CARF Working Paper

CARF-F-263

POWER LAWS IN REAL ESTATE PRICES DURING BUBBLE PERIODS

Takaaki Ohnishi
Canon Institute for Global Studies
Takayuki Mizuno
University of Tsukuba
Chihiro Shimizu
Reitaku University
Tsutomu Watanabe
The University of Tokyo

December 2011

CARF Working Papers can be downloaded without charge from:
http://www.carf.e.u-tokyo.ac.jp/workingpaper/index.cgi

POWER LAWS IN REAL ESTATE PRICES DURING BUBBLE PERIODS

TAKAAKI OHNISHI

The Canon Institute for Global Studies, 11th Floor, Shin Marunouchi Building
1-5-1 Marunouchi, Chiyoda-ku, Tokyo 100-6511, Japan
and
Graduate School of Economics, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
ohnishi.takaaki@canon-igs.org

TAKAYUKI MIZUNO

Faculty of Engineering, Information and Systems, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
mizuno@cs.tsukuba.ac.jp

CHIHIRO SHIMIZU

Faculty of Economics, Reitaku University
2-1-1 Hikarigaoka, Kashiwa-shi, Chiba 277-8686, Japan
cshimizu@reitaku-u.ac.jp

TSUTOMU WATANABE

Graduate School of Economics, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
watanabe@e.u-tokyo.ac.jp

How can we detect real estate bubbles? In this paper, we propose making use of information on the cross-sectional dispersion of real estate prices. During bubble periods, prices tend to go up considerably for some properties, but less so for others, so that price inequality across properties increases. In other words, a key characteristic of real estate bubbles is not the rapid price hike itself but a rise in price dispersion. Given this, the purpose of this paper is to examine whether developments in the dispersion in real estate prices can be used to detect bubbles in property markets as they arise, using data from Japan and the U.S. First, we show that the land 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 before and after the bubble period. Second, in the U.S. data we find that the tail of the house price distribution tends to be heavier in those states which experienced a housing bubble. We also provide evidence suggesting that the power-law tail observed during bubble periods arises due to the lack of price arbitrage across regions.

Keywords: Econophysics; Power law; Bubbles; House prices; Land prices; Price disper-

sion.

1. Introduction

Property market developments are of increasing importance to practitioners and policymakers alike. The financial crises of the past two decades have illustrated just how critical the health of this sector can be for maintaining financial stability. For example, the recent financial crisis in the U.S. first reared its head in the form of subprime loan problems. The financial crises in Japan and Scandinavian countries in the 1990s were both triggered by the collapse of bubbles in real estate markets. More recently, the rapid rise in real estate prices - often supported by a strong expansion in bank lending - in a number of emerging market economies has become a concern for policymakers. Given these experiences, it is critically important to understand the relationship between property markets, financial issues, and financial crises.

One of the most urgent tasks in this respect is the development of methodologies to detect bubbles in property markets, which is key to reliably forecasting potential financial crises and, if possible, mitigating them. Consider the case of residential property. In real estate economics, the "fundamental" value of a house is considered to be the present discounted value of current and future income flows resulting from renting the house to someone (see, for example, Ref. 1). Under normal circumstances, house prices remain close to their fundamental value due to price arbitrage. However, in some cases, arbitrage forces are not present, and prices therefore deviate substantially from their fundamental value. This is what is called a housing bubble. To detect housing bubbles defined in this way, reliable estimates of the fundamental value of properties are required. This, in turn, means we need to know market participants' expectations of rental prices in the coming years. However, such expectations are not observable and the fundamental value of properties therefore is rather difficult to estimate. Thus, it is next to impossible for researchers or policymakers to tell, when they observe a price hike, whether this is the result of a rise in fundamental property values or of some other factor(s).

An alternative approach to detecting real estate bubbles is to make use of information on the cross-sectional dispersion of real estate prices. A wide-held believed is that the prices of *all* properties rise equally during a bubble period. However, this is not the case. What happens instead is that prices go up considerably for some properties, but less so for others, and as a result, price inequality across properties increases during a bubble. In other words, a key characteristic of real estate bubbles is not the rapid price hike itself but a rise in price dispersion. In fact, examining property prices in Tokyo during the bubble in the late 1980s, Ref. 2 found that price dispersion increased substantially; more specifically, the cross-sectional distribution of house prices had a power-law tail during that period.

The purpose of this paper is twofold. First, we provide further evidence that price dispersion increases during real estate bubbles. The first episode we examine is the

land price bubble in Tokyo, while the second is the recent housing bubble in the U.S. We carefully examine changes in the price distribution, especially changes in its tail part, in each case. Second, we investigate why the price distribution has a power law tail during bubble periods. Our hypothesis is that the power-law tail arises due to the mixture of lognormal distributions with different mean and variance; namely, prices are lognormally distributed for each of the sub-areas, but the distribution in each sub-area has a different mean and variance. To test this hypothesis, focusing on the land price bubble in Tokyo, we divide the entire sample area into sub-areas to see whether the distribution of prices in each sub-area is close to a lognormal, as well as whether the mean and variance are different across sub-areas.

The rest of the paper is organized as follows. Section 2 provides a review of the literature on cross-sectional price distributions during real estate bubbles. Section 3 provides new evidence that land price dispersion increased in Tokyo during the bubble period. Section 4 then examines the house price distribution in the U.S. Finally, Section 5 concludes the paper.

2. Literature Review

2.1. *How to detect real estate bubbles*

What is the shape of the cross-sectional distribution of real estate prices? How does it evolve over time? These are one of the most important research questions in real estate economics as well as in econophysics. The first paper that examines the shape of the distribution of real estate prices is Ref. 3, which shows that the cross-sectional distribution of land prices in Japan has a power law tail. The presence of a power law tail of house price distributions is confirmed for Taiwan by Ref. 4 and for U.K. by Ref. 5. The other related works include Refs. 6–10.

Houses and land constitute an important part of individual wealth. Given that the distribution of wealth owned by individuals has a power law tail, their results may not be so surprising. More importantly, however, previous studies have shown that the distribution of real estate prices change over time. For example, in real estate economics, Ref. 11 employs the U.S. house price data to ask whether the change in the price distribution comes from a change in the distribution of house characteristics such as size, location, etc, or from a change in the implicit prices associated with those characteristics. However, the most important contributions in this respect are Refs. 3, 6, which have found that the exponent associated with the power law tail of the Japan's land price distribution exhibited a substantial fall during the bubble period (i.e., the upper tail was heavier during the bubble period), and that the exponent was very close to unity just before the bubble burst.

The finding by Refs. 3, 6 gives us a hint on a new method for detecting real estate bubbles. The traditional method is based on the analysis of the time series of a real estate price index (e.g., the Case-Shiller index for house prices). The recent examples include Refs. 12–14, which seek to identify a housing bubble in the U.S. by examining whether price acceleration was faster than exponential or not. In

4   *T. Ohnishi et al.*

contrast, the finding by Refs. 3, 6 suggests to make use of price information at the cross-sectional dimension, rather than at the time series dimension, so as to detect real estate bubbles.

To proceed further along this line of research, we need to address the following questions. First, why does the distribution of real estate prices have a power law tail? Second, what is the mechanism through which the tail of the distribution becomes heavy during real estate bubbles? Ref. 2 addresses the first question, arguing that the observed power law tail stems from the mixture of different distributions. The land price data used by Refs. 3, 6–9 are collected from various sites in Japan, including sites located in large cities like Tokyo, as well as sites located in rural areas with much lower population densities. As one can easily imagine, the price observations in large cities and those in rural areas are from different distributions.

