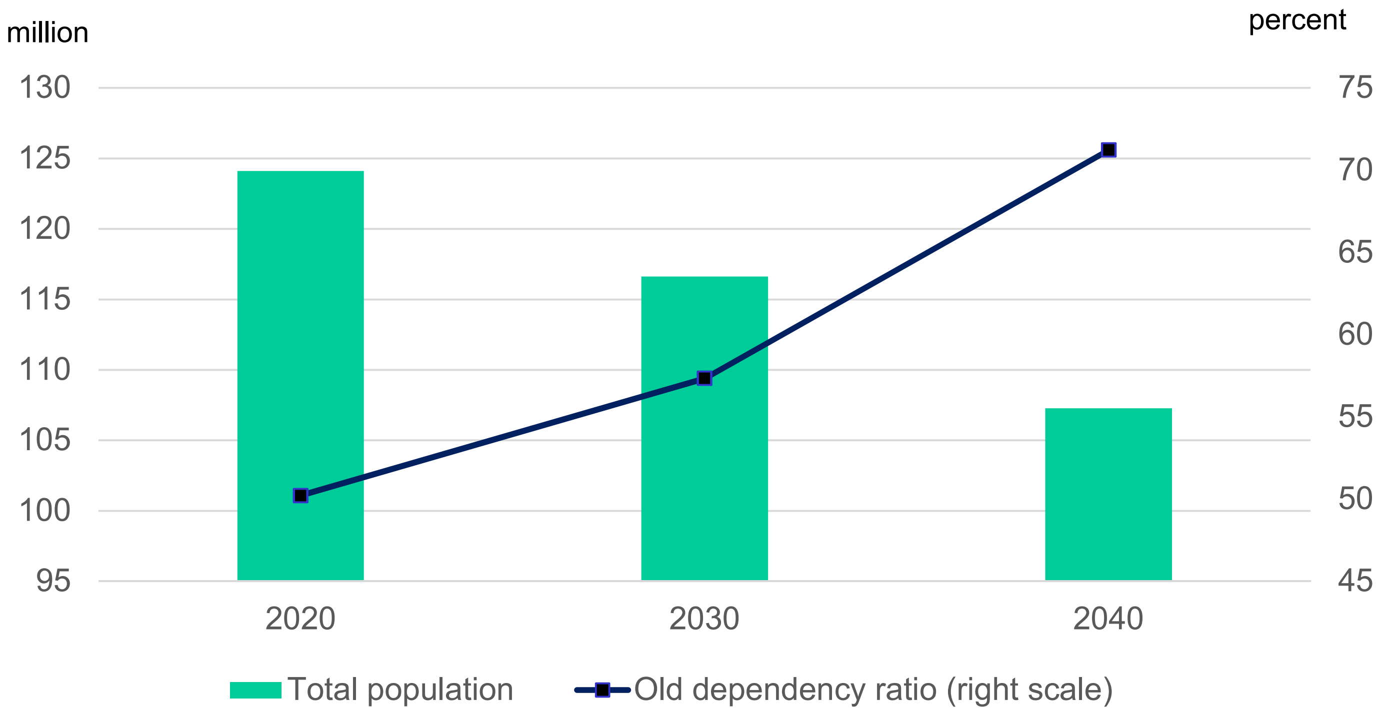


Source: Ministry of Health, Labor and Welfare

Shimizu,C and T.Watanabe(2010), "Housing Bubble in Japan and the United States," Public Policy Review Vol.6, No.2,pp.431-472

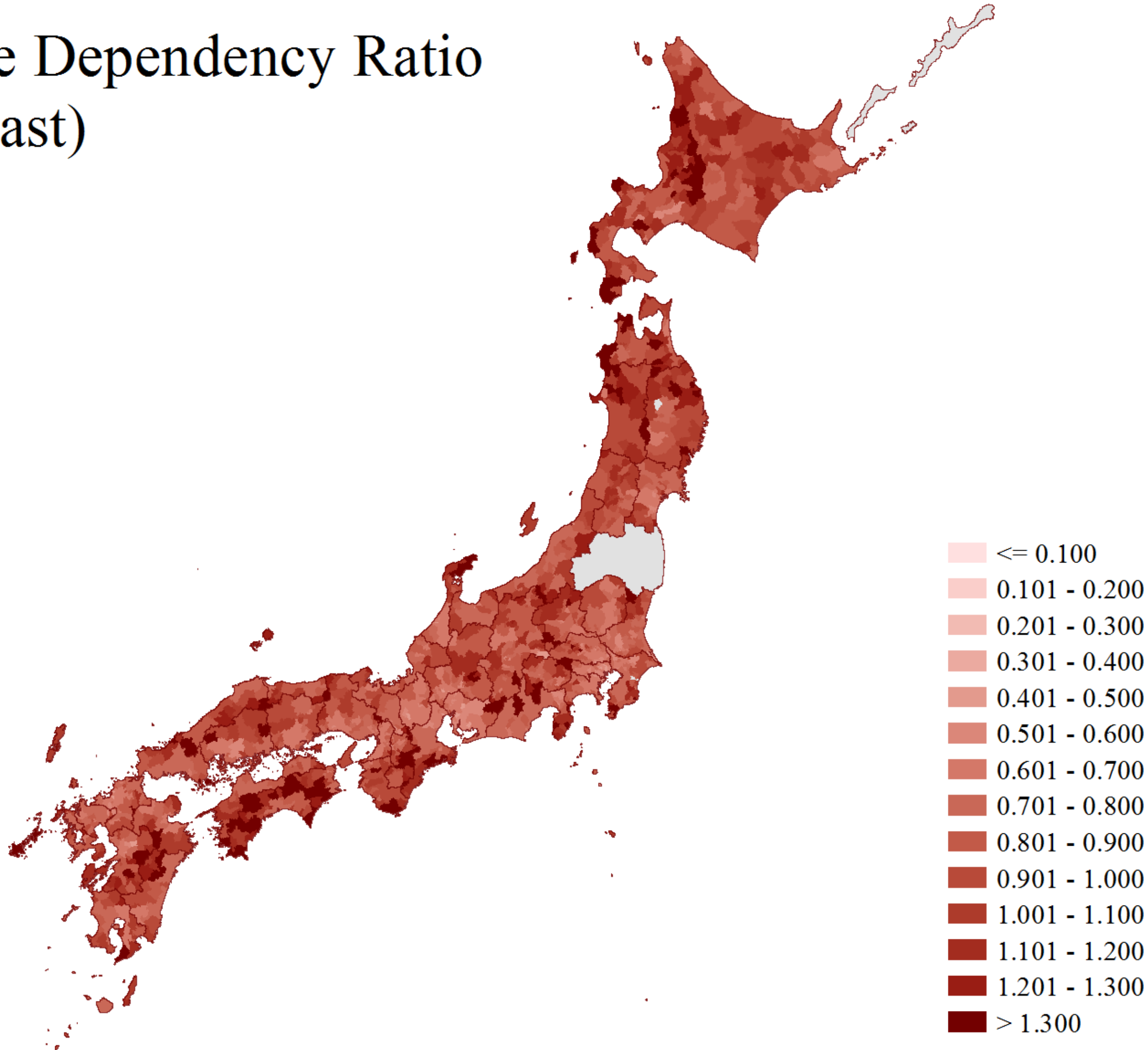
Old Age Dependency Ratio

- Assumption on future population
  - The **medium variant projection** on demographic changes calculated by **IPSS(National Institute of Population and Social Security Research)**



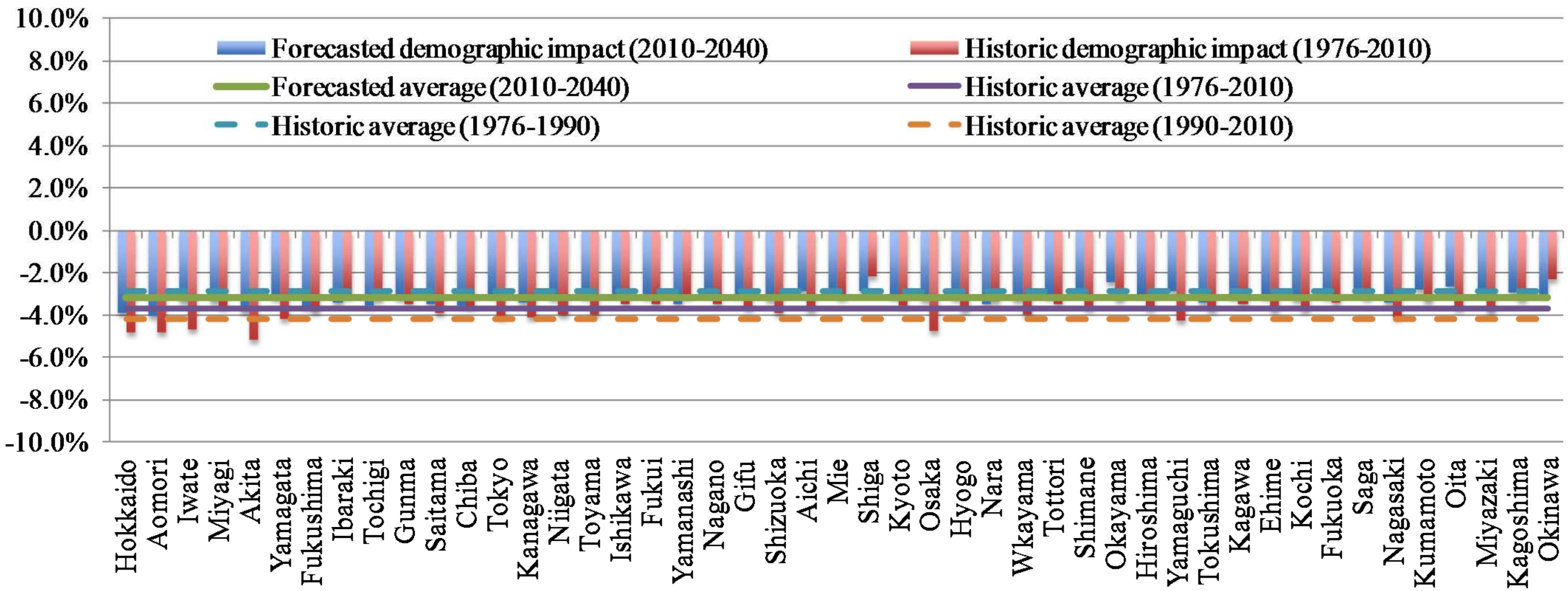
Note : IPSS projection is based on natural increases/decreases calculated from the survival probability and the number of births by cohort and social increases/decreases due to movement between regions. .

The Old Age Dependency Ratio 2040 (Forecast)



Source: Authors' calculation. The map is provided by Ministry of Land, Infrastructure, Transport and Tourism, "National Land Numerical Information: Administrative Zones Data."

