Scientific Research

The Virtual Repeat Sale Model for the House Price Index for New Building in China

Weihua Jin, Shengping Jin*

Department of Statistics, Wuhan University of Technology, Wuhan, China
Email: *spjin@whut.edu.cn

Abstract

By using the characteristics of the new building in China, this article constructs the virtual repeat sale method to produce virtual repeat data which is similar to the repeat sale model on the house price index. Case-Shiller procedure and OFHEO method are used to calculate the house price index for new building in China. A discussion is given and furthering models are needed to take advantage of the virtual repeat sale data.

Keywords

House Price Index, Virtual Repeat Sale Method, OFHEO Index

1. Introduction

There are mainly two house price indexes in China nowadays: the China Real Estate Index System and 70 large and medium-sized city real estate price index. There is something wrong in the data quality of the house price index, for example, the China Real Estate Index System uses only the price offered by developers instead of the real price making a bargain, 70 large and medium-sized city real estate price index replaces the cover of all buildings by sampling. Moreover, there are shortages on the theory and method to compile index, which calculates the price index depended by average and they are the same as the first or second generation methods.

The most outstanding character of the real estate has hetero-quality [1]. The ones in the area of real estate always underline that the position and environment around have great influence on house price. For an extreme example, the average price of new sale house in the center of a city is ten thousand yuan per one square meter, however, the new sale houses that are all in urban by chance have the same price a year later because of re-building the old area and there are no new place, but can you say that the price before a year hadn't risen? There is no difference in terms of average price that is full of misunderstanding according to the example. Considering

---

*Corresponding author.

the strong difference in quality of house, the science way to calculate price index is to take the change price in the same quality house into consideration and separate influence from the changeable quality to sale price making a bargain, that the real necessary of market is reflected, meanwhile, the right market signal can be transferred.

Developed countries pay more attention to the two points referred upon so far in order to overcome the problem resulted from the first or second generation methods. The third generation method develops into two models, and one is hedonic method that combines house price with series of main quality (such as position, floor, square, towards, material, environment surrounding), the other is repeated-sale method [2]-[6] that is used to consider the change of price making a bargain in different period time about the same building. But Chinese cities change very fast and new buildings have dominant, that commons pay attention to house price of new buildings, so the repeated-sale method abroad can't be applied into directly. Certainly it's hard to use hedonic method, because the exact degree can't be ensured and the cost is also humorous based on collecting the character data of each house wholly.

However Chinese real estate has its own character, that new residences have structure stability and almost all communities are developed by large area usually. The difference between Chinese estate and estate abroad should be used efficiently to construct house price of new building in China.

2. Virtual Repeat Sale Data

We can mine the information of repeat trade like second-hand trade according to the product structure and possess in setting price of Chinese residence. Because of the necessary to building stability, the same building, unit and towards of house are almost same except different floors. The house price would rise by floors upward in possess of setting price for high floor or a little high house, which can be used to make up likely repeat data.

Suppose that the same building, unit and toward (simply called "three common") have been sold two floors, which is written by $l_1, l_2 (l_1 < l_2)$ that p_1, p_2 represent prices during the month t.

2.1. Rule 1: Interpolation Method

If a "three common" building were sold on the lth floor in other month, the formula below would be chosen to calculate the supposed price on the lth floor in tth month:

$$p_1 + \frac{p_2 - p_1}{l_2 - l_1}(l - l_1)$$ (1)

where that works out the fantasy price on the condition of $l_1 < l < l_2$ is called interpolation method.

2.2. Rule 2: Extrapolation Methods

If a "three common" building were sold on the lth floor in other month and on the condition of $l < l_1 < l_2$ or $l_1 < l_2 < l$, the formula (1) could be selected to compute, which that calculate the supposed price on the condition of $l < l_1 < l_2$ or $l_1 < l_2 < l$ is called extrapolation method.

