Using Repeat Sales Model to Estimate Housing Index in Price Engineering

Yonghong Xu1*

Xia’men University, 422 Siming nan Road,Xia’men, 361005,Fujian ,China

Abstract

I suggest a series of rules in this paper which made Repeat Sales Model available in measuring index in the New Commodity Housing Market in price engineering. Then I use these rules to build Repeat Sales Model based on sales data from the department of housing management to estimate the Housing Price Index of Lian-qian district of Xia’men city of china, some topics in price engineering are argued in empirical studies, which include: how to select data, which is the best estimator, what is the real index, and what biases does new index have?

© 2011 Published by Elsevier B.V. Open access under CC BY-NC-ND license.

Key words Repeat Sales Model; Housing Price Index; Real Estate Index; Price Engineering

1. Introduction

Housing serves as a major source of individual wealth in most country. Hence, changes in its value may influence consumer spending and saving decision, in turn affecting overall economic activity. In order to depict the housing market, most government intends to sponsor engineering research works on housing price index. But the price of housing is harder to measure than that of most other goods and assets, because of two key distinguishing characteristics.

First point and most importantly, houses are heterogeneous. No two houses are identical, if only because they cannot occupy quite the same location. This means that we cannot always reliably predict the sales price of a given dwelling from the price of another. Second point, houses are generally sold infrequently, and we cannot easily observe the market price of a given dwelling without it being sold. Because of these two problems, the classical index models would not serves well in this field [1] [2].

Worldwide, the most frequently used methods for calculating house price indices are: (1) hedonic price models suggested by A.T. Court (1939)[3]; (2) Repeat Sales Model suggested by Bailey, Muth, and Nourse (1963)[4]; (3) variants on and hybrids of the latter two, also called hybrid model, suggested by Case and Quigley (1991)[5][6].

I attempt to apply First point in Xia’men, fujian province to find a new way to build housing price index in china, because of there is seldom research on Repeat Sales Model in china, especially on applicant research. Further more,
Repeat Sales Model using transaction date, and most other research always worked on appraisal data, from literature research, transaction date was proved more accuracy than appraisal data.

The article is organized as follows. Section 2 describes the repeat sales regression and then discusses some of the problems in application and interpretation of the method. Section 3 describes the data of we used. Section 4 presents the results, and section 5 concludes.

2. Methodology

2.1. Repeat Sales Model

The repeat sales methods use the houses which transact more than two times to estimates the market price tendency. It treats the appreciation of a house within in period $t$, $r_{i,t}$, as the natural log of ratio of the price of the house at the end of the period over the price of the house at the beginning of the period.

$$ r_{i,t} = \ln \left( \frac{p_{i,t}}{p_{i,t-1}} \right) \quad (1) $$

$P$ is price of the house. We will assume that the log return for a certain asset $i$ in period $t$ may be represented by the logged price relative of an index of house market $\mu_t$, and an error term $\epsilon_{i,t}$.

$$ r_{i,t} = \mu_i + \epsilon_{i,t} \quad (2) $$

We wish to use sales data about houses to estimate the index $\mu_t$ over some interval $t = 1, 2, 3, \cdots, T$. Our observed data consist of purchase and sales price pairs, $B_i$ and $S_i$, of the individual house comprising the index, as well as the dates of purchase and sale: $b_i$ and $s_i$. Thus,

$$ y_i = \ln (S_i / B_i) = \sum_{t=b_i}^{s_i} r_{i,t} = \sum_{t=b_i}^{s_i} \mu_i + \sum_{t=b_i}^{s_i} \epsilon_{i,t} \quad (3) $$

Equation (3) can also be rewrite to matrix form, like,

$$ y = X\mu + \epsilon \quad (4) $$

Where $X$ is a designed matrix each row of the designed matrix corresponds to one observation in repeat sample and first none zero element in row $i$ appears immediately after the purchase time while the last none zero element appears at the time of the sale.