This suggests that one should look at the distribution of prices collected from a particular area, in which price observations come from a single distribution. Then, the next question to be addressed is what would be the shape of the price distribution for that particular area. Ref. 2 argues that it should be a lognormal because of the following reason. The standard theory in real estate economics is based on the idea that the value of a house is the sum of the values of attributes such as its floor space, its age, the commuting time to the nearest station, and so on. That is, the log price of a house, $\log P$, is the sum of K components:

$$\log P = \sum_{k=1}^{K} x_k \tag{1}$$

where x_k is a random variable representing the value of k-th attribute of a house. According to the central limit theorem, the sum of the random variables on the right hand side converges to a normal distribution as K goes to infinity, if the average of the variances of x_k converges to a finite constant and

$$\lim_{K \to \infty} \frac{\max_{k \leq K} \{s_k\}}{K \bar{s}_K} = 0 \tag{2}$$

where s_k is the standard deviation of x_k and $\bar{s}_K$ is the square root of the average of the variances of x_k. Note that Eq. (2) requires that no single term dominates the average variance. Therefore, if one eliminates the contribution of a dominant component to prices, the price distribution should follow a lognormal distribution.

To test this prediction, Ref. 2 constructs a dataset containing house prices only in Tokyo, and eliminates the contribution of the size of a house to its price. It was found that the distribution of size adjusted prices in Tokyo is quite close to a lognormal distribution for most of the sample period, which is in sharp contrast with the results in Refs. 3, 6, 8. At the same time, it was found that, during the bubble period, the price distribution still deviates from a lognormal, exhibiting a heavy upper tail. This result is the same as the ones in Refs. 3, 6, 8, and in this
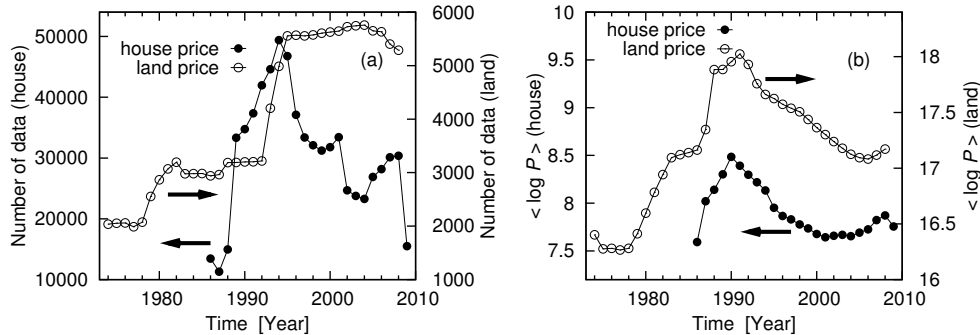
Fig. 1.   The number of observations (a) and the mean of log prices (b) for houses and land. House prices are in ten thousand yen, while land prices are in yen.

sense, the question regarding the mechanism behind the presence of a heavy tail of the price distribution during bubble periods remains unsolved.

2.2. *Changes in the distribution of house prices over time*

In the rest of this section, we provide a brief review of the main results of Ref. 2. Let us start with looking at how the cross-sectional house price distribution evolved over time. In Ref. 2, we investigated changes in the distribution of prices for condominiums traded in Tokyo. The dataset we used is compiled from individual listings in a widely circulated real estate advertisement magazine, which is published on a weekly basis by Recruit Co., Ltd., one of the largest vendors of residential lettings information in Japan. The dataset covers the Greater Tokyo Area for the period 1986 to 2009, including the bubble period in the late 1980s and its collapse in the first half of the 1990s. It contains 724,416 records, each of which consists of the selling price of a house, as well as the attributes of a house, including its size, location, etc. The number of observations for each year is more than 10 thousand, as shown by the closed circle in Fig. 1(a).

Fig. 1(b) shows that the mean of log house prices, represented by the closed circle, exhibited a rapid rise in the latter half of the 1980s, reaching its peak in 1990. Then it turned to a gradual decline, which continued until it hit the bottom in 2001. This is a famous episode of real estate bubble in Tokyo. To see how the house price distribution changed during this period, we show in Fig. 2 the cumulative distribution function (CDF) of house prices for each year in 1986 to 2009. The vertical axis represents the fraction of houses whose prices are greater than the value indicated on the horizontal axis. The figure shows that the tail of the CDF in each year has a power law tail; namely,

$$P(> P) \propto P^{-\alpha} \qquad (3)$$

where $P(> P)$ is the cumulative probability, and α is a power law exponent. This result looks similar to the findings in previous studies, including Refs. 3, 6, 8. Note
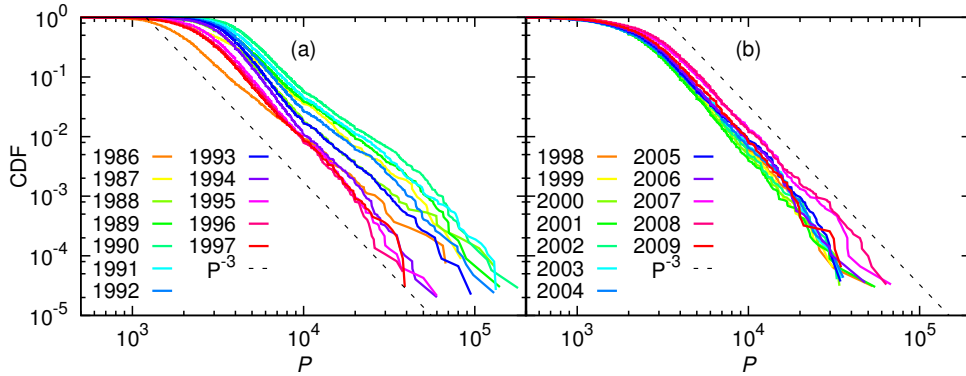
Fig. 2.   Cumulative distribution function of house prices for each year in 1986-1997 (a) and in 1998-2009 (b). House prices are in ten thousand yen. Colors represent different years. The dotted reference line represents power law scaling with an exponent of $\alpha = 3$.
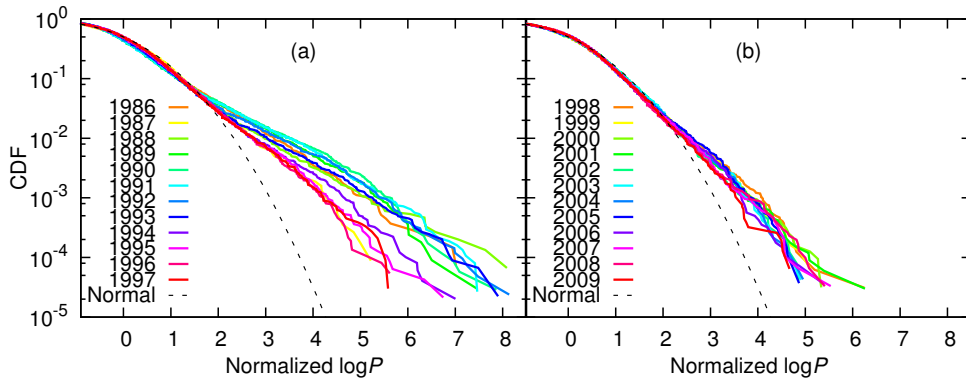

Fig. 3.   Cumulative distribution function of normalized log prices for each year in 1986-1997 (a) and in 1998-2009 (b). Colors represent different years. The dotted line represents the standard normal distribution.

that the CDF shifted to the right during the bubble period (i.e., 1986-1990), and it shifted back to the left after the collapse of the bubble (i.e., 1991-1995). However, the shape of the distribution, especially the power law exponent, seems to remain unchanged across years.