2.3. Rule 3: Strengthen Method

Supposed the "three common" building had been sold out k floors that are called $l_1 < l_2 < \cdots < l_k$, which the price is written by $p_1, p_2, \cdots, p_k$ and the month is called t. If a three commons building were sold on the lth floor in other month, the outer push could be chosen on the condition of $l < l_1$ or $l_k < l$. We choose the interpolation or extrapolation method moderately in other case, which is performed detail as rule 4 below.

2.4. Rule 4: The Set of Threshold Value

The condition of $l_2 - l_1 \leq \delta_1$ is necessary when apply the interpolation method based on the inequality of $l_1 < l < l_2$.

The condition of $l_2 - l_1 \leq \delta_1$, $l_1 - l \leq \delta_2$ is necessary when apply the extrapolation method based on the inequality of $l < l_1$.

The condition of $l_k - l_{k-1} \leq \delta_1$, $l - l_k \leq \delta_2$ is necessary when apply the extrapolation method based on the inequality of $l_k < l$.

And δ_1, δ_2 are set based on, such as $\delta_1 = 4$, $\delta_2 = 2$.

Except those conditions upon, there are still other conditions to use the interpolation or extrapolation, for example, buildings must be high-floor or a little high-floor or the total floors (called l_{total}) should be 15 floors at least, which the inequality $l_1 \geq 3$, $l_k < l_{\text{total}}$.

3. Repeat Sale Method Using the Virtual Data and an Illustration

3.1. Virtual Repeat Sale Method

Similar to the BMN model and OFHEO house price index [5] [7], the model of virtual repeat sale method is represented by:

$$\ln\left(P_{it}\right) = \beta_t + H_{it} + N_i$$

where P_{it} is the real transaction or virtual sale price (yuan per square meter) at time t for the house i. β_t is the logarithm of the market price at time t. H_{it} is a Guassian process and N_i is a white noise for house i.

$$\Delta V_i = \ln\left(P_{it}\right) - \ln\left(P_{is}\right) = \beta_t - \beta_s + H_{it} - H_{is}$$

$$\Delta V_i = \sum_{\tau=0}^{T} \ln\left(P_{i\tau}\right) X_{i\tau} = \sum_{\tau=0}^{T} \left(\beta_\tau + H_{i\tau} + N_i\right) X_{i\tau} = \sum_{\tau=0}^{T} \beta_\tau X_{i\tau} + \varepsilon_i \qquad (2)$$

where $X_{i\tau}$ is a dummy variable that equals 1 if the price of house i was observed for a second time at time τ, −1 if the price of house i was observed for the first time at time τ, and zero otherwise.

In the Case-Shiller procedure, H_{it} is a Gaussian random walk. Therefore, each step of the random walk is assumed to be independent of the previous step. This is not the case for the OFHEO index; the steps are assumed to be dependent. This means that the errors in the regression when fitting the model (2) are not independent which a violation of a standard regression assumption is.

For two sales of a house at time t and s $(t > s)$, recall:

$$H_{it} - H_{is} = \sum_{\tau=s+1}^{t} \upsilon_{i\tau}$$

$$Var\left[H_{it} - H_{is}\right] = E\left[\left(\sum_{\tau=s+1}^{t} \upsilon_{i\tau}\right)^2\right] - \left(E\left[\sum_{\tau=s+1}^{t} \upsilon_{i\tau}\right]\right)^2$$

$$= E\left[\sum_{\tau=s+1}^{t} \upsilon_{i\tau}^2\right] - \sum_{\tau=s+1}^{t}\sum_{\tau'=s+1,\tau'\neq\tau}^{t} E\left[\upsilon_{i\tau}\upsilon_{i\tau'}\right]$$

$$= (t-s)\left(E\left[\upsilon_{i\tau}^2\right] - E\left[\upsilon_{i\tau}\upsilon_{i\tau'}\right]\right) + (t-s)^2 E\left[\upsilon_{i\tau}\upsilon_{i\tau'}\right]$$

where $Var(\cdot)$ is the variance of a random variable.