Historic and Forecasted Demographic Impacts on Land Prices

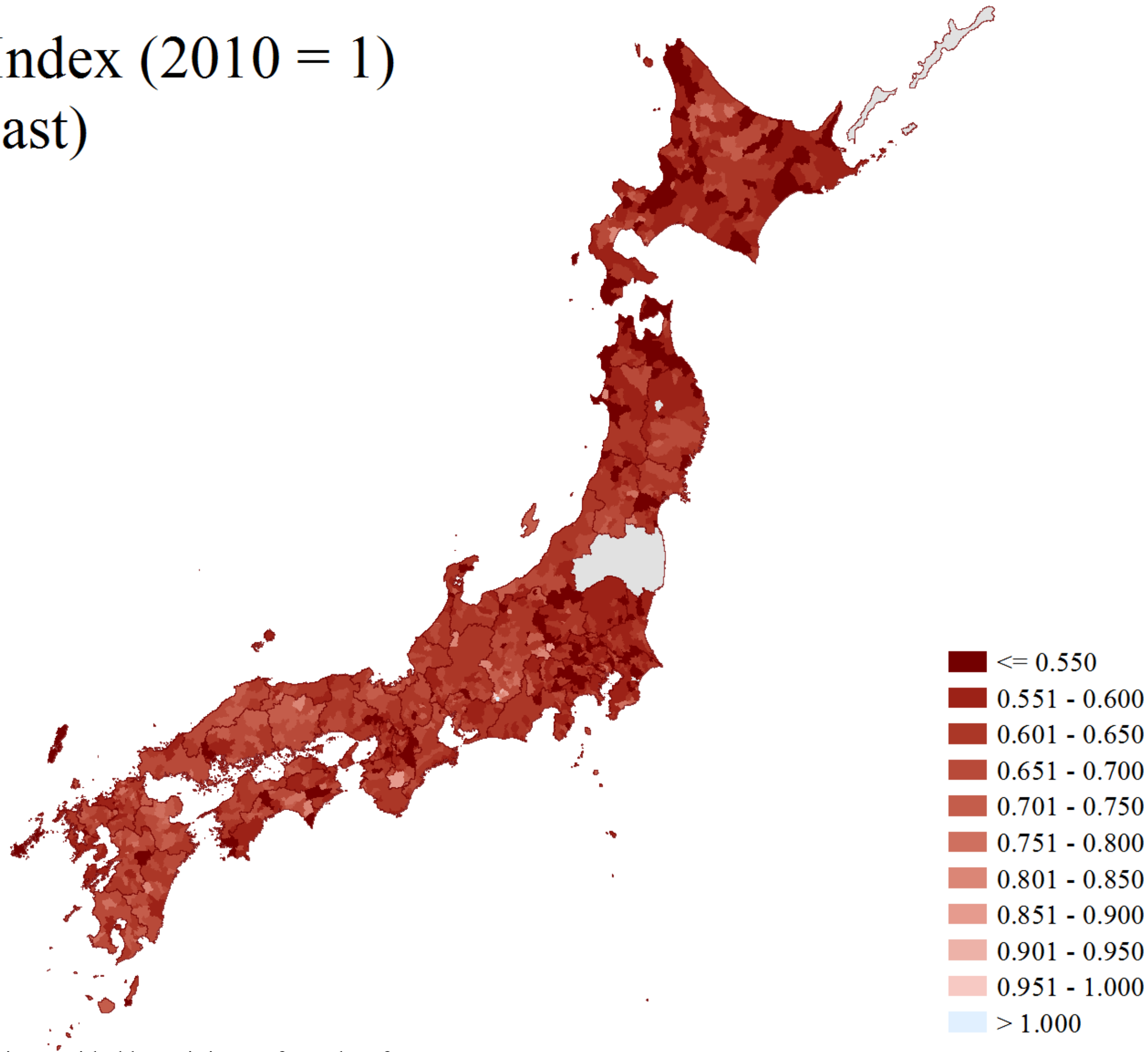


The contribution of demographic changes:

1976-2010 : -3.8 percent per year

2010-2040 : -2.4 percent per year

Land Price Index (2010 = 1)
2040 (Forecast)

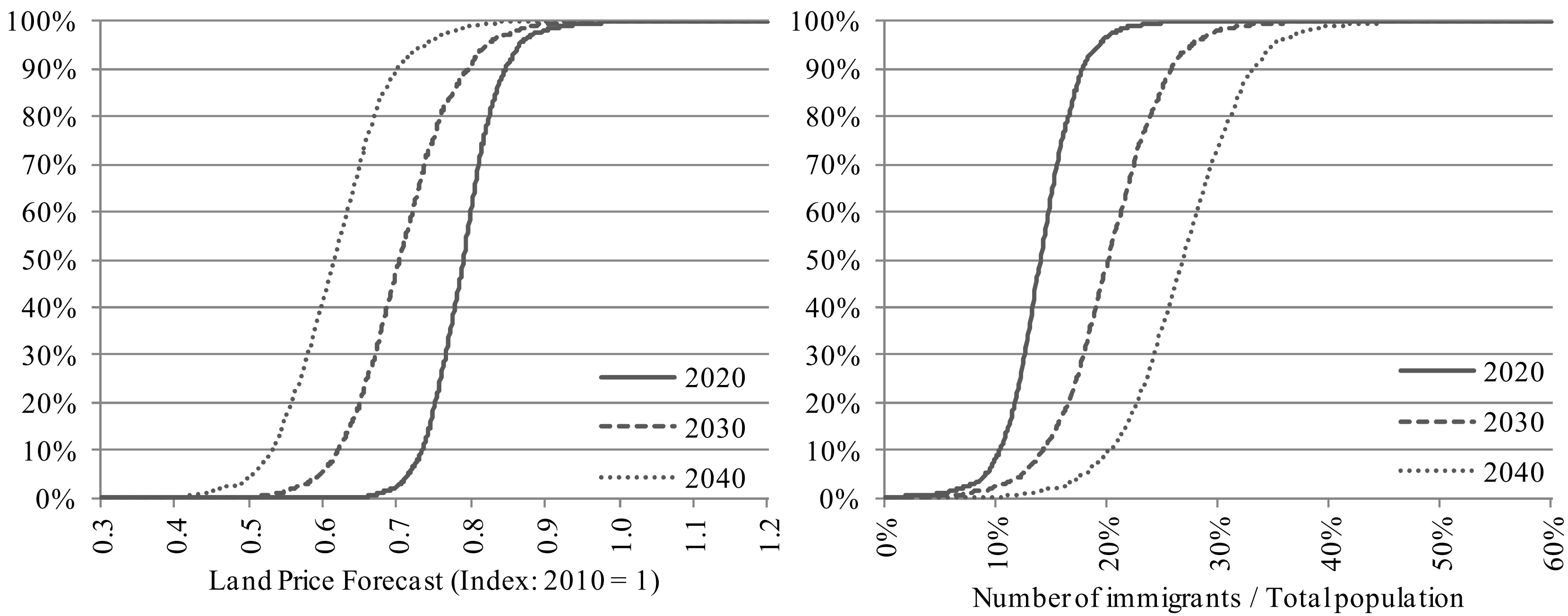


Source: Authors' calculation. The map is provided by Ministry of Land, Infrastructure, Transport and Tourism, "National Land Numerical Information: Administrative Zones Data."

Simulation:

- First, to examine the extent to which housing demand would be created by accepting immigrants and to what degree this would offset the decrease in residential land prices, we will estimate **the number of immigrants** (foreign workers) that would be needed to maintain the 2010 land price level.
- Second, we will estimate the extent to which housing demand would be created and residential land price decreases offset if **the retirement age were raised from 65 to 70 or 75** in order to utilize the labor power of the older generation.
- Third, we will estimate the extent to which housing demand would be created and residential land price decreases offset if **the female employment rate were raised to the same level as the male employment rate** in order to utilize the labor power of women.

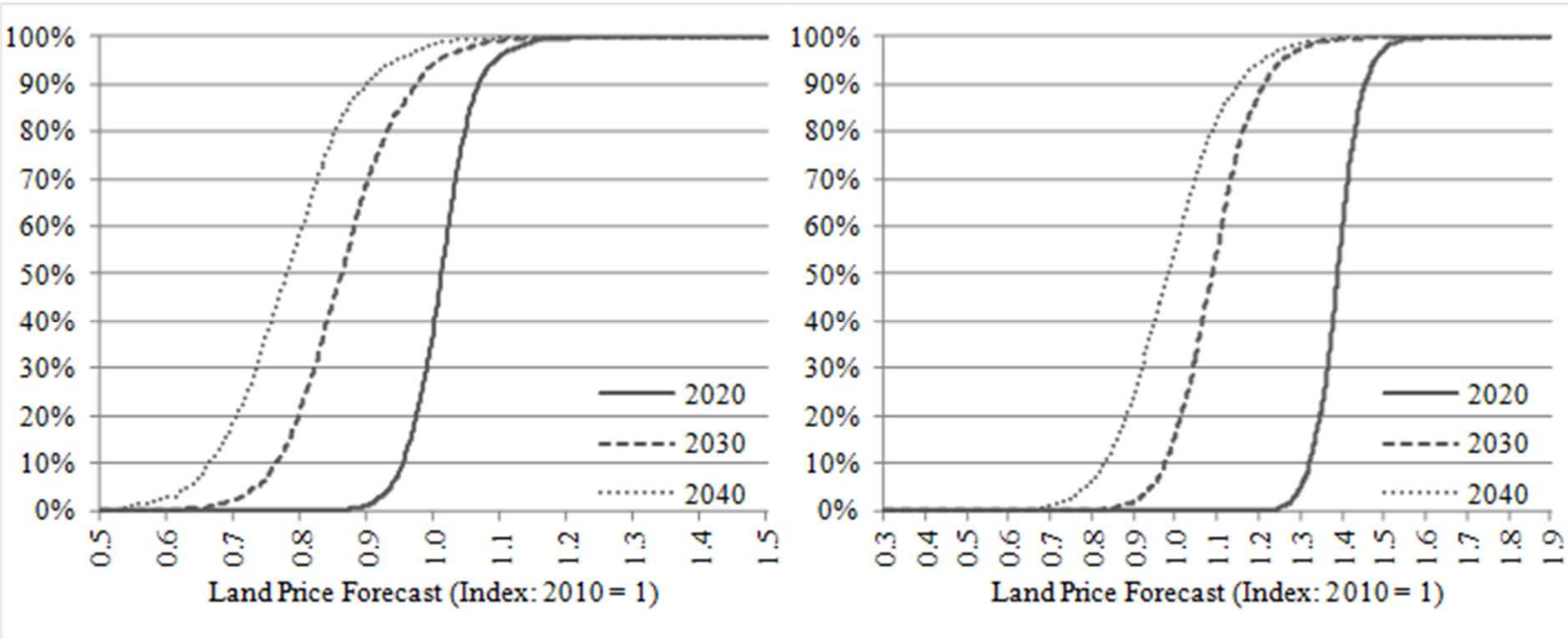
Simuration1 : Immigration Impact.



a) Simulation Results Based on Demographic Factors b) Proportion of Immigrants in Total Population

Residential Land Price Simulation Results and Proportion of Immigrants in Total Population

Simuration2 : Retirement Year.