To see how far the price distributions are away from a lognormal, we calculate a normalized log price, which is defined by

$$\text{Normalized} \log P = (\log P - \langle \log P \rangle) / \left\langle (\log P - \langle \log P \rangle)^2 \right\rangle_+, \tag{4}$$

where $\langle \cdot \rangle$ denotes the average over all observations, while $\langle \cdot \rangle_+$ is the average over observations only with $\log P > \langle \log P \rangle$. Note that the denominator of the right hand side of Eq. (4) differs from a usual definition of the standard deviation in that we discard observations below the average. We employ this definition to accurately
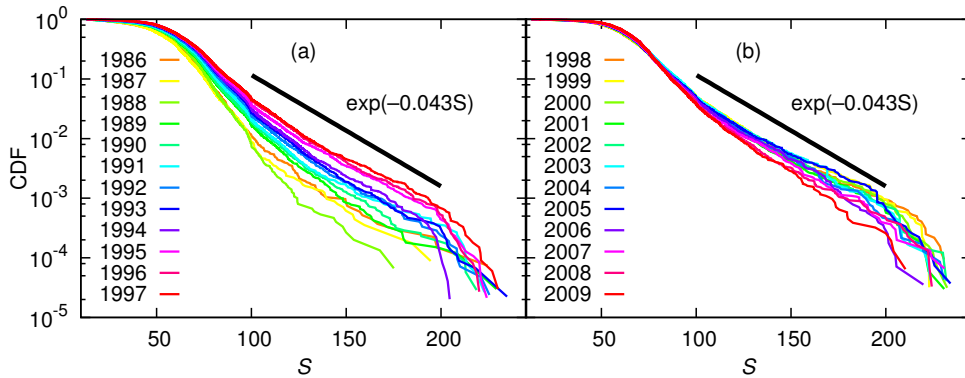
Fig. 4.   Cumulative distribution function of house sizes for each year in 1986-1997 (a) and in 1998-2009 (b). Colors represent different years. The reference line is a straight line in a semi-log plot, with a slope of -0.043, which is estimated by an OLS regression using all observations with $100 < S < 200$.

evaluate the upper tail of the distribution, which may be different from the lower tail of the distribution (i.e., the distribution may be asymmetric). Fig. 3 shows the CDFs for the normalized log prices, indicating that the upper tail of the CDF in each year deviates significantly from that of the standard normal distribution, which is shown by the dotted line.

2.3. *Distributions of size adjusted prices*

Why does the price distribution deviate from a normal distribution even when we focus on prices only in Tokyo? Ref. 2 argues that this occurs because there exists a dominant determinant of house prices, and therefore the condition represented by Eq. (2) is violated. In real estate economics, it is empirically shown that there are three important determinants of property prices: the land area of the property; the livable floor space area of the structure; and the location of the property. Based on this, Ref. 2 eliminates the contribution of the size of a house to its price.

Fig. 4 shows the CDFs of house sizes S, which is measured in terms of floor space (square meters) of a condominium. We see that the CDF in each year follows an exponential distribution; that is,

$$P(> S) \propto \exp(-\lambda S) \tag{5}$$

for $S > 100$, where λ is an exponent. The exponent λ is almost identical across years, and the estimate of λ, which is obtained by applying an ordinary least squares (OLS) to the observations with $100 < S < 200$, is 0.043.

Next, we examine the relationship between the price of a house and its size. Fig. 5 shows the mean of log prices conditional on the house size S, namely, $\langle \log P | S \rangle$. We see that there exists a positive linear relationship between the log price and the size
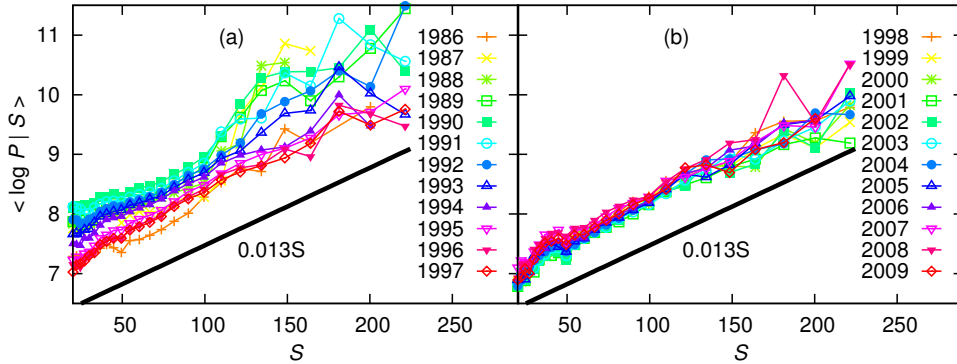
8   *T. Ohnishi et al.*

Fig. 5.   The mean of $\log P$ conditional on S for each year in 1986-1997 (a) and in 1998-2009 (b). Colors represent different years. The reference line is the straight line with a slope of 0.013, which is estimated by an OLS regression line using all observations.

of the form

$$\log P = aS + Q \tag{6}$$

where a is a positive coefficient and Q is a random variable to which we refer as a size-adjusted price. We see in Fig. 5 that the coefficient a is identical across years and it is about 0.013. Note that the relationship is not linear but exponential. This may be easy to understand if one recognizes that high end houses are equipped with luxurious facilities per square meter (such as a bath room with whirlpool bath), which obviously push up the costs.

The fact that P follows a power law distribution with an exponent of α implies that $\log P$ follows an exponential distribution with an exponent of α. On the other hand, the fact that S follows an exponential distribution with an exponent of λ implies that aS follows an exponential distribution with an exponent of λ/a. The estimate of α is 3, while the estimate of λ/a is 3.3 (= 0.043/0.013). Therefore, each of the left hand side of Eq. (6) and the first term on the right hand side follows an exponential distribution with an almost identical exponent. Given this fact, it can be shown that each side of Eq. (6) follows a distribution with an identical upper tail if the size-adjusted price Q follows a normal distribution or an exponential distribution with an exponent below 3.

To see whether this holds or not, we calculate Q for individual houses. The CDFs of Q are shown in Fig. 6. Note that the estimates of Q are normalized in the same way as we did with $\log P$. We see in Fig. 6 that, for 1998-2008, the CDFs of normalized Q are all identical to the CDF of the standard normal distribution, confirming that the above prediction is correct. However, for 1986-1997, the upper tail of the Q distribution is considerably heavier than that of the normal distribution. The deviation from the normal distribution is particularly large for 1988, 1992, 1993, and 1994, indicating that the deviation tends to be large during the bubble period, as well as the early stage of its collapse. In this sense, we still do not understand
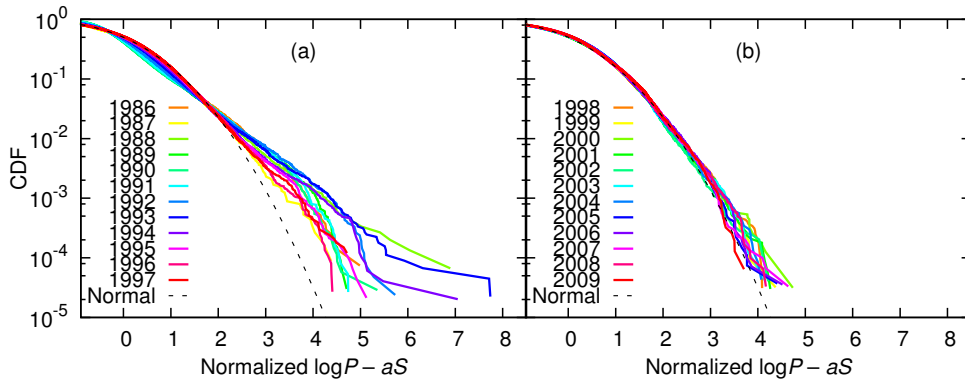
Fig. 6.   Cumulative distribution function of normalized size-adjusted prices for each year in 1986-1997 (a) and in 1998-2009 (b). Colors represent different years. The dotted line represents the cumulative distribution function for the standard normal distribution.

why price distributions deviate from the normal distribution during bubble periods. In the next section, we address this issue in more detail.

3. Application to the Japanese Land Price Data

3.1. *Land price data*

In this section, we examine changes in the distribution of land prices in Tokyo. This dataset contains land prices for about 30 thousand sites across Japan. The price of each site is assessed by the Land Appraisal Committee once a year and made available to the public. The dataset contains information about various attributes of land, including its address, size, shape, types of land use (residential, commercial, and industrial), etc. The land prices contained in this dataset are not actual transaction prices but appraisal prices, which may be different from transaction prices. This is clearly a potential shortcoming of this dataset compared to the house price data used in the previous section. However, this dataset covers the period 1974-2008, much longer than the period covered by the house price data used in the previous section. Most importantly, this dataset contains observations before and after the real estate bubble in Tokyo, thus allowing us to compare the price distribution during a bubble with the price distributions before or after the bubble.