Their three-step procedure is described below.

Step 1. Fit the model in (2) by OLS. Step 1 gives the BMN index.

Step 2. Compute the residuals of the regression in (2), and denote these as $\hat{d}_i$. Fit the model in (3)

$$\hat{d}_i^2 = A(t-s) + B(t-s)^2 + \eta_i \qquad (3)$$

where $\eta_i \sim N\left(0, \sigma^2\right)$ which are identical independent distribution.

Step 3. The predicted values of the squared deviations from (3), $\tilde{d}_i^2$ are used to derive the weights needed to obtain GLS estimates of the β_t parameters in the following regression:

$$\frac{\Delta V_i}{\sqrt{\tilde{d}_i^2}} = \sum_{\tau=0}^{T} \beta_\tau \frac{X_{i\tau}}{\sqrt{\tilde{d}_i^2}} + \frac{\varepsilon_i}{\sqrt{\tilde{d}_i^2}} \qquad (4)$$

Index numbers for periods $t = 1, 2, \cdots, T$ are given by:

$$I_t = 100\mathrm{e}^{\hat{\beta}_t}$$

where $\hat{\beta}_t$ are the parameter estimates of (4).

3.2. An Example

There were 6354 origin records for a sample data in 2012 and in Xiangyang city, Hubei, China. 1463 virtual sale records were constructed by the method described in Section 2.

By using the virtual repeat sale method in Section 3.1, index numbers for months $t = 1, 2, \cdots, 12$ are computed as following **Table 1**.

Figure 1 is residuals of the regression for BMN model. And **Figure 2** is residuals of the regression for OFHEO model.

Comparing **Figure 1** and **Figure 2**, it is clear that BMN model is better than OFHEO model. This phenomenon is opposite to usually ones which is given by the repeat sale model. So using the virtual repeat sale data is different significant from the original repeat sale model and need to research further.

Figure 3 is depicted for the indexes of BMN and OFHEO model. In this example, the red one is more convinced us than the blue one.

4. Conclusion Remarks

The article come up with that the virtual repeat sale method produces virtual repeat data and the calculation method similar to OFHEO, and gives out a kind of virtual repeat trade model to compute house index, which tries

Table 1. The indexes given by BMN and OFHEO.

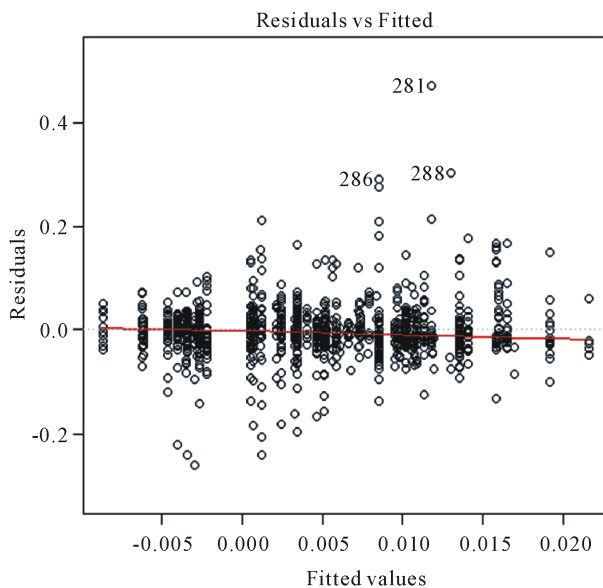| Month	1	2	3	4
BMN	100.00	100.86	100.51	100.97
OFHEO	100.00	100.67	100.02	100.20
Month	5	6	7	8
BMN	101.19	100.56	101.60	101.94
OFHEO	100.65	100.15	100.90	101.32
Month	9	10	11	12
BMN	101.66	102.19	101.72	101.31
OFHEO	100.96	101.47	101.01	100.61