There are two forms of designed matrix, geometric specification of $X$ and arithmetic specification. Geometric specification of $X$ has $x_i$ equal to $[\cdots 0 \ -1 \ 0 \cdots 1 \ 0 \cdots]$ with time period $b_i = -1$ and $s_i = 1$ while arithmetic specification has the $i$th row with $[\cdots 0 \ 1 \ 1 \cdots 1 \ 0 \cdots]$. The geometric specification was used by Bailey, Muth, and Nourse (1963) [4], Shiller (1991) and the arithmetic specification was used by Case and Shiller (1989) [5] [6] and Crone and Voith (1992) [7].
2.2. The estimators

2.2.1. Unweighted repeated sales estimator (OLS)

This is the original formulation of the repeat sales regression by Bailey, Muth, and Nourse (1963) [4]. It uses ordinary least squares to estimate the logged return series by assume the $\varepsilon_{i,t}$ in equation (3) is $\varepsilon_{i,t} \sim i.i.d \ N(0, \sigma^2_{i,t})$.

$$\hat{\boldsymbol{\mu}}_{\text{OLS}} = (X'X)^{-1}X'y$$

(5)

By ignoring the differing variances of the observations, the procedure overweighs those that contain relatively less information and underweight those that contain relatively more information about the fluctuations of the $\mu$ series.

2.2.2. Maximum likelihood estimator (GLS)

Webb (1981a, 1981b, and 1981c) [8][9][10] found that error term $\varepsilon_{i,t}$ was not independent with holding period, so he proposed GLS estimator and showed it to be the uniform minimum variance estimator under similar assumptions. In GLS procedure, the weights are chosen as the square root of the inverse of the holding period for each logged price, GLS estimator can be write as:

$$\hat{\boldsymbol{\mu}}_{\text{GLS}} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

(6)

Where $\Omega = P'P$, and $P^{-1}$ is weight matrix with $1/\sqrt{s_i - b_i - 1}$ as its diagonal elements.

The GLS estimator is also the maximum likelihood estimator under the assumption that the errors are distributed normally i.i.d.

2.2.3. Three-stage weighted repeat sales (WRS)

Case and Shiller (1987) [5][6] observe that, in the case of real estate assets, the error variance for an individual house grows with the length of the holding period. However, they suggest that there may also be a fixed component of error variance that is unrelated to the length of the holding period. To jointly estimate the fixed and the variable components of variance, they suggested a model in which the errors from an OLS regression are used. It contains three stages. Stage 1 in the WRS model is the OLS estimate. The squared errors from stage 1($\hat{\varepsilon} = y - X\hat{\boldsymbol{\mu}}_{\text{OLS}}$) are then regressed upon a constant and a vector of holding periods for each house. Thus, stage 2 tests the proposition that the error variance grows linearly with the holding interval, $\hat{\varepsilon}^2 = \alpha - \psi \beta + \tau$ and that there is a fixed component to the house specific variance unrelated to the holding period, $\hat{\varepsilon}^2 = \hat{\alpha} - \psi \beta$. Case and Shiller use the inverse of the square root of the fitted value in stage 2 as the weights in the third stage, a GLS procedure. The WRS improves on the GLS estimator by introducing a constant term in the weight estimate. WRS estimator can be writing as:

$$\hat{\boldsymbol{\mu}}_{\text{WRS}} = (X'\Omega_{\text{WRS}}^{-1}X)^{-1}X'\Omega_{\text{WRS}}^{-1}y$$

(7)

Where $\Omega_{\text{WRS}}$ is diagonal matrix with diagonal elements $1/\hat{\varepsilon}^2$.

The WRS method is widely used in the housing finance industry. Goetzmann (1992) [11] proposed several alternative estimators including a “Stein-like” estimator first proposed by Hill and Zeimer (1984) [12] and several Bayesian estimators and evaluates them along with WRS and the conventional estimators. However these estimators tent to be computationally cumbersome in large samples and hence have not been widely adopted.