This dataset was used by previous studies including Refs. 3, 6–9. We depart from the previous studies in some important respects. First, we focus on lands located in the Greater Tokyo Area. Second, we focus on lands of residential use. In this way, we intentionally pick up homogeneous lands. Finally, we focus on lands with its size smaller than 300 square meters because of the following reason. Fig. 7 shows the CDFs of land sizes S for those lands which are located in the Greater Tokyo Area and are of residential use. We see that the CDFs are exponentially distributed with an identical exponent, but they exhibit a sudden drop at $S = 330$. This suggests the possibility of sampling error by the data provider, although we are not quite sure
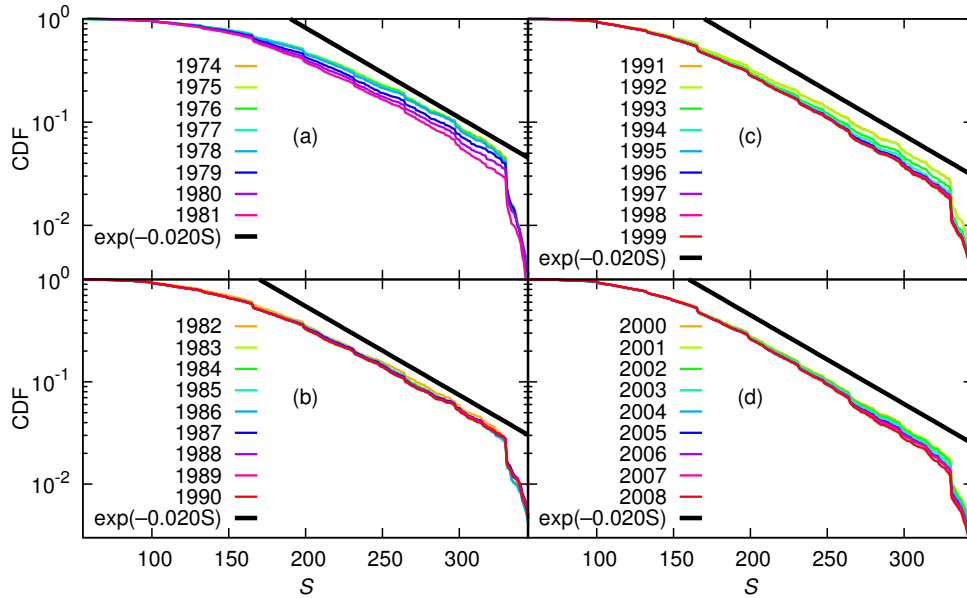
10   *T. Ohnishi et al.*

Fig. 7.   Cumulative distribution function of land sizes for each year in 1974-1981 (a), in 1982-1990 (b), in 1991-1999 (c), and in 2000 -2008 (d). Colors represent different years. The reference line is a straight line in a semi-log plot, with a slope of -0.020, which is estimated by an OLS regression using all observations with $80 < S < 330$.

why this happens. To make sure that our analysis is not affected by this error, we discard the observations with land sizes above 300 square meters. The remaining number of observations is $2,000$ to $5,000$ per year, as shown by the open circle in Fig. 1(a).

3.2. *Changes in the distribution of land prices over time*

Fig. 1(b) shows fluctuations in the mean of log prices for land. The land price started to rise in 1979 and exhibited a discrete jump in 1987, which was followed by a gradual increase that continued until the price reached its peak in 1991. Then the price started to fall in 1992, and it continued to decline gradually until 2006, when the Japanese economy started to recover from a long stagnation.

To see how the distribution of land prices evolved over time during the bubble period, we show in Fig. 8 the CDFs of land prices. The distributions have a power law tail for 1987 to 1994. The power law exponent in 1987-1990 is around 2, but it declines gradually over time from 1991 to 1994. These results are perfectly consistent with the results on the house price distribution presented in the previous section. A more interesting finding is that the distribution of land prices did not have a heavy tail before the bubble took place in the second half of the 1980s. Note that we have no comparable result for house prices due to data limitation. This new result is particularly informative because we are more confident that the price distribution
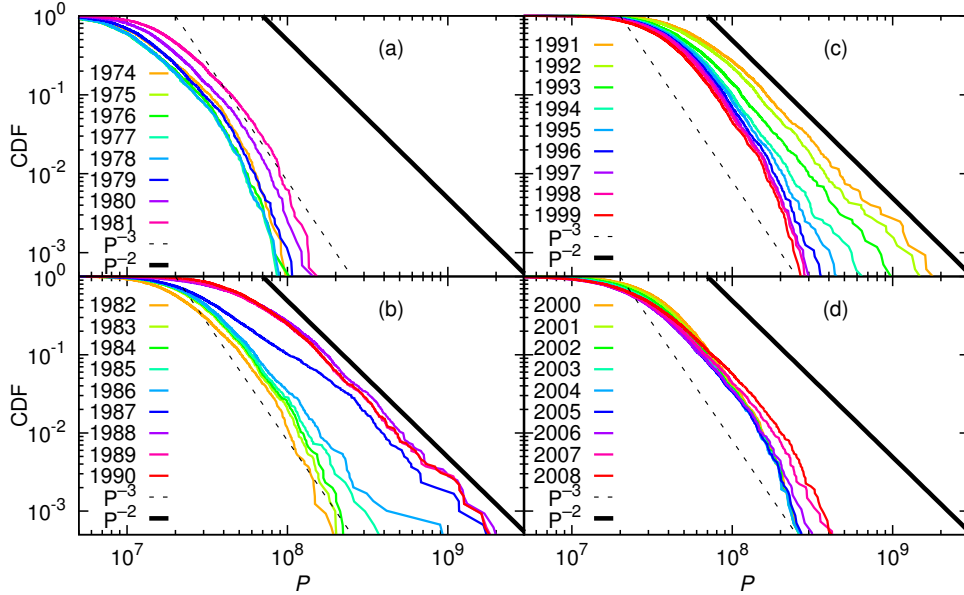
Fig. 8. Cumulative distribution function of land prices for each year in 1974 -1981 (a), in 1982-1990 (b), in 1991-1999 (c), and in 2000-2008 (d). Land prices are in yen. Colors represent different years. The two reference lines represent power law scaling with an exponent of $\alpha = 2$ (the solid reference line) and $\alpha = 3$ (the dotted reference line).

tends to have a heavy tail *only* during the bubble period. The CDFs of normalized land prices in Fig. 9 show this point clearer; namely, the CDFs are almost identical to the CDF of the standard normal distribution except the period of bubble and its collapse (i.e., 1987-1995).