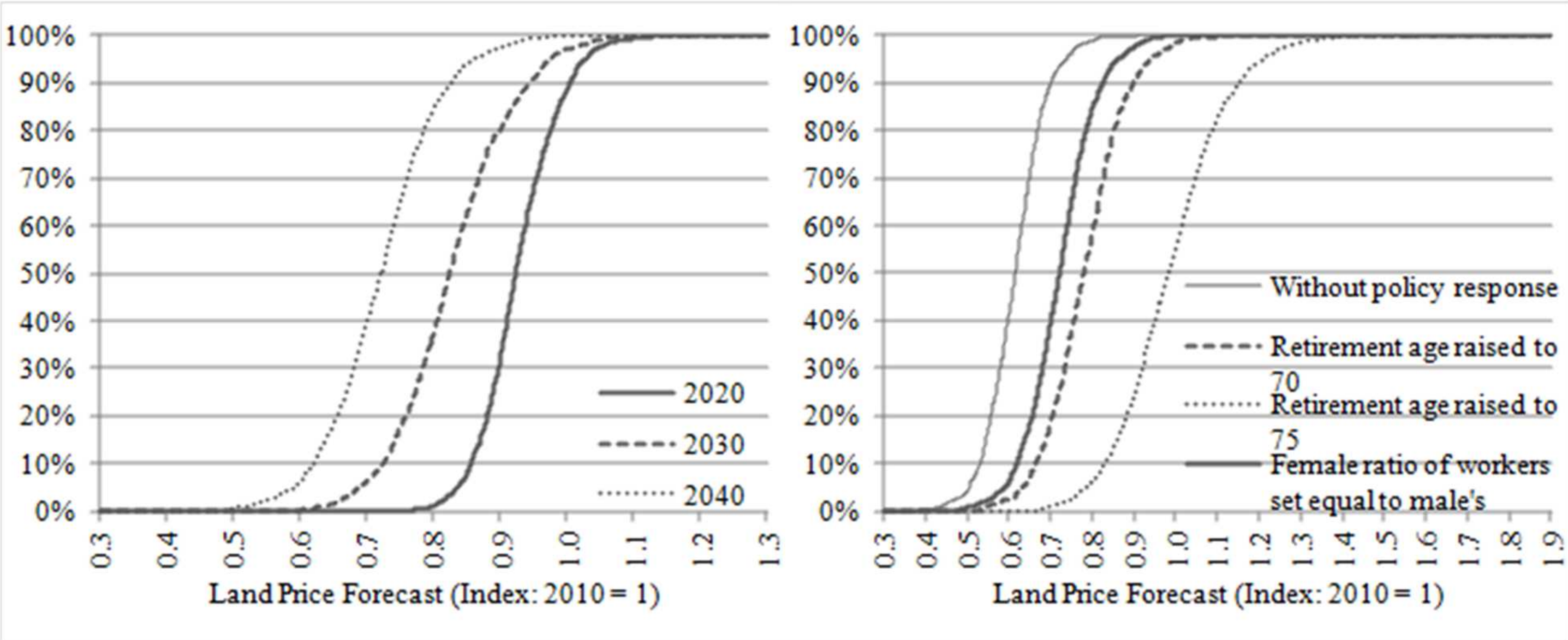


a) Effect If Retirement Age Raised to 70

b) Effect If Retirement Age Raised to 75

Residential Land Price Simulation Results (N = 1,683)

Comparison of Policy Effect.



a) Effect if Female Employment Rate Raised

b) Comparison of Different Policy Effects

Residential Land Price Simulation Results 3 (N = 1,683)

Interpretation:

- The results showed that a) around 40 million immigrants would be needed by 2040 to maintain housing asset values as of 2010. In other words, the ratio of foreigners in the total Japanese population would need to increase to approximately 30%. At the present time, it is extremely difficult to imagine Japan accepting this kind of society.
- Meanwhile, b) to promote the social advancement of women, even assuming a fixed birthrate, it would be necessary to provide more childcare alternatives for families in order to maintain the housing asset value to a little extent. If the aim of society were to increase the birthrate, then greater infrastructural development such as new daycares would be required, which would incur significant costs.

- The foregoing arguments suggest that the most effective and least costly policy would be c) extending the retirement age, especially given that life expectancy is expected to lengthen due to advances in medical technology.
- In reality, this solution could only be achieved by shaping the future by implementing multiple related policies. Moreover, it is necessary to recognize that the effects of such a policy would have only a temporary impact, as the baby boomers in Japan (born in 1947-1949) are entering their retirement age of 65 in 2015 and raising the retirement age has the effect of delaying this phenomenon for 5-10 years.
- Indeed, if the number of births and productive-age population do not increase, the problem would remain un-solved.

2. Machine Learning and Big Data.

- (1) What is big data?
- Big data is not a new phenomenon, but one that is part of a long evolution of data collection and analysis. Among numerous definitions of big data that have been introduced over the last decade, the one provided by Mayer-Schönberger and Cukier (2013) appears to be most comprehensive: Big data is **“ the ability of society to harness information in novel ways to produce useful insights or goods and services of significant value”** and **“things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value .”**

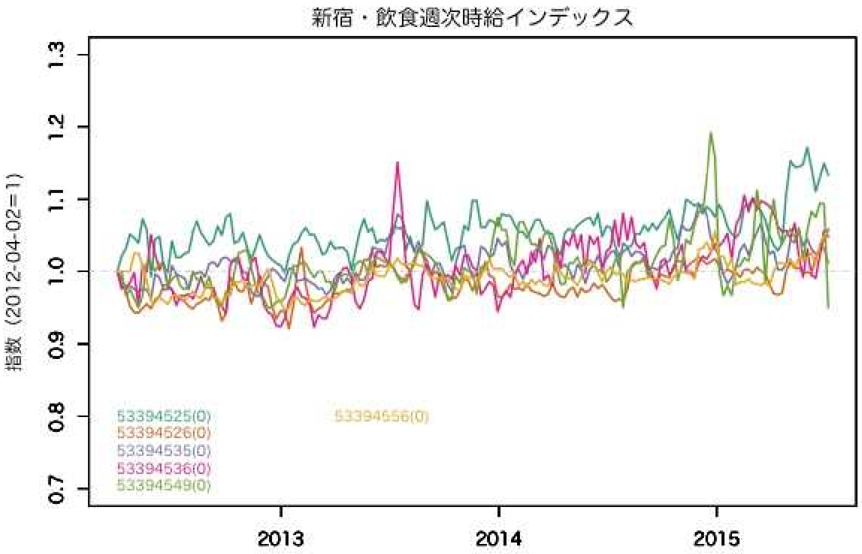
- Mayer-Schönberger,V and Cukier K, (2013), *Big Data: A revolution that will transform how we live, work and think*, UK: Hachette.

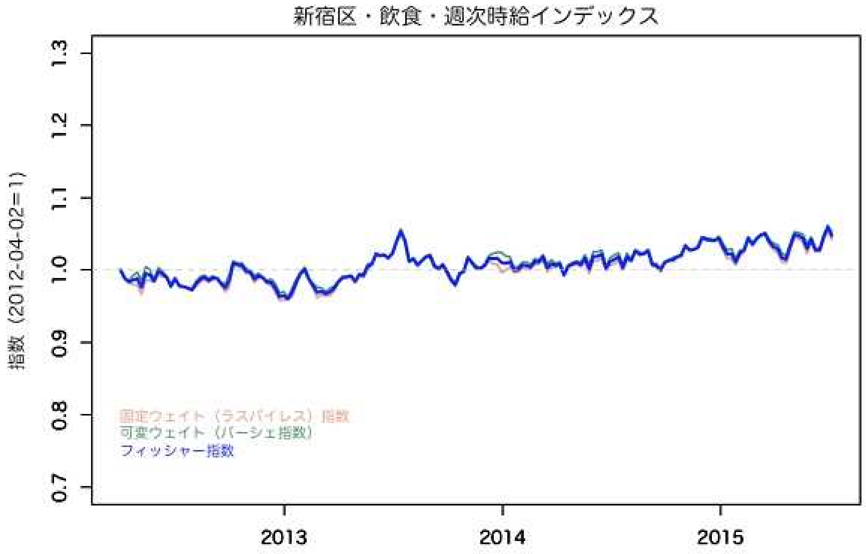
- In the community of analytics, it is widely accepted that big data can be conceptualized by the following three dimensions (Laney, 2001):

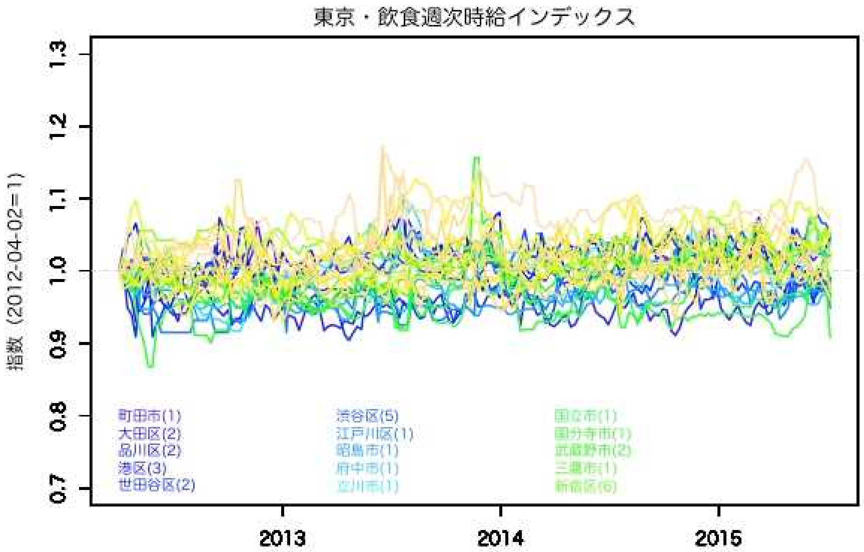
- **a). Volume**
- **b). Velocity**
- **c). Variety**

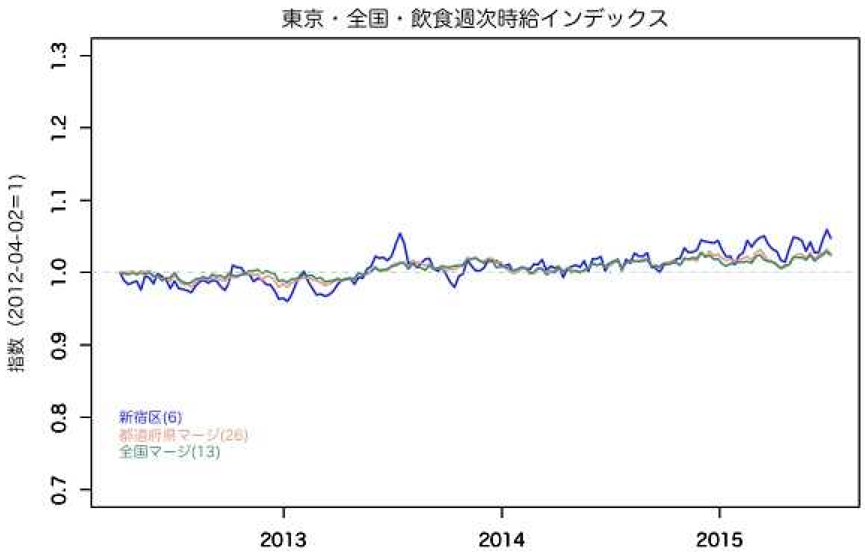
- Laney D,(2001), *3D data management: Controlling Data Volume, velocity, and variety*, US:META Group.