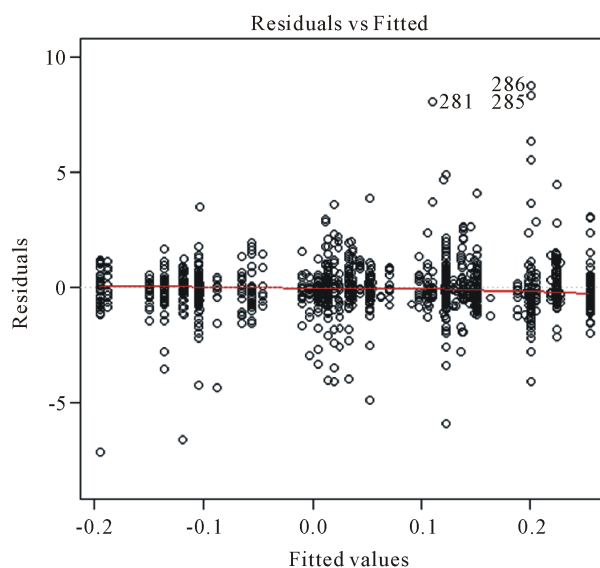**Figure 1.** The residuals for BMN model.

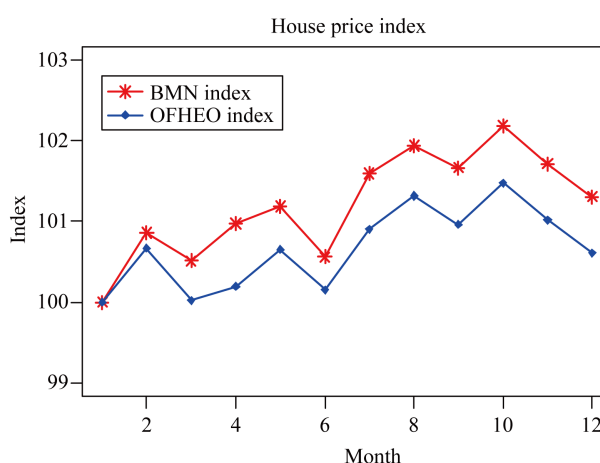**Figure 2.** The residuals for OFHEO model.

Figure 3. The indexes of BMN and OFHEO.

to calculate according to one year data of a city, based on the character of new building in Chinese city.

As the weakness of the traditional repeat sales methods, perhaps the most obvious issue is that single sales are excluded, thus reducing the sample size significantly. The number of observations which are eliminated is staggering. So, further research is needed to use all data and virtual repeat data.

Acknowledgements

References

[1]   Can, A. (1997) Spatial Dependence and House Price Index Construction. *Journal of Real Estate Finance and Economics*, **14**, 203-222. http://dx.doi.org/10.1023/A:1007744706720

[2]   Grimes, A. and Young, C. (2012) A Simple Repeat Sales House Price Index: Comparative Properties under Alternative Data Generation Processes. Motu Economic and Public Policy Research Trust and the Authors.

[3]   Bailey, M.J., Muth, R.F. and Nourse, H.O. (1963) A Regression Method for Real Estate Price Index Construction. *Journal of American Statistical Association*, **58**, 933-942. http://dx.doi.org/10.1080/01621459.1963.10480679

[4] Case, K.E. and Shiller, R.J. (1989) The Efficiency of the Market for Single-Family Homes. *The American Economic Review*, **79**, 125-137.

[5] Calhoun, C. (1996) OFHEO House Price Indices: HPI Technical Description.

[6] Feng, Y.J. and Gong, T.T. (2013) The Application of Repeated Sales Model to Calculating House Price Index among Small Cities. *6th International Conference on Information Management*, *Innovation Management and Industrial Engineering*, 468-471.

[7] Nagaraja, C.H., Brown, L.D. and Zhao, L.H. (2011) An Autoregressive Approach to House Price Modeling. *The Annals of Applied Statistics*, **5**, 124-149. http://dx.doi.org/10.1214/10-AOAS380

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.