It should be noted that the CDFs in Fig. 9 are for prices that are normalized, but they are not size adjusted. This suggests that the role of size adjustment is much smaller for land prices than for house prices. To investigate this in more detail, we show in Fig. 10 the mean of log prices for land conditional on the land size. There is a linear relationship between the log price and the size, which is similar to the result for house prices presented in Fig. 5. However, the estimated slope for land is 0.00501, which is much smaller than the slope for houses (the corresponding estimate was 0.013). This result suggests that the size of land may not be a dominantly important determinant of land prices. Probably this reflects the fact that the building-to-land ratio, whose ceiling is set by the regulator, differs area by area even within the Greater Tokyo Area (for example, it is high in the downtown while it is low in the suburbs of Tokyo with high end residential properties). To confirm this result, we show in Fig. 11 the CDFs of size adjusted land prices. The CDFs are basically identical to those without size adjustment. Based on these results, in what follows, we use land prices that are not size adjusted.
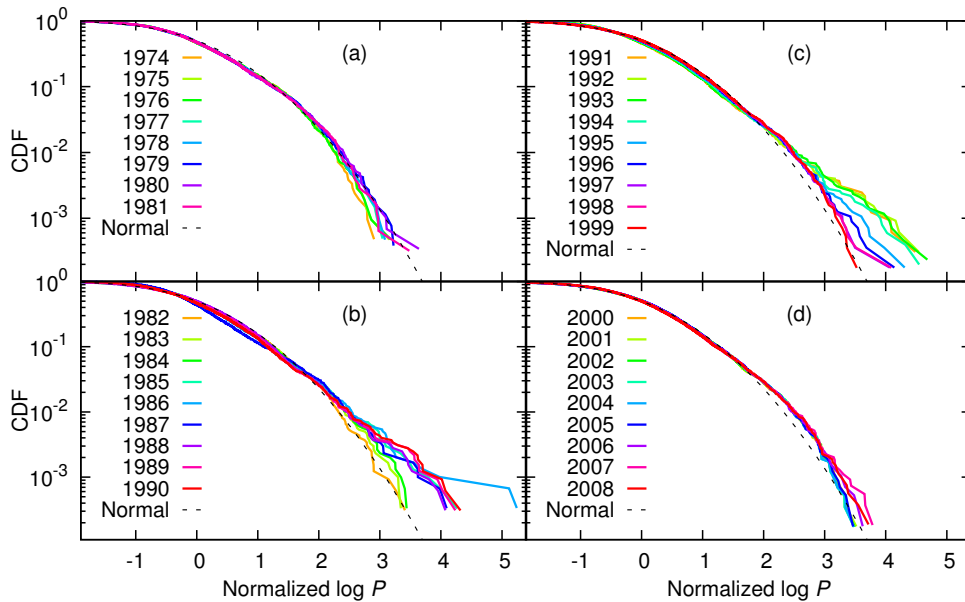
12   *T. Ohnishi et al.*

Fig. 9.   Cumulative distribution function of normalized log land prices for each year in 1974-1981 (a), in 1982-1990 (b), in 1991-1999 (c), and in 2000-2008 (d). Colors represent different years. The dotted line represents the cumulative distribution function of the standard normal distribution.

3.3. *Land price distributions within and across sub-areas*

Why did the land price distribution in the Greater Tokyo Area have a power law tail only during the bubble period? Our hypothesis is that the power law tail arises due to the mixture of lognormal distributions with different mean and variance. Specifically, land prices are lognormally distributed for each of the sub-areas ("regions") in the entire Greater Tokyo Area, but the distribution in each sub-area has a different mean and variance. In other words, prices are homogeneous within a region, but heterogeneous across regions. It is this spatial heterogeneity that produces a power law tail. Note that our decision in the previous subsection to focus on lands located only in the Greater Tokyo Area was based on this argument. It turned out that this strategy works well for the "peace" time, but not for the bubble period, suggesting that the size of a region within which prices are homogeneous would be smaller during the bubble period.

To test the above hypothesis, we divide the entire area into regions, each of which contains n lands. Then we apply the following algorithm:

  (i) Choose randomly a land as the seed for a region.
 (ii) Pick $(n-1)$ nearest neighbor lands and label them as the lands belonging to the region defined in step (i).
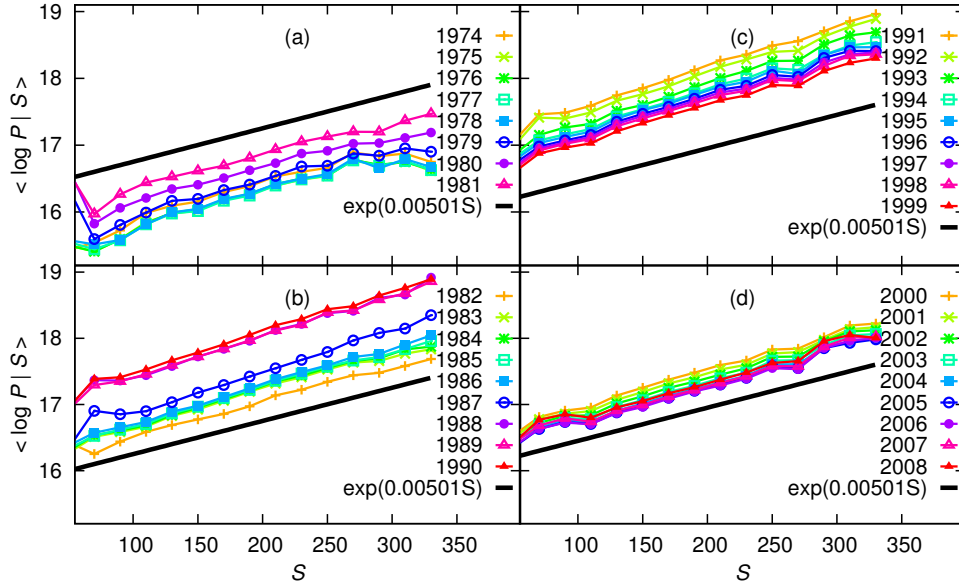(iii) Repeat steps (i) and (ii) until all lands are labeled.

Fig. 10.   The mean of $\log P$ conditional on S for each year in 1974-1981 (a), in 1982-1990 (b), in 1991-1999 (c), and in 2000-2008 (d). Colors represent different years. The slope of the reference line, 0.00501, is estimated by applying an OLS to all observations.

The result of this exercise may depend on the choice of a land in step (i). To make sure that it is not the case, we repeat the above algorithm 128 times. Fig. 12 shows the CDF of normalized land prices for different region sizes. Fig. 12(a) presents the CDF of normalized prices in 1976, which is a typical year in the pre-bubble period, while Fig. 12(c) shows the CDF of normalized prices in 1987, a typical year during the bubble period. Both panels show five different CDFs, each of which corresponds to a different region size. Note that the black line, labeled by $n = +\infty$, is the CDF for the entire area, which is the same as the one in Fig. 9.

Starting with Fig. 12(a), we see that the distributions are all very close to a normal distribution when n is 64 or larger than that. This implies that, in the pre-bubble period, land prices in a region of the small size, say $n = 64$, are lognormally distributed, and the mean and variance of the distribution in each region is identical, consequently land prices are lognormally distributed even in a region of the larger size. In this sense, land prices are spatially homogeneous within each region, as well as across regions. Turning to Fig. 12(c), we see that, for $n = +\infty$, the CDF is far away from the CDF of the standard normal distribution, but the CDF comes closer to the Gaussian CDF as n becomes smaller, and finally converges to it when $n = 64$. This occurs because land prices in a region of the small size, say $n = 64$, are lognormally distributed even during the bubble period, but the mean and variance is not identical across regions, so that land prices deviate from the lognormal in a region of the larger size. That is, land prices are spatially heterogeneous during the
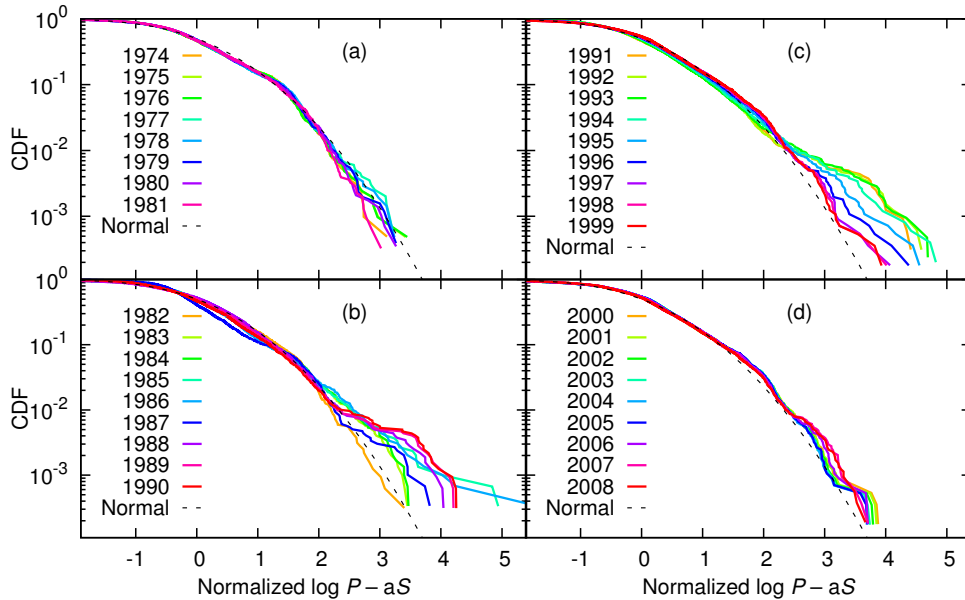
14   *T. Ohnishi et al.*

Fig. 11.   Cumulative distribution function of normalized size-adjusted land prices for each year in 1974-1981 (a), in 1982-1990 (b), in 1991-1999 (c), and in 2000-2008 (d). Colors represent different years. The dotted line represents the cumulative distribution function of the standard normal distribution.

bubble period, and it is this heterogeneity that produces a power law tail.