Real Time Index in Labor Market in Japan:

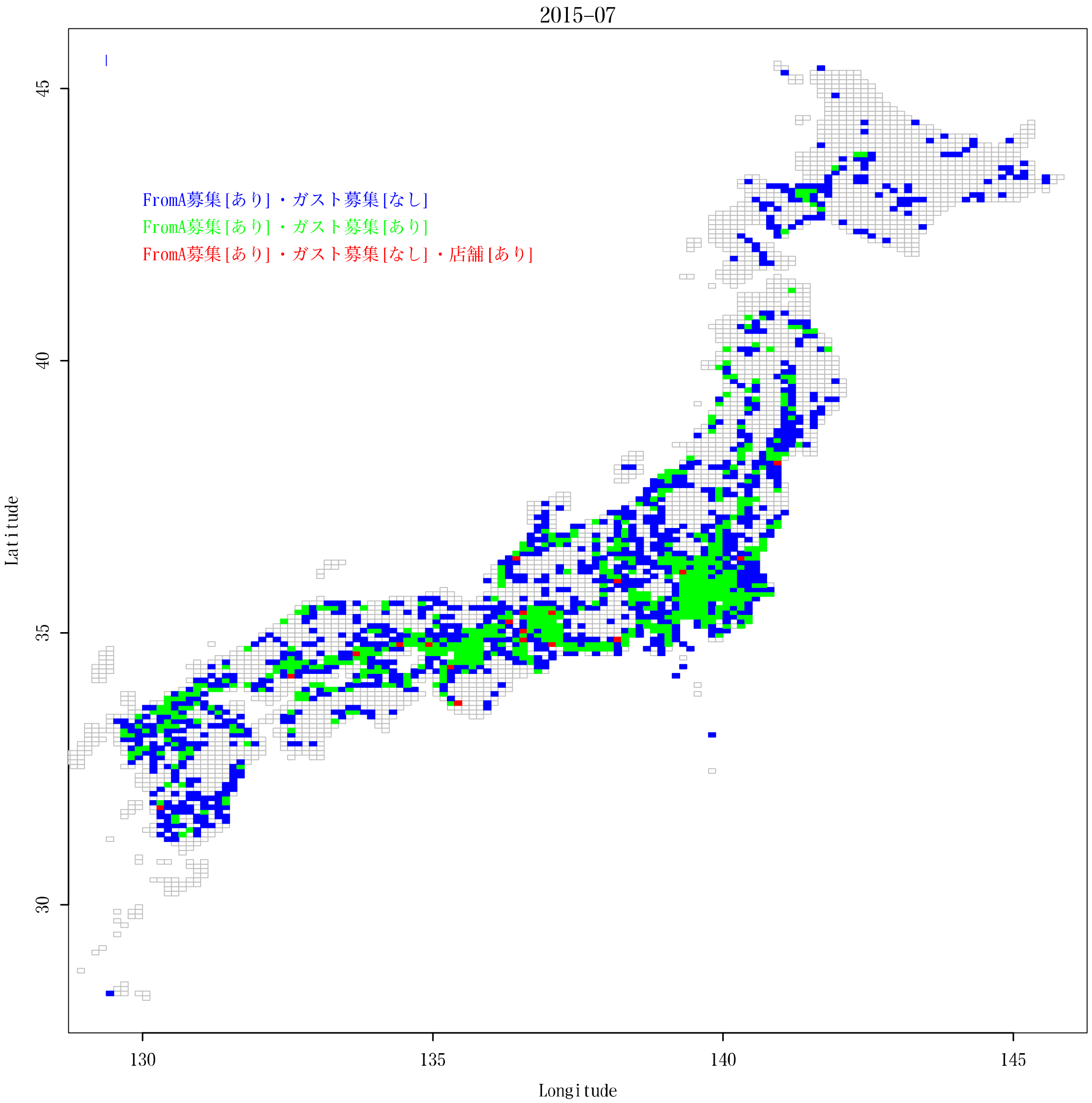






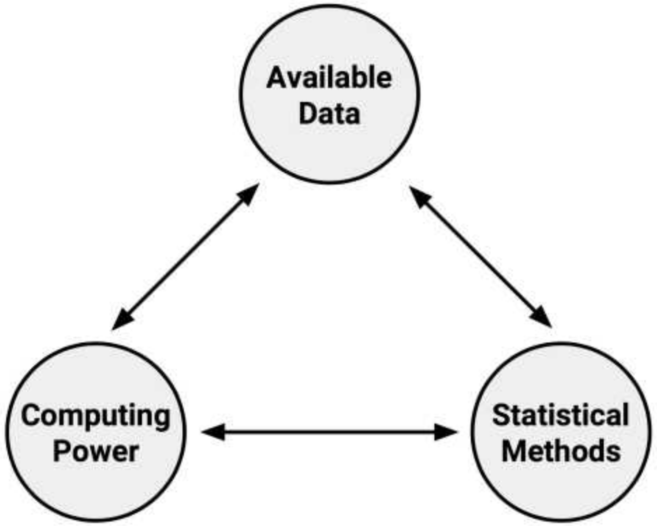


Heat Map


2015-07
FromA募集[あり]・ガスト募集[なし]
FromA募集[あり]・ガスト募集[あり]
FromA募集[あり]・ガスト募集[なし]・店舗[あり]
45
40
35
30
Latitude
130
135
140
145
Longitude

(2) Machine Learning.

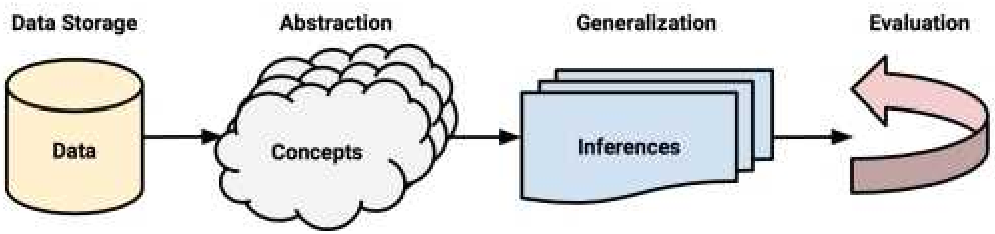
- The development of computer algorithms to transform data into intelligent action → machine learning .
- Available data, statistical methods, and computing power rapidly and simultaneously evolved.
- Growth in data necessitated additional computing power, which in turn spurred the development of statistical methods to analyze large datasets.



Brett Lanz (2015)

How Machine Learn?

- Regardless of whether the learner is a human or machine, the basic learning process is similar. It can be divided into four interrelated components:
- **a) Data storage**: utilizes observation, memory, and recall to provide a factual basis for further reasoning.
- **b) Abstraction**: involves the translation of stored data into broader representations and concepts.
- **c) Generalization**: uses abstracted data to create knowledge and inferences that drive action in new contexts.
- **d) Evaluation**: provides a feedback mechanism to measure the utility of learned knowledge and inform potential improvements.



Brett Lanz (2015)

Knowledge representation

- a) Mathematical equations
- b) Relational diagrams such as trees and graphs
- c) Logical if/else rules
- d) Groupings of data known as clusters

- →The choice of model is typically not left up to the machine. Instead, the learning task and data on hand inform model selection.

(3) Machine learning in practice

- **a) Data collection** : The data collection step involves gathering the learning material an algorithm will use to generate actionable knowledge.

- **b) Data exploration and preparation** : The quality of any machine learning project is based largely on the quality of its input data.

- **c) Model training** : By the time the data has been prepared for analysis, you are likely to have a sense of what you are capable of learning from the data.

- **d)Model evaluation** : Because each machine learning model results in a biased solution to the learning problem, it is important to evaluate how well the algorithm learns from its experience.

- **e)Model improvement** : If better performance is needed, it becomes necessary to utilize more advanced strategies to augment the performance of the model. Sometimes, it may be necessary to switch to a different type of model altogether.

(4) Matching input data to algorithms

Model	Learning Task
Supervised Learning Algorithms	
Nearest Neighbor	Classification
Naive Bayes	Classification
Decision Trees	Classification
Classification Rule Learners	Classification
Linear Regression	Numeric prediction
Regression Trees	Numeric prediction
Model Trees	Numeric prediction
Neural Networks	Dual use
Support Vector Machines	Dual use
Unsupervised Learning Algorithms	
Association Rules	Pattern detection
k-means clustering	Clustering
Meta-Learning Algorithms Bagging	Dual use
Boosting	Dual use
Random Forests	Dual use

Classification Method

- Several classification methods that have been popular in big data applications;

- a) *k* nearest neighbour algorithm,
- b) regression models,
- c) Bayesian networks,
- d) artificial neural networks and
- e) decision trees.

Typical classification task involving only two categories.

- A bank loan officer may want to evaluate whether approving a mortgage application is risky or safe, and consequently he/she needs to determine a label for the corresponding applicant (‘risky’ or ‘safe’).
- In practice, a classification task is implemented through the following three stages:
- **Stage 1** : Specify a suitable algorithm for classification, that is, a classifier.
- **Stage 2** : Optimize the selected classification algorithm using a set of training data.
- **Stage 3** : Make predictions using the optimized classification algorithm.

3. Validation.

- Forecasting Housing Prices.

- a) Regression models,

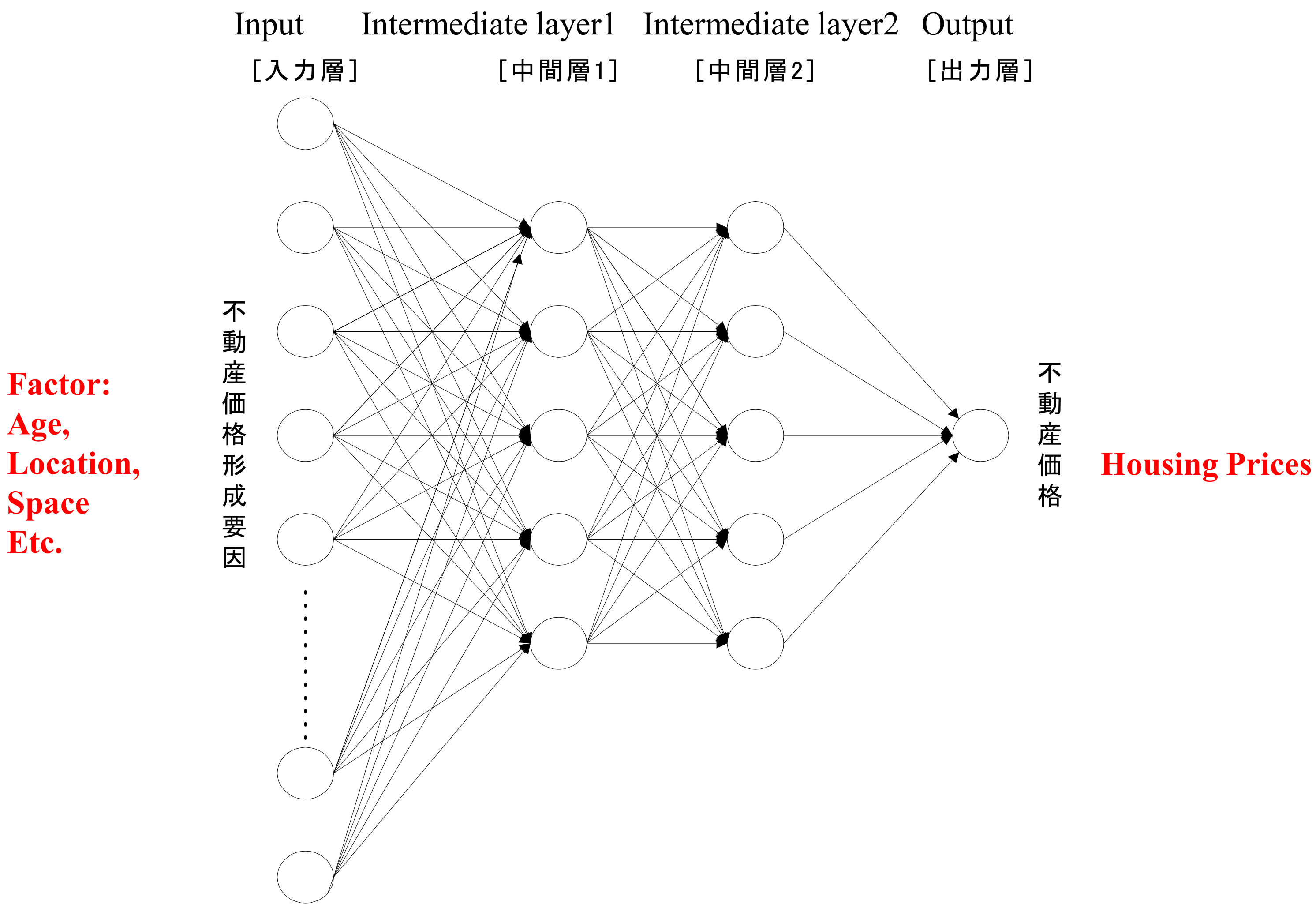
- b) Artificial neural networks and

- c) Decision trees.