2.3. Adjustment

Logarithmic transformation brought us convenience in operation, but the log transformation results in a downward bias of the arithmetic mean at each point. Unfortunately, there is no way to solve explicitly for the arithmetic average given the geometric average, although we may approximate it, under the assumption that the property returns in each period are log normally distributed. So we adjust this bias by using the estimated variance.
\[ I_{t,\text{adj}} \approx \exp(\mu_t + \frac{1}{2}\var(\mu_t)) \]  

(8)

3. Data

3.1. homogeneous-like sample

For the purpose of this study, I choose the Lian-qian District as a case study. Lian-qian District is the important house market of Xiamen affiliated to Fujian Province (China) between 1999 and 2005, it covers almost 1/2 house transaction of Xiamen. In China, the focus of the housing market is the price tendency of new constructed house. Theoretically, the Repeat Sales Model is set for the two-hand deal, and it is not available in the new constructed house market. But by setting a series of rules, I found that Repeat Sales Model can also be reset to suit the new constructed house market.

Unlike single house in America, Chinese house is constructed as a building, like figure 1. For the safety consideration of the building, the floor plan is always identical in every floor, and the house in the same floor usually priced identically which imply the different rooms in the same floor of the same building is identical in living quality. Like figure 1, there are four buildings in a community, A12 and A12 always be sold in the same price for every M2, A21 and A22 are also have the same characteristic, but the price for every M2 of A51 is not the same as B51, C51, D51. Because there are not in the same building.

![Fig.1. Chinese house construction](image)

So if A11 and A12 are sold in different period, we can treat it as the same house exchange in two times. To find the homogeneous-like sample like A11 and A12, we suggested some rules: (1) homogeneous-like sample must be chose from the same floor. (2) homogeneous-like sample must price the same in the same period. (3) if the same community have some building like A, for example building B in figure 1, we can expanded homogeneous-like sample from these similar building. We can also set A11 /A12 /B11 /B12 as homogeneous-like sample.

3.2. Data exclude rules

I got 5312 data between 2003/1/1 and 2005/3/31 for this research. To construct a market housing price index, I also set some rules to exclude the price anomalies. These including: (1) the house which is supply as welfare for civil servant or for low income groups. (2) The house is not availability for the public, like some house offer to internal staff. (3) A large change happens to the house sales pair. (4) There were no physical changes to the property, there may have been a recording error in one of the sale prices, or an excessive price change caused by idiosyncratic, non-market factors. Rule (4) is different to observe; so I set three standards to capture the non-market factors. standard 1 define the non-market factors from the single price which is lower than 85% of homogeneous-like sample in the same period, while standard 2 is 90% and standard 3 is 95%.

4. Empirical results

4.1. Standard for excluding non-market factors

A review of the results for standards for excluding non-market factors in this section. Table 1 presents a summary of the regression effect in different for excluding non-market factors. Standard 1 brought 8.66% sample loss, but the adjust R² from OLS, GLS and WRS are all below 0.4. It shows that samples in this research can hardly find
tendency because of the non-market factors interference. From Table 1, Standard 3 is most convenience one by considering sample loss and regression statistics.

Furthermore, we also can easy find evidence to support Standard 3. In our economic system, house always stands for an expensive wealth or a large sum of money. A small discount in house transaction means thousands of or millions of money. It is hard for buyers to ask for discounts exceed 5% in bargaining. Over the past few years, the housing price in all city in china is rising, real estate developer usually reluctant even reject to offer any discount in price, so the actual situation of supply and demand support Standard 3.

Table 1. Regressions statistics in different Standards for excluding non-market factors

<table>
<thead>
<tr>
<th>Standard for excluding non-market factors</th>
<th>Sample left</th>
<th>Ratio of Lost sample</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>OLS</td>
</tr>
<tr>
<td>All</td>
<td>2345</td>
<td>0</td>
<td>0.1785</td>
</tr>
<tr>
<td>Standard 1</td>
<td>2142</td>
<td>0.0866</td>
<td>0.3511</td>
</tr>
<tr>
<td>Standard 2</td>
<td>1896</td>
<td>0.1915</td>
<td>0.4725</td>
</tr>
<tr>
<td>Standard 3</td>
<td>1750</td>
<td>0.2537</td>
<td>0.6101</td>
</tr>
</tbody>
</table>

4.2. Standard for excluding non-market factors

From Table 1, GLS and WRS have not show apparent advantage in regression in different samples. The adjusted R2 in GLS is below than OLS in any forms of dataset both include non-market transaction datasets and excluding non-market transaction dataset. The adjusted R2 in WLS is exceeding that in OLS, but it is not notable. So we suggest OLS to estimate Repeat Sales Model in new constructed house market in china, though adjusted R2 in OLS is a little below WRS, it is much more easy than WRS.