It should be noted that the CDF for the case of $n = 16$ in Fig. 12(a) is located inside the CDF of the standard normal distribution. This occurs because log prices, when they are normalized, never exceed $\sqrt{n-1}$. This finite size effect plays a dominantly important role when n is 16 or smaller than that, but it may be the case that the finite size effect plays some role even when n is greater than 16. If that is the case, the convergence of the land price distribution to the normal distribution observed for 1987 may be an artificial fact caused by the finite size effect. To investigate this possibility, we randomly replace the address of a land with the address of another land, thereby removing the possible relationship between the prices of lands which are located closely. Using the randomly reshuffled data constructed in this way, we examine how the CDF of normalized log prices changes depending on the region size.

The result of this exercise is presented in Figs. 12(b) and (d). Note that we produce 24 sets of randomly reshuffled observations, to each of which we apply the same grouping method as before. The CDFs in Fig. 12(b), which is for 1976, is almost the same as the CDFs in Fig. 12(a). This is not so surprising, given that, for 1976, lands in the entire area are homogeneous even before randoml reshuffling. However, the CDFs in Fig. 12(d) differ from their counterpart in that the convergence to the normal distribution is not observed. The finite size effect does not explain this
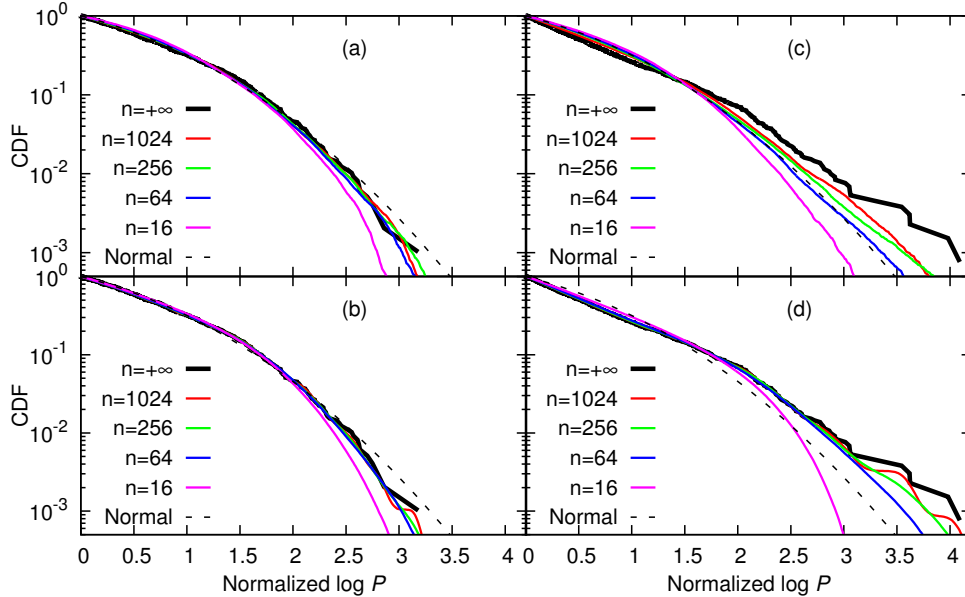
Fig. 12.   Cumulative distribution functions of normalized log prices for the original and randomly reshuffled data. Panels (a) and (b) are for prices in 1976, with (a) for the original data and (b) for the randomly reshuffled data. Panels (c) and (d) are for prices in 1987, with (c) for the original data and (d) for the randomly reshuffled data. Colors represent different values of n (i.e., the size of each region). The dotted line represents the CDF of the standard normal distribution.

difference; we should observe the same convergence in Fig. 12(d) if the convergence observed in Fig. 12(c) is due to the finite size effect.

Next, we quantitatively evaluate how prices are heterogeneous across regions by measuring the degree to which the CDF of normalized prices deviates from the CDF of the standard normal distribution. We denote this measure of deviation as $D\sqrt{N}$, where D is the maximum difference between the empirical CDF and the Gaussian CDF, and N is the number of observations. This is a statistics used in the Kolmogorov–Smirnov test, which is a nonparametric test to determine if a sample comes from a population with a specific distribution. In our case, the null hypothesis is that price observations are drawn from the standard normal distribution. The null is rejected at the 5 percent significance level if $D\sqrt{N} > 1.36$.

The deviation from the standard normal distribution is shown in Fig. 13(a). The dotted line represents the 5 percent significance level; namely, the null is accepted if the deviation is smaller than that level, and it is rejected otherwise. We see that the deviation in each year tends to be smaller as n decreases. Specifically, the deviation in the case of $n = +\infty$ is very large in every year and exceeds the 5 percent significance line. However, for $n = 1024$, the deviation falls below the 5 percent significance line in all years except 1986, 1987, 1991, and 1993, which corresponds to the period of the bubble and its collapse. Finally, when n is 256
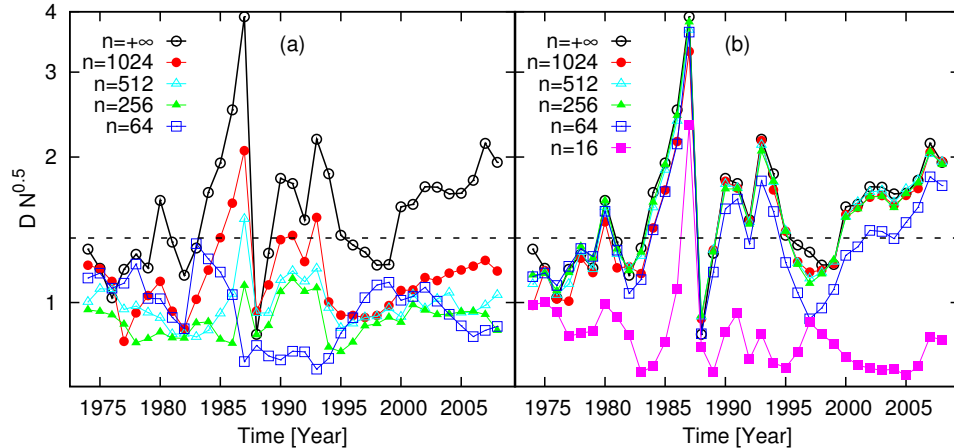
16   *T. Ohnishi et al.*

Fig. 13.   The deviation from the standard normal distribution for the original data (a), and for the randomly reshuffled data (b). Colors represent different sizes of a region, n. The dotted line is 1.36, representing a 5 percent significance level.

or smaller than that, the deviation falls below the 5 percent significance line for all years including the bubble period. This is in sharp contrast with the results obtained from the randomly reshuffled data, which is shown in Fig. 13(b). Not surprisingly, the deviation in each year does not depend on the value of n, since spatial heterogeneity is already removed through random reshuffling. Note that the deviation for the case of $n = 16$ is far smaller than the deviation in the other cases due to the finite size effect.