Evaluation Bias

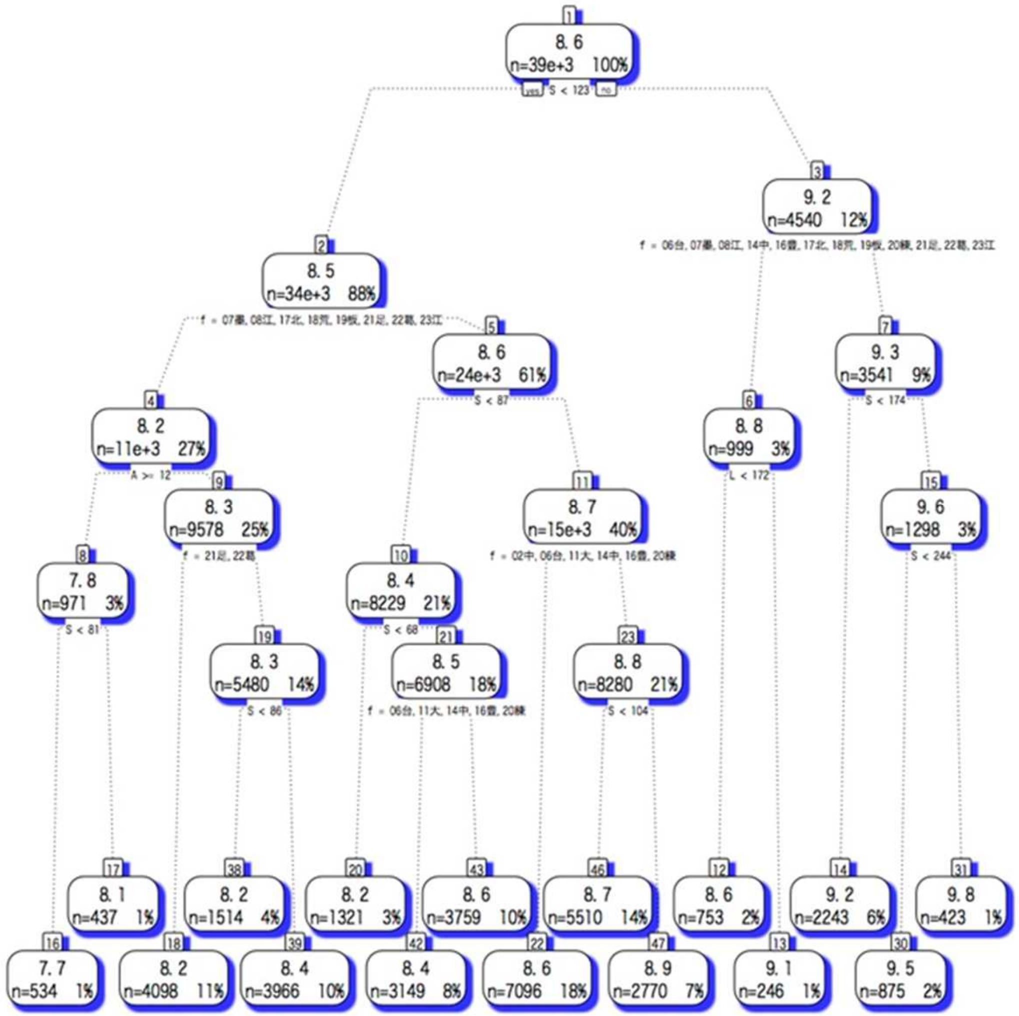
- Evaluation Bias is a necessary evil associated with the abstraction and generalization processes inherent in any learning task. In order to drive action in the face of limitless possibility, each learner must be biased in a particular way.
- Consequently, each learner has its weaknesses and there is no single learning algorithm to rule them all. Therefore, **the final step in the generalization process is to evaluate or measure the learner's success in spite of its biases and use this information to inform additional training if needed**.

Artificial neural networks



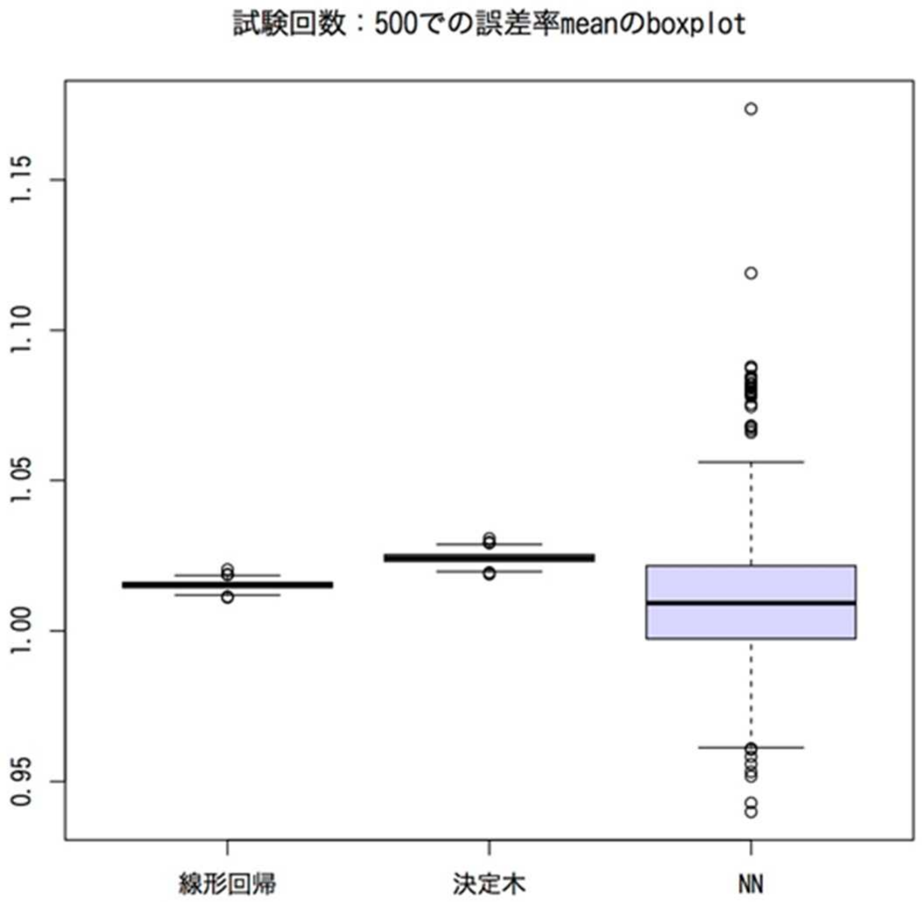
Chihiro Shimizu(2016), Introduction to Statistics for Market Analysis, Asakura-shoten. (in Japanese)

Decision Tree: Artificial intelligence /AI



Chihiro Shimizu(2016), Introduction to Statistics for Market Analysis, Asakura-shoten. (in Japanese)

Validation with Prediction Power: Boxplot of Number of trials 500 times with Sampling and Replacement.

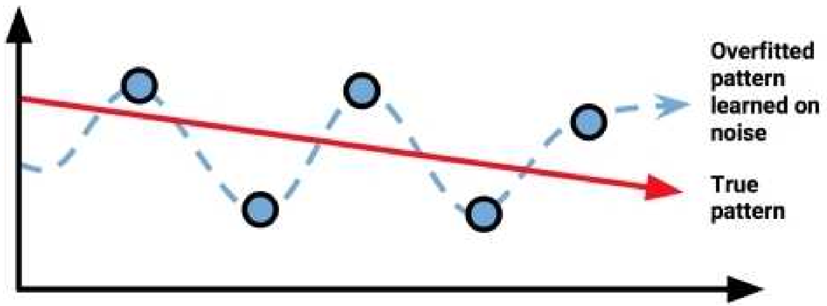


Regression **Decision Tree** **Neural Network**

Chihiro Shimizu(2016), Introduction to Statistics for Market Analysis, Asakura-shoten. (in Japanese)

Overfitting:

- A model that seems to perform well during training, but does poorly during evaluation, is said to be overfitted to the training dataset, as it does not generalize well to the test dataset.
- Solutions to the problem of overfitting are specific to particular machine learning approaches. For now, the important point is to be aware of the issue. How well the models are able to handle noisy data is an important source of distinction among them.



Brett Lanz (2015)

Introduction to Statistics03

- **Multiple Regression**

P. Newbold, W.L. Carlson, B. M. Thorne (2010), "Statistics for Business and Economics"7th edition.
Chapter 12: Multiple Regression

Today's Case:

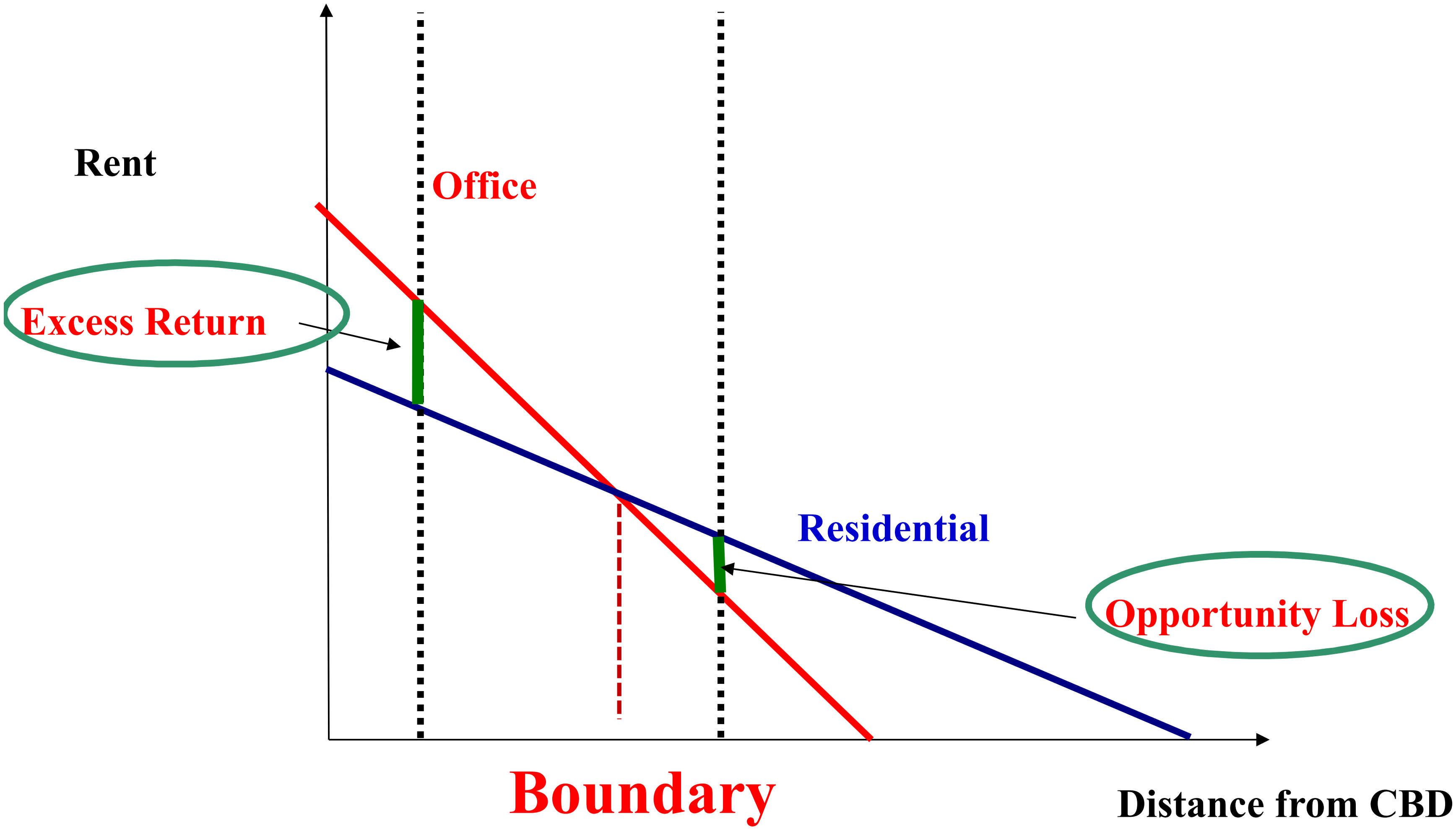
- Shimizu, C., K. Karato and Y. Asami (2010), "Estimation of Redevelopment Probability using Panel Data-Asset Bubble Burst and Office Market in Tokyo-,"Journal of Property Investment & Finance, 28(4), 285-300.