4.3. Contractions of different index

Figure 2 shows three housing price index built in different method. “RSM-HPI” stands for the housing price index that is constructed with Repeated Sales Model but without adjusting by equation (8), while “Adjusted RSM-HPI” stands for the Housing Price Index that is built with Repeated Sales Model and adjusted by equation (8). To show the merit of Repeated Sales Model, I calculate a housing price with classical index theory which calculated by mean price without considering heterogeneity of house. The housing price index with classical index theory is “HPI from mean price” in Figure 2.

“RSM-HPI” and “Adjusted RSM-HPI” depict an upward tendency of housing price from 2003/1/1 to 2005/3/31. “Adjusted RSM-HPI” located at the top of “RSM-HPI”, and the gap between these two indexes become larger with time. These evidence shows that log transformation will brought bias in index, and this problem will be worse with time if we have not adjusted it. In our economic word, we should adjust the housing price index with equation (8) when using Repeat Sales Model.

“RSM-HPI” and “Adjusted RSM-HPI” show different price tendency with “HPI from mean price”, with reviewing the housing price tendency and economic develop over the past few years, we find that “HPI from mean price” depict a false tendency of housing price.
Fig. 2 Housing price index from different method

4.4. Index bias

Cho (1996) [13] [14] identifies five types of bias that can found in the context of RSM: renovation bias, hedonic bias, trading-frequency bias, sample-selection bias and aggregation bias. Renovation bias arises due to the possibility of structural changes to the property between sales in a repeat sample. If there are structural changes over time, then the RSM no longer yields a constant quality index. Hedonic bias is related to renovation bias and arises when the values of some structural and location attributes vary over time. Sample selection bias in RSM arises on account of the fact that the repeat sales sample may not be representative of all properties. Aggregation bias arises due to the flexibility of setting the time interval in estimating the RSM. Any specific interval that is used can both includes irrelevant information as well as excludes information: a quarterly index includes irrelevant information if the underlying process is monthly. However if the underlying process is yearly, then the quarterly index excludes relevant information and is biased on that accounts.

In this paper, data are from the new constructed house market. The new house is always without renovation, so the index would be without renovation bias. The new house is always sold in a short time span, so hedonic bias might be disappearing. Furthermore I use all sample in index estimate without sampling, so the index has no sample selection bias. Once using Repeat Sales Model in new house market, most bias disappears.

5. Concluding remarks

This paper employs Repeat Sales Model to build housing price index in Xia’men, fujian province of china. In order to make this model suit to new constructed house market to depict the key problem of Chinese housing, I set a series of rules to find the homogeneity in new constructed houses. Then I use these rules to build Repeat Sales Model based on sales data from the department of housing management to estimate the Housing Price Index of Lian-qian district of Xia’men city. From empirical results I find several points for housing price index practice: (1)if a housing is sale lower than 95% of the similar housing in the same time, it might be a non-market sample, I suggested to exclude it in index calculate. (2) GLS and WRS have not showed nominal excellence in new constructed house market, we suggest using OLS to estimate Repeat Sales Model. (3) Repeat Sales Model shows it merit in deal with heterogeneity in house market. When use it, we should adjust the log transformation bias, or we would get an underestimate index. (4) In new constructed house market, some bias of Repeat Sales Model will disappear.

6. Copyright

All authors must sign the Transfer of Copyright agreement before the article can be published. This transfer agreement enables Elsevier to protect the copyrighted material for the authors, but does not relinquish the authors' proprietary rights. The copyright transfer covers the exclusive rights to reproduce and distribute the article, including reprints, photographic reproductions, microfilm or any other reproductions of similar nature and translations. Authors are responsible for obtaining from the copyright holder permission to reproduce any figures for which copyright exists.

Acknowledgements

The project was supported by the Natural Science Foundation of Fujian Province of China (No. 2011J05166)

References