In sum, the analysis in this subsection shows that, during the bubble period in Tokyo, the distribution of land prices within a region of the small size is close to a lognormal distribution, but its mean and variance differs across regions. Heterogeneity across regions in this sense is an important characteristic of real estate bubbles, and it is the source of the power law tail observed during bubble periods. Put differently, our analysis suggests that a bubble period is special in that there remains ample opportunity of spatial price arbitrage, which is not present for the other years. This may be accounted for by the shortfall of liquidity available to the market participants during a bubble.[15,16]

4. Application to the U.S. House Price Data

4.1. *The U.S. house price data*

In this section, we apply our method to the U.S. house price data. The data we use is collected from the website of Zillow.com, an online real estate marketplace launched in early 2006.[17] The dataset contains price information, as well as information on the attributes of houses, such as size, address, year built, and the number of bedrooms and bathrooms, for a large number of houses for sale as of September
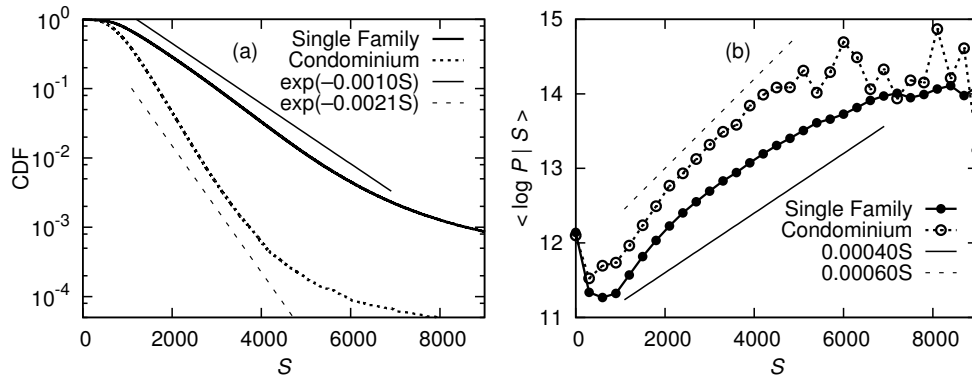
Fig. 14.   The house size distribution and the price-size relationship in the U.S. Panel (a) shows the cumulative distribution function of house sizes, measured by square feet, for single family houses (solid line) and condominiums (dotted line). The straight reference line for single family houses is of a slope of 0.0010, which is estimated by applying an OLS regression to the observations for $1000 < S < 7000$. The straight reference line for condominiums has a slope of 0.0021, which is estimated for the observations $1000 < S < 5000$. Panel (b) shows the mean of $\log P$ conditional on S for single family houses (closed circle) and condominiums (open circle). The straight reference lines are estimated by applying an OLS regression to the observations with $1000 < S < 7000$ for single family houses, and to the observations with $1000 < S < 5000$ for condominiums.

2011. Unfortunately, the historical data is not available for U.S. house prices, so that we cannot do the same analysis as we did in the previous sections. Still, we are able to examine how the U.S. house price distribution looks like as of September 2011, as well as how it differs across states with and without the experience of a housing bubble.

Fig. 14(a) shows the CDFs of house sizes, measured by square feet, for single family houses and for condominiums. We see that the CDF for single family houses is on a straight line in a semi-log plot for the range of $1000 < S < 7000$, while the CDF for condominiums is on a straight line for the range of $1000 < S < 5000$, indicating that house sizes are exponentially distributed for single family houses as well as for condominiums. The exponent of the distribution for single family houses is 0.0010, which is estimated by applying an OLS to the observations with $1000 < S < 7000$. Similarly, the exponent for condominiums is 0.0021, which is estimated using the observations with $1000 < S < 5000$. Turning to Figure 14(b), it shows the mean of log prices conditional on the size, namely, $\langle \log P|S \rangle$. We see that the log price of a house and its size are linearly correlated with a coefficient of 0.00040 for single family houses and a coefficient of 0.00060 for condominiums. Note that Fig. 14 are produced using the entire observations in the U.S., but we confirmed that the features shown in Fig. 14 are commonly observed for each state (which is not reported due to the space constraint). In the rest of this subsection, we use the observations with $S < 7000$ for single family houses and the observations with $S < 5000$ for condominiums. The number of observations is $862,485$ for single family houses, and $109,619$ for condominiums.
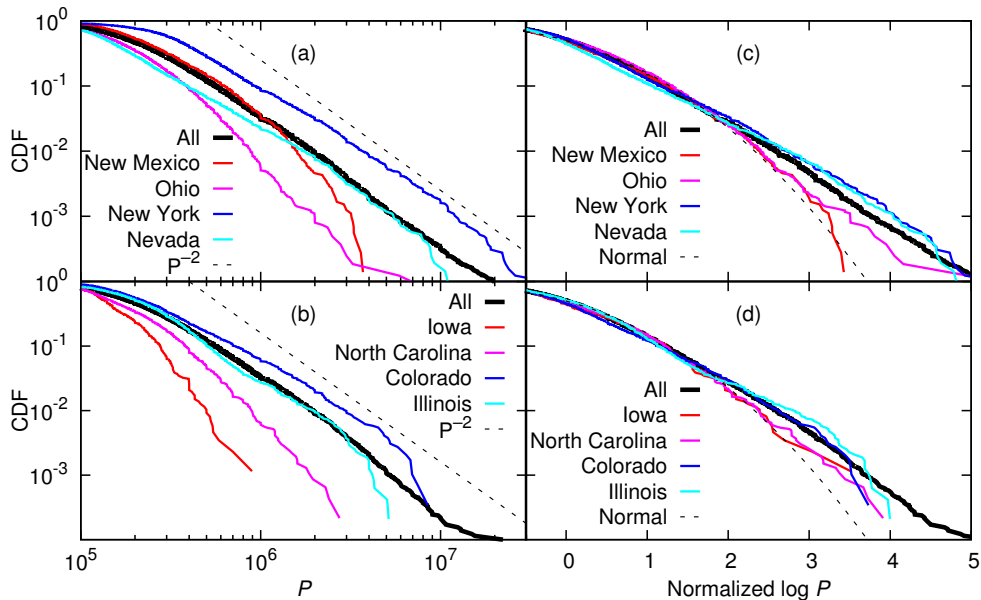
18   *T. Ohnishi et al.*

Fig. 15.   Cumulative distribution functions of prices. For single family houses, the CDF of original prices is shown in panel (a), while the CDF of normalized prices is in panel (c). For condominiums, the CDF of original prices is shown in panel (b), while the CDF of normalized prices is in panel (d). House prices are in US dollars. The black line is for the CDF of prices for the entire area, while the colored lines are for the CDFs of prices for individual states. The straight reference lines in panels (a) and (b) represent power law scaling with an exponent $\alpha = 2$. The dotted reference lines in panels (c) and (d) is the standard normal distribution.

4.2. *U.S. house price distributions by state*

Fig. 15(a) and (b) show the CDFs of prices for single family houses and those for condominiums. The black line represents the CDF for the entire observations, showing a power law tail with an exponent of around 2. On the other hand, the colored lines are for the CDFs of individual states. We see that the tail of the distribution substantially differs across states. That is, for single family houses, a power law tail is observed for New York and Nevada, but not for New Mexico and Ohio. Similarly, for condominiums, a power law tail is observed for Colorado and Illinois, but not for Iowa and North Carolina. The CDFs of normalized prices are presented in Fig. 15(c) and (d), showing that the distribution of normalized prices deviates substantially from the standard normal distribution in some states but not so much in the other states. We also calculated the distribution of size-adjusted prices as we did in the previous section, but we found that the distribution of size adjusted prices is almost the same as the distribution of original prices. In what follow, we show the results obtained using the original prices.
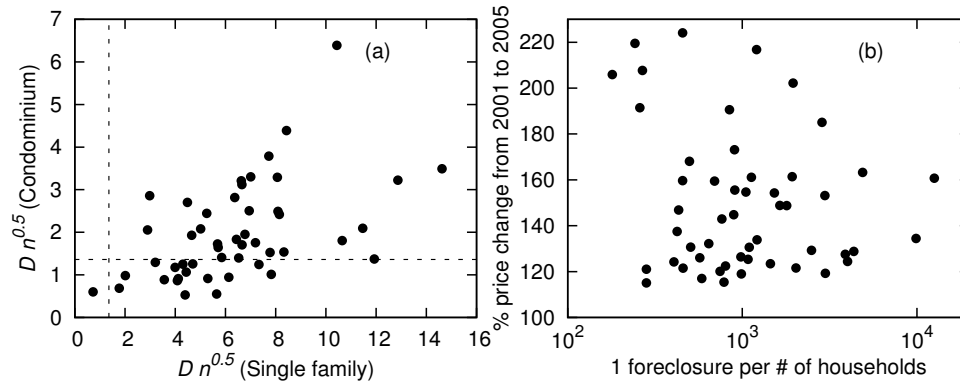
Fig. 16.   The deviation from the normal distribution for single family houses versus the deviation from the normal distribution for condominiums (a), and the foreclosure rate in October 2011 versus the house price inflation rate in 2001Q1-2005Q4 (b). Kendall's τ coefficient is 0.40 with p-value < 0.01 for panel (a), while it is -0.03 with p-value 0.39 for panel (b). The horizontal and vertical dashed lines in panel (a) is a 5 percent significance level below which the null (i.e., no deviation from the standard distribution) is not rejected.