1.Motivations: The purpose of this research

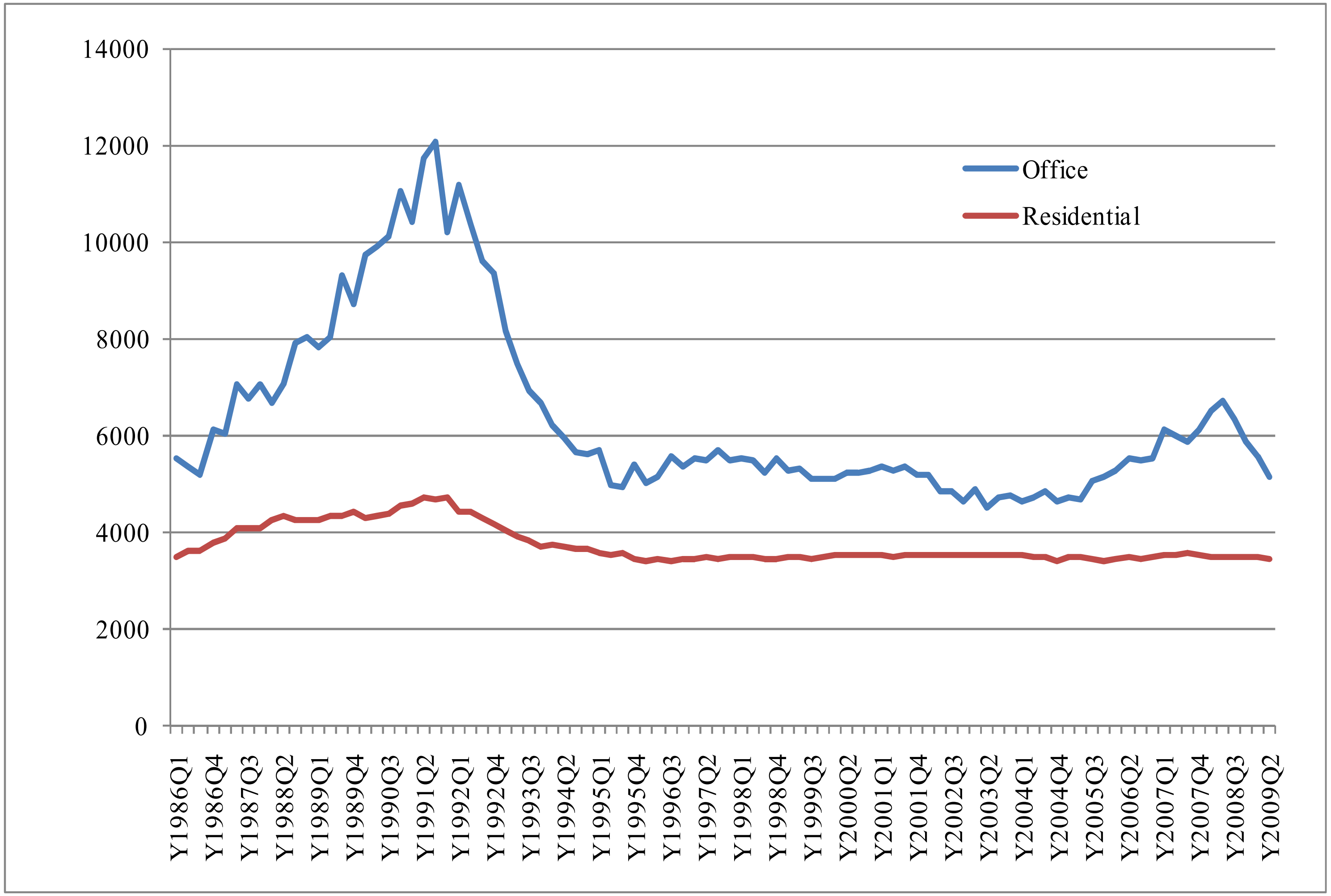
- 1.What happened in the Real Estate Market of Tokyo during the "lost decade" ?
- -What have we learned from these ups and downs in the real estate market?
- -Have recent real estate investment risk management efforts incorporated these lessons?

- 2. What are economic conditions for the redevelopment/conversion of buildings?
- - In the 1980s bubble, repeated rounds of speculative real estate transactions targeted urban areas in particular, and numerous lots were converted in poor ways.
- - Upon the collapse of the bubble, poorly located office spaces and the pencil buildings built on residential tracts suffered high vacancy rates.

Rent Curve :


Rent
Office
Excess Return
Residential
Opportunity Loss
Boundary
Distance from CBD

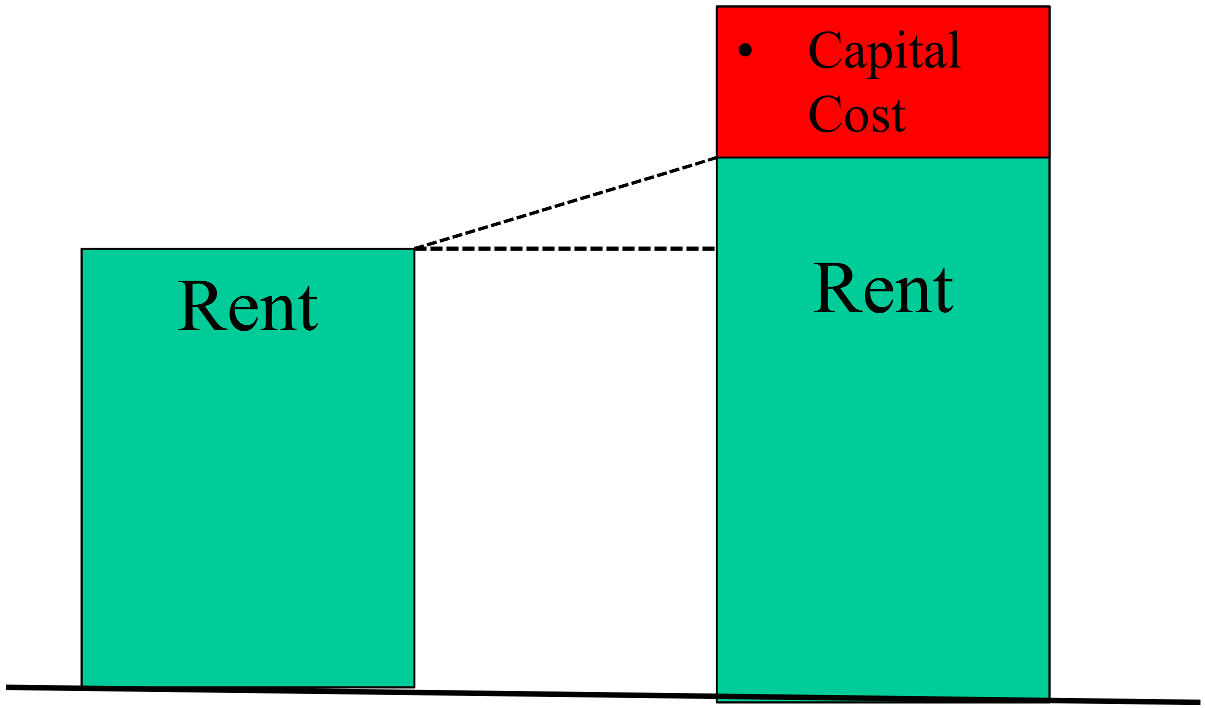
Hedonic Index of Office rent and Residential Rent



2.Theoretical Framework : Conditions for redevelopment

- Wheaton (1982) assumes that housing stock developed at one point in time exists at multiple time points.

- Existing Income
- Post Redevelopment Iincome



Empirical Analysis;

- Rosenthal and Helsley (1994) used an empirical analysis to verify Wheaton’s conditions for redevelopment.
- McGrath (2000) conducted an empirical analysis of commercial real estate by considering redevelopment conditions while taking into account soil pollution risks of land for redevelopment.

Panel Data Analysis: Office Use to Residential Use

- **Panel Data 1991→1996→2001**
- Bubble Bursting Period

- **Conversion from Office Use to Residential Use**
- a)- The conversion of offices into housing apparently occurs after landowners acknowledge land-use failures and closely examine profitability of land when it is used for office buildings and for housing.

- b)- We can ignore variables of urban planning constraints in the office-to-housing conversion case.

- c)- We can ignore land intensification costs.

Theoretical Framework:

Capital K and constant land area $\overline{L}$ are invested to produce a building with a total floor space of Q.

$$Q = F\left(K, \overline{L}\right) \qquad (1)$$

the landowner destroys the existing building at a cost of c per floor area. Given the discount rate i and the rent R^R for floor area Q, the maximized profit per land area for the new building for housing

$$\max_K r^R = \frac{R^R F\left(K, \overline{L}\right) - iK - c\overline{Q}}{\overline{L}} \qquad (2)$$

Conditions for Redevelopment

the production function: $F\left(K, \overline{L}\right) = AK^{\alpha}\overline{L}^{\beta}$

the optimization condition in Equation (1)

$$R^{R}\left(\partial F\left(K, \overline{L}\right)/\partial K\right) = i\,.$$

the redevelopment condition (2) can be rewritten as follows.

$$\Delta = \left(1 - \alpha\right)R^{R}Q - \left(R^{C} + c\right)\overline{Q} \geq 0 \qquad (3)$$

Econometric Model

- the binary choice model with panel data

$$\tilde{\Psi}_{it} = \gamma\Delta_{it} + u_{it}$$
$$u_{it} = \delta + \mu_i + \varepsilon_{it} \qquad i = 1, 2, \cdots, n \quad t = 1, 2$$

u_{it} : error component
δ : coefficient of the common constant term
μ_i : each group's random effect
ε_{it} : random variable

$\tilde{\Psi}_{it} > 0 \ (\Psi_{it} = 1)$ → redevelopment

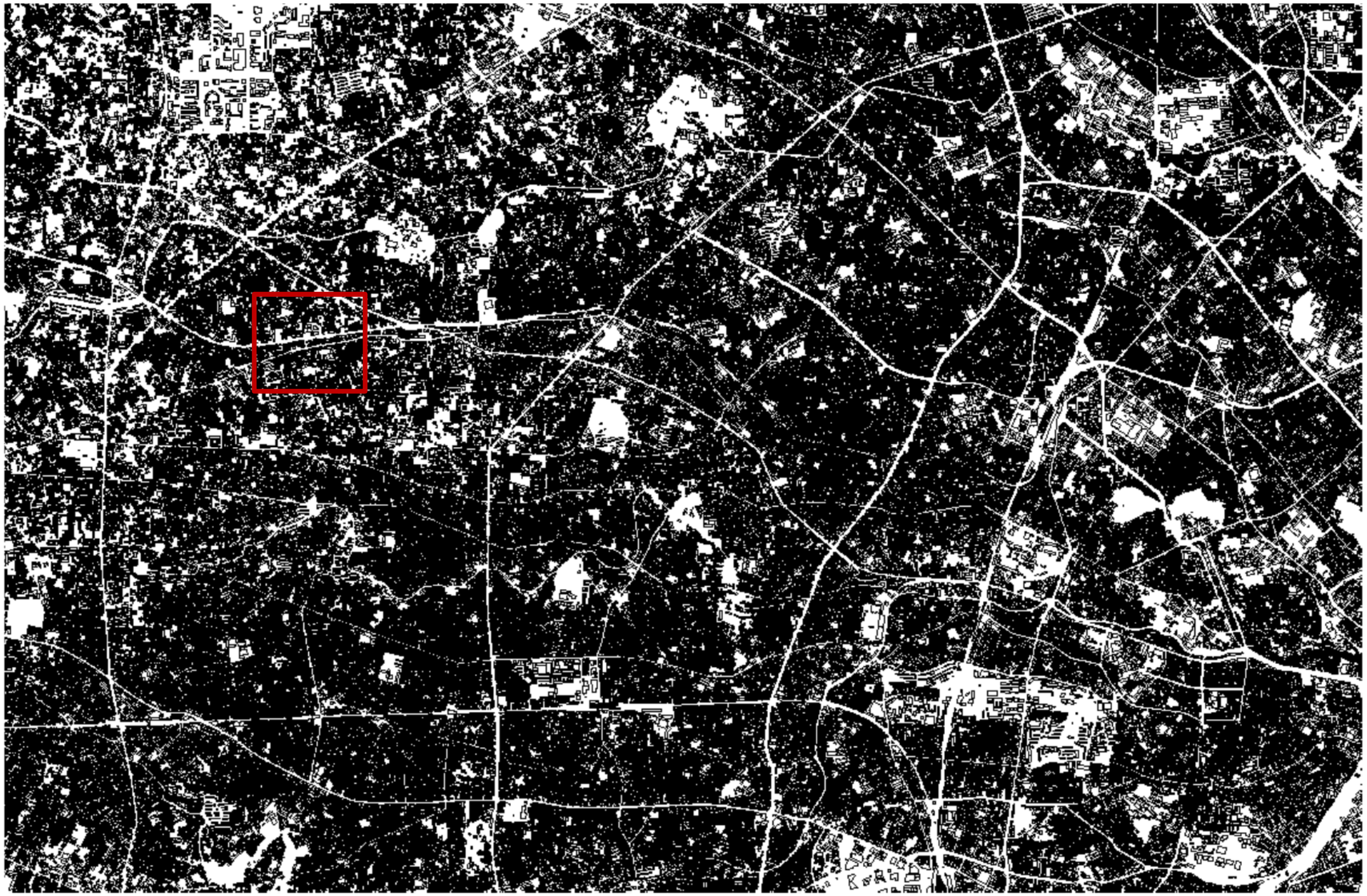
$\tilde{\Psi}_{it} \leq 0 \ (\Psi_{it} = 0)$ → continue with the present use

the redevelopment probability**:**

$$\Pr(\Psi_{it} = 1) = \Pr\left(\tilde{\Psi}_{it} > 0\right) = \Pr(\varepsilon_{it} > -\gamma\Delta_{it} - \delta - \mu_i) = \Phi(\gamma\Delta_{it} + \delta + \mu_i)$$

3. Data

- **Land uses and use conversions**
- **There is about 1.7 million(1,665,152) buildings**



3. Data

- **Land uses and use conversions**

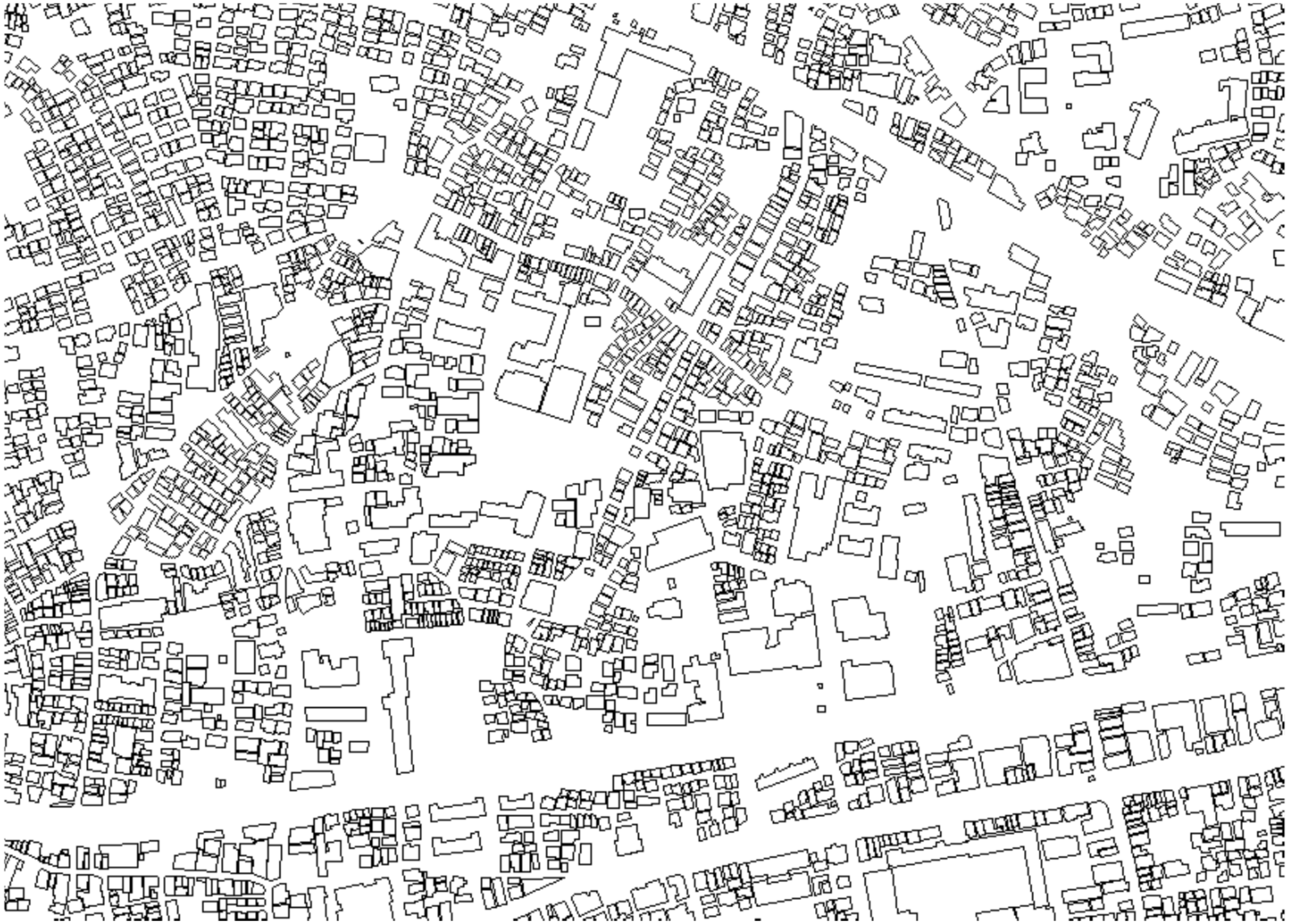


Figure 1. Office Buildings (1991) = 40,516

Of the 40,516 office buildings that existed in 1991, 2,607 were redeveloped or converted into housing by 1996, with the remaining 37,909 buildings used still for offices.

Of the office buildings that existed in 1996, 3,576 were redeveloped or converted into housing by 2001.
The remaining 36,940 office buildings remained as offices.

Table 2. Descriptive Statistics of Office and Housing Rent Data

	Office		Housing	
	Average	Standard deviation	Average	Standard deviation
Rent (yen/m^2)	4,851.48	1,925.12	3,248.26	824.9
Contractual space (m^2)	264.02	309.87	41.03	20.63
Distance to Tokyo centre (minutes)	12.46	6.25	10.53	7.17
Number of years after construction (years)	16.19	10.29	9.26	7.28
Distance to station (minutes)	4.13	2.91	6.76	3.89
Total floor space (m^2)	3,426.36	4,520.41	–	–
Number of observations=	13,147		488,348	

4. Estimation Results

- **4.1. Rent functions for office and housing uses**
- -Hedonic Equations

- **4.2. Condition for profit gaps**

- **4.3. Random probit model estimation**
- -Floor Space Production Function
- - Random probit model

Table 5. Office and Housing Rent Function Estimation Results

Method of Estimation	OLS			
Dependent Variable	*OR*: Rent of Office (in log)		*RC*: Rent of Condominiur	
Property Characteristics (in log)	Coefficient	t-value	Coefficient	t-value
Constant	8.374	181.483	0.253	–24.999
FS: Contractual space	0.19	59.102	–0.197	–141.297
BY: Number of years after construction	–0.093	–24.174	–0.070	–259.324
WK: Distance to nearest station	–0.219	–46.556	–0.034	–70.827
ACC: Time distance to Tokyo centre	–0.112	–25.362	–0.066	–117.539
TA: Total floor space	0.051	16.932	–	–
SRC: SRC building dummy	0.199	34.02	0.013	29.494
Ward (city) Dummy	Yes		Yes	
Railway/Subway Line Dummy	Yes		Yes	
Time Dummy	Yes		Yes	
Adjusted R square=	0.608		0.758	
Number of observations=	13,147		488,348	

Figure 2. Spatial Distribution (Housing rent > Office rent):1995

2.33%
(40,516Buildings)

Figure 3. Spatial Distribution (Housing rent > Office rent):2000

17.89%
(40,516Buildings)