4.3.   *The deviation from the normal distribution*

We estimate the deviation from the normal distribution, $D\sqrt{N}$, for each state. The results are shown in Fig. 16(a) as a scatter-plot of the deviation from the normal distribution for single family houses versus that for condominiums. The figure shows that the estimated deviation for single family houses is statistically significant for all states except one (North Dakota), while the deviation for condominiums is statistically significant for more than two thirds of the states. It also shows that the deviation for single family houses and that for condominiums are highly correlated, with the Kendall's τ coefficient being 0.40, which is significant at the 1 percent significance level.

The deviation from the normal distribution can be regarded as a statistic to measure the size of a housing bubble. Let us compare this with the other traditional statistics which are believed to be related with the size of a housing bubble. The first statistic is the foreclosure rate, which is defined as the total number housing units in a state divided by the number of properties that received foreclosure filings in that state. We use the foreclosure rate by state, which is published by RealtyTrac in October 2011.[18] Note that the foreclosure rate is one of the most reliable indicators of a housing bubble, although it is not observable until the bubble bursts. The second statistic is the house price inflation rate from 2001Q1 to 2005Q4, during which the U.S. real estate market was said to to be in a bubble. We calculate the inflation rates by state using the quarterly house price indexes published by the Federal Housing Finance Agency.[19]

First, we compare the foreclosure rate and the house price inflation rate, which is shown in Fig. 16(b). Somewhat surprisingly, the Kendall's τ coefficient is -0.03,
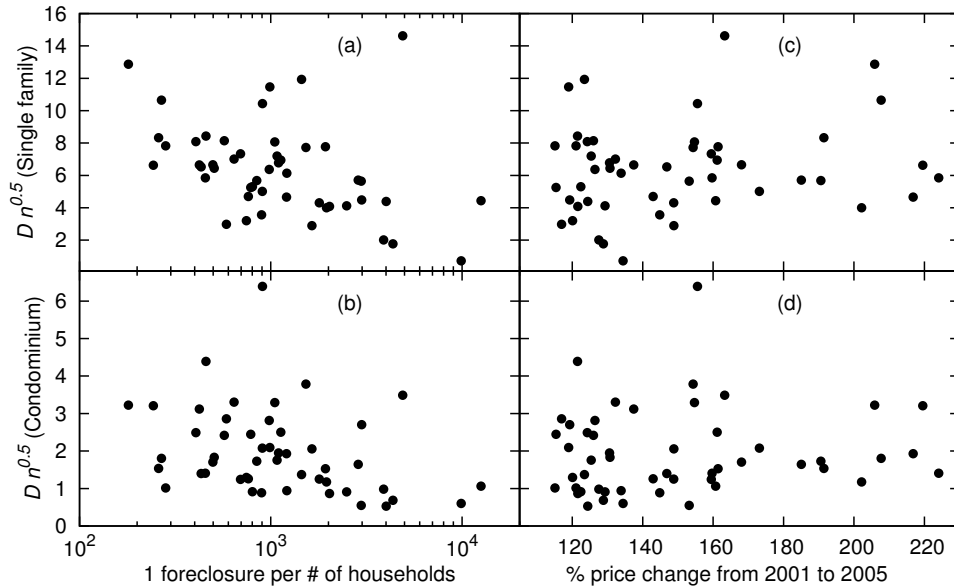
20   *T. Ohnishi et al.*

Fig. 17.   The deviation from the normal distribution versus the foreclosure rate (panels (a) and (b)), as well as versus the house price inflation rate in 2001-2005 (panels (c) and (d)). Kendall's τ coefficient is -0.33 with p-value < 0.01 for panel (a), -0.28 with p-value < 0.01 for panel (b), 0.09 with p-value 0.17 for panel (c), and 0.07 with p-value 0.39 for panel (d).

so that there is no statistically significant correlation between them. One may want to predict the foreclosure rate in a state, which is observable only *after* the bubble bursts, from the house price inflation in that state, which is observable *before* the bubble bursts. However, our result says that it is next to impossible. Next, we compare our estimate of the deviation with the foreclosure rate, which is shown in Fig. 17(a) and (b). The Kendall's τ coefficient is now -0.33 for single family houses and -0.28 for condominiums, so that the correlation is statistically significant for both cases at the 1 percent significance level. This suggests that the deviation measure may provide useful information to forecast the foreclosure rate. Finally, we compare the deviation measure with the house price inflation rate, which is shown in Fig. 17(c) and (d). The Kendall's τ coefficient is 0.09 for single family houses and 0.07 for condominiums, indicating the absence of statistically significant correlation. In sum, these results from the U.S. house price data suggest the usefulness of our deviation measure in detecting housing bubbles.

5. Conclusion

How can we detect real estate bubbles? To address this question, we started with a review of the literature on the dispersion of real estate prices during bubble periods. Then, we provided new evidence on price dispersion. First, we showed that the land 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 before and after the bubble period. Second, we showed that the land price distribution within a region of the small size was close to a lognormal even during the bubble period, but the mean and variance of the lognormal distribution was substantially different across regions. Based on this finding, we argued that this spatial heterogeneity is the source of the power law tail observed during bubbles. We also argued that this spatial heterogeneity can be regarded as evidence for the lack of price arbitrage across regions during bubble periods. Third, we found from the recent U.S. data that the tail of the house price distribution tends to be heavier in those states which experienced a housing bubble.

Acknowledgments

References

1. C. Himmelberg, C. Mayer, and T. Sinai, *J. Econ. Persp.* **19**, 67 (2005).
2. T. Ohnishi, T. Mizuno, C. Shimizu, and T. Watanabe, *Research Center for Price Dynamics Discussion Paper Series*, No. 61 (2010).
3. T. Kaizoji, *Physica A* **326**, 256 (2003).
4. C.-I. Chou and S. P. Li, arXiv:1008.1376v1 [physics.soc-ph] (2010).
5. N. MacKay, arXiv:1012.3039v1 [physics.soc-ph] (2010).
6. T. Kaizoji and M. Kaizoji, *Physica A* **344**, 138 (2004).
7. T. Yamano, *Eur. Phys. J. B* **38**, 665 (2004).
8. A. Ishikawa, *Physica A* **371**, 525 (2006).
9. A. Ishikawa, *Prog. Theor. Phys. Suppl.* **179**, 103 (2009).
10. K. Nishinari, M. Iwamura, Y. Saito, and T. Watanabe, *J. Phys.: Conf. Ser.* **221**, 012006 (2010).
11. D. P. McMillen, *J. Urban Econ.* **64**, 573, (2008).
12. W.-X. Zhou and D. Sornette, *Physica A* **329**, 249 (2003).
13. W.-X. Zhou and D. Sornette, *Physica A* **361**, 297 (2006).
14. W.-X. Zhou and D. Sornette, *Physica A* **387**, 243 (2008).
15. O. A. Lamont and R. H. Thaler, *J. Polit. Econ.* **111**, 227 (2003).
16. A. Shleifer, and R. Vishny, *J. Finance* **52**, 35, (1997).
17. http://www.zillow.com
18. http://www.realtytrac.com/trendcenter/
19. http://www.fhfa.gov/