Figure 3. Spatial Distribution (Housing rent > Office rent):2005

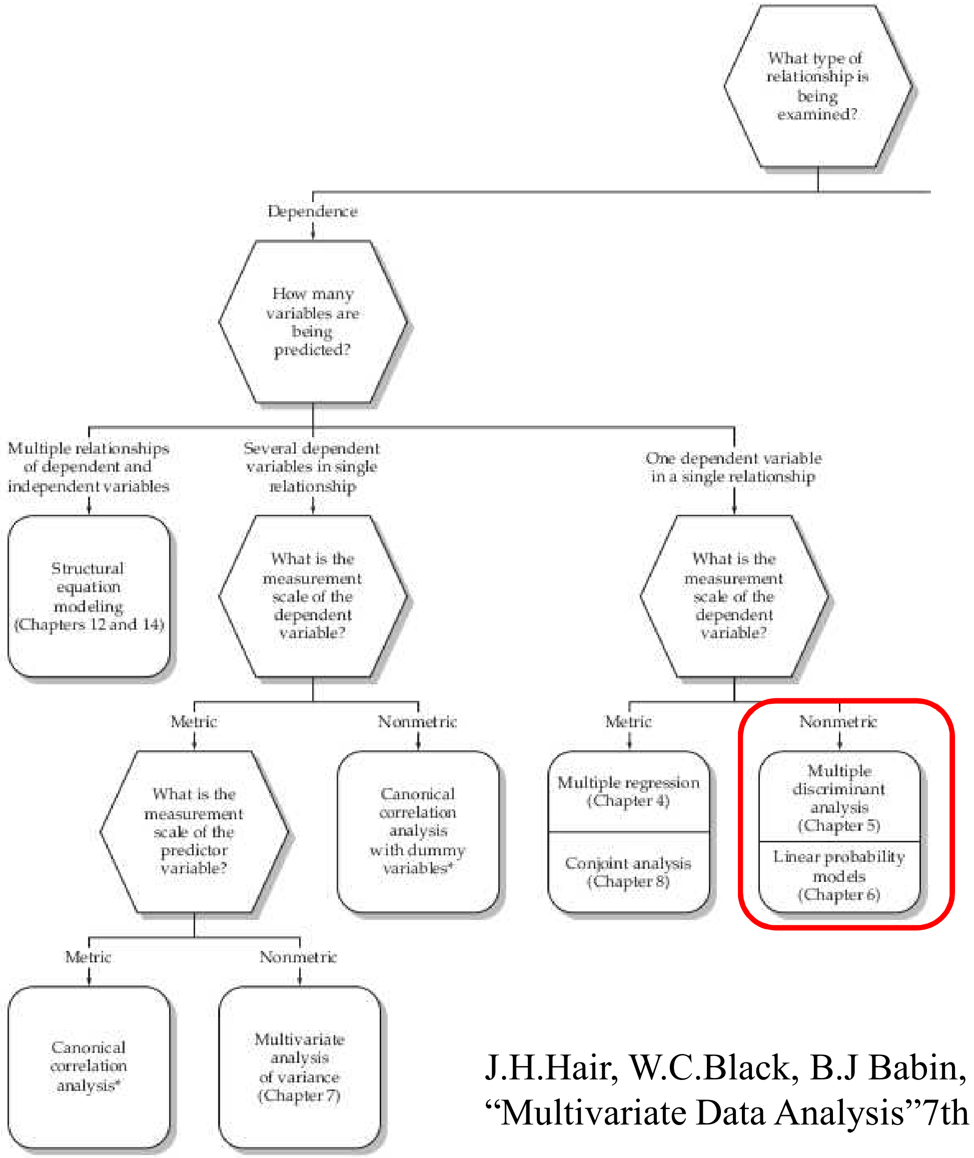
27.58%
(40,516Buildings)

Table6.Panel Data Outline

Year	Variable		Obs.	mean	std. dev.	min	max
1996	R^R	yen	40516	8399	2836	2837	26542
	R^C	yen	40516	4720	765	3018	6451
	Δ	million yen	40516	−10.91	55.11	−2712.34	−0.01
	Ψ	-	40516	0.06	0.25	0	1
	$\Delta(\Psi=1)$	million yen	2607	−2.25	7.07	−153.73	−0.01
	$\Delta(\Psi=0)$	million yen	37909	−11.51	56.90	−2712.34	−0.01
2001	R^R	yen	40516	6402	2162	2163	20232
	R^C	yen	40516	4808	779	3073	6570
	Δ	million yen	40516	−7.19	37.66	−1878.82	0.02
	Ψ	-	40516	0.09	0.28	0	1
	$\Delta(\Psi=1)$	million yen	3576	−1.44	4.21	−101.27	0.00
	$\Delta(\Psi=0)$	million yen	36940	−7.75	39.38	−1878.82	0.02

Note. R^R is the housing rent, R^C is the office rent, Δ is the difference in income Eq.(b)

Selecting a Multivariate Technique



J.H.Hair, W.C.Black, B.J Babin, R.H Anderson (2010), "Multivariate Data Analysis"7th edition.

Table 7. Probit Estimation of Redevelopment Probability

	Total	Region 1	Region 2	Region 3
Δ	0.3181	0.0576	0.4447	0.3407
	(0.0093)	(0.0058)	(0.0250)	(0.0219)
Constant	−13.5617	−5.7765	−9.3597	−9.7961
	(0.4317)	(0.1630)	(0.5139)	(0.6578)
σ	10.5011	2.9883	7.6478	8.0016
	(0.3327)	(0.0903)	(0.4046)	(0.4998)
ρ	0.9910	0.8993	0.9832	0.9846
	(0.0006)	(0.0055)	(0.0017)	(0.0019)
Number of obs.	81032	30110	19898	30468
Individual Number of groups	40516	15055	9949	15234
Wald (chi squared)	1160.1	98.8	315.3	242.8
	[.000]	[.000]	[.000]	[.000]
Log likelihood	−15071.5	−2567.9	−3792.0	−8043.3

Note. Standard errors are presented in parentheses. The dependent variable in the probit equals one if the parcel is redeveloped, zero if parcel remains in its current use. ρ is a correlation coefficient of random effect. Wald statistics test null hypothesis which all parameter is zero. Brackets [ ] are p-value for Wald test.

Table 7. Probit Estimation of Redevelopment Probability

	Total	Region 1	Region 2	Region 3
Δ	0.3181	0.0576	0.4447	0.3407
	(0.0093)	(0.0058)	(0.0250)	(0.0219)
Constant	−13.5617	−5.7765	−9.3597	−9.7961
	(0.4317)	(0.1630)	(0.5139)	(0.6578)
σ	10.5011	2.9883	7.6478	8.0016
	(0.3327)	(0.0903)	(0.4046)	(0.4998)
ρ	0.9910	0.8993	0.9832	0.9846
	(0.0006)	(0.0055)	(0.0017)	(0.0019)
Number of obs.	81032	30110	19898	30468
Individual Number of groups	40516	15055	9949	15234
Wald (chi squared)	1160.1	98.8	315.3	242.8
	[.000]	[.000]	[.000]	[.000]
Log likelihood	−15071.5	−2567.9	−3792.0	−8043.3

Note. Standard errors are presented in parentheses. The dependent variable in the probit equals one if the parcel is redeveloped, zero if parcel remains in its current use. ρ is a correlation coefficient of random effect. Wald statistics test null hypothesis which all parameter is zero. Brackets [ ] are p-value for Wald test.

5.Conclusion and Future work:

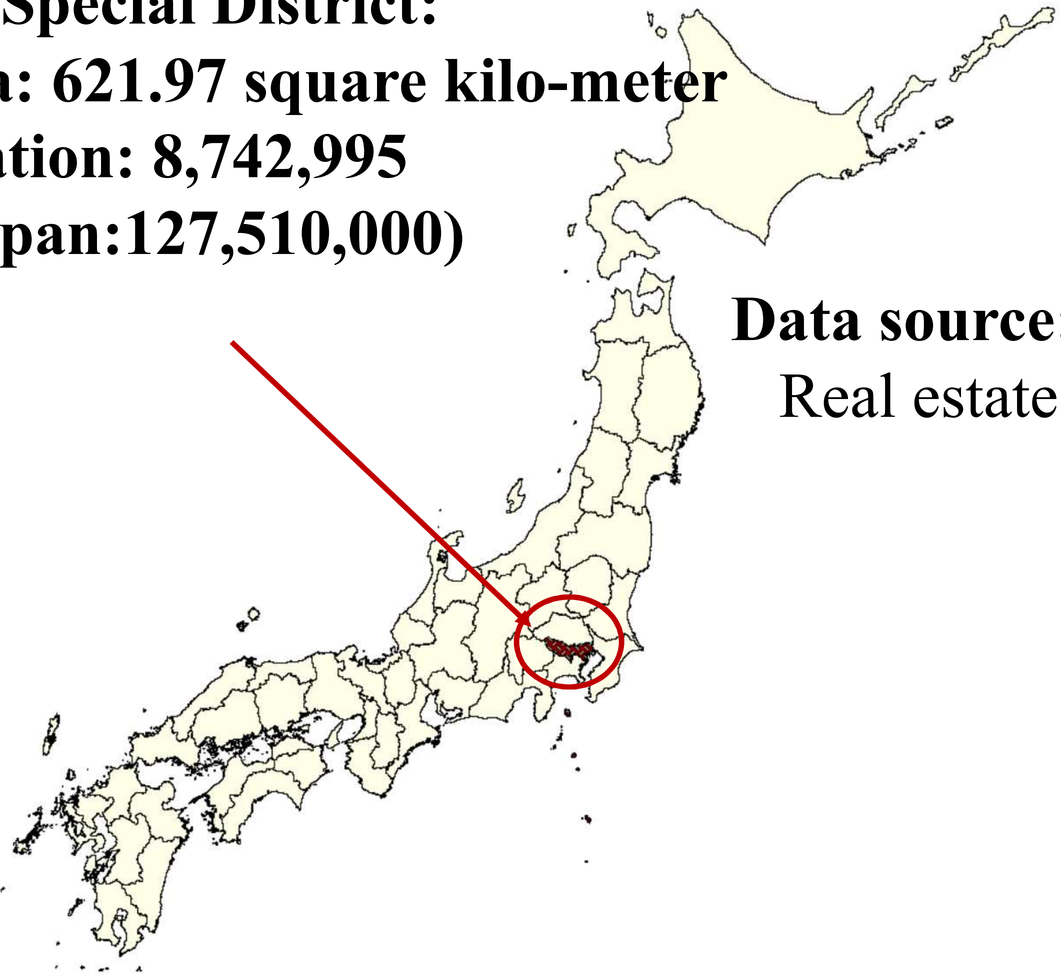
- This is the first empirical study using panel data to analyse conditions for redevelopment.

- We found that if random effects are used to control for individual characteristics of buildings, the redevelopment probability rises significantly when profit from land after redevelopment is expected to exceed that from present land uses. This increase is larger in the central part of a city.

- Limitations stem from the nature of Japanese data limited to the conversion of offices into housing. In the future, we may develop a model to generalize land-use conversion conditions.

Tokyo Special District:

Tokyo Special District:
Area: 621.97 square kilo-meter
Population: 8,742,995
(All Japan:127,510,000)



Data source:
Real estate advertisement magazine
(1986-2008: 23 years)

Panel Data Analysis: Office Use to Residential Use

- **Conversion from Office Use to Residential Use**

- a)- The conversion of offices into housing apparently occurs after landowners acknowledge land-use failures and closely examine profitability of land when it is used for office buildings and for housing.

- b)- We can ignore variables of urban planning constraints in the office-to-housing conversion case.

- c)- We can ignore land intensification costs.

Table 4. Floor Space Production Function

Cobb–Douglas Production Function: $$Q = AK^{\alpha}L^{\beta}$$

Q: stands for the total floor space, K: for construction costs and L :for the site area size

	coef.	t-value
Constant term	24.140	2.673
log K	0.390	10.704
log L	0.670	15.077
Annual trend	−0.011	−2.396
Ward dummy	Yes	
Adj. R^2	0.959	

Note: The annual trend indicates an estimated coefficient of the trend term representing the time of completion

References:

- Brett Lanz (2015), *Machine Learning with R*.

- Schen Liu, James McGree, Zengyuan Ge and Yang Xie (2016), *Computational and Statistical Methods for analyzing Big Data with Applications*.

- P. Newbold, W.L. Carlson, B. M. Thorne (2014), *Statistics for Business and Economics 8th edition*.

清水千弘 : Chihiro Shimizu, PhD

シンガポール国立大学不動産研究センター　教授

Professor, Institute of Real Estate Studies
National University of Singapore
21 Heng Mui Keng Terrace, #04-02
Singapore 119613
Tel: (65) 6601 4925 Fax: (65) 6774 1003
Email: cshimizu@nus.edu.sg
Website: www.ires.nus.edu.sg