New Methodology for Constructing Real Estate Price Indices Applied to the Singapore Residential Market

Liang Jiang, Peter Phillips and Jun Yu

SMU

Motivations

- Real estate prices are one of the key indicators of economic activity,
- Housing market is a highly important sector,
- An upsurge of interest in studying real estate markets in the literature.

Challenges

- Houses are distinctive.
- House transactions are infrequent and sales data are unbalanced.
- The average quality of properties in the market may change.

Standard Hedonic Model

- Basic idea: observed house prices may be regarded as the composite sum of elements that reflect implicit structural and locational prices (Rosen, 1974).
- The standard hedonic model (Ghysels et al., 2012) is:

$$y_{i,j,z} = c + \beta_{t(i,j,z)} + \gamma' X_{i,z} + \varepsilon_{i,j,z} \tag{1}$$

Specification bias

- The choice of the covariates is limited by data availability and involves subjective judgements,
- $\epsilon_{i,j,z}$ might be correlated with $X_{i,z}$ because not all covariates are observed,
- can lead to spurious regression effects in which the irrelevant hedonic variables are significant,
- A parametric function is needed to relate $Y_{i,j,z}$ to $X_{i,z}$.

Repeat-Sales Methods

- Basic idea: by looking at the difference in sale prices of the same house, it avoids the difficulties of choosing hedonic information and specifying functional forms.
- The Case-Shiller Model (Case and Shiller, 1987,1989)

$$
\begin{aligned}
y_{i,j,z} - y_{i,j-1,z} &= \beta_{t(i,j,z)} - \beta_{t(i,j-1,z)} + \sum_{k=t(i,j-1,z)-1}^{t(i,j,z)} u_{i,z}(k) \\
&\quad + \varepsilon_{i,j,z} - \varepsilon_{i,j-1,z} \qquad (2)
\end{aligned}
$$

- May be motivated by the following specification:

$$
y_{i,j,z} = \beta_{t(i,j,z)} + f(X_{i,z}, \gamma) + \mu_z + \sum_{k=0}^{a_{t(i,j,z)}} u_{i,z}(k) + \varepsilon_{i,j,z} \quad (3)
$$

Sample Selection Bias

- Single sale homes are omitted from the analysis, reducing the sample size significantly.
- Sample selection bias: are repeat-sales representative?
- Criticized by Clapp et al. (1991), Gatzlaff and Haurin (1998), Nagaraja et al.(2010) etc.

Other Literature

- Case and Quigley (1991),
- Quigley (1995) and Englund et al. (1998),
- Hill et al. (1997),
- Shiller (2008),
- Nagaraja, Brown and Zhao (2011) etc.

Novelty

- We propose a new methodology to construct real estate price indices that addresses some of the criticisms of the hedonic and repeat-sales methods.
- In our approach, the model is designed to control for hedonic information in a general way and pair sale prices at the individual building level, instead of the individual house level as is done in the repeat-sales method.

Four Main Advantages

- First, the method makes use of all the real estate information in the sample: robust to sample selection bias and gains in efficiency;
- A number of fixed effects are included in the framework to control for unobserved hedonic information and the functional form linking price and hedonic information is left unspecified: robust to specification bias;
- The new model puts greater weight on pairs whose time gaps between sales are smaller, similar to repeat-sales methods; but since our pairs are constructed at the building level, the time gaps in our pairs are much smaller than those in pairs for repeat-sales methods.
- Our model involves a simple and convenient GLS estimation procedure that is easy to implement and computationally efficient.

New Model

- Basic idea: to create the sale pair at the individual building level.
- The new model is:

$$y_{i,j,p} = \beta_{t(i,j,p)} + f(Z_p) + \gamma' X_{i,p} + \mu_p + \sum_{k=t(1,1,p)+1}^{t(i,j,p)} u_{k,p} + h_{i,p} + \epsilon_{i,j,p}, \quad (4)$$

New Model

- Take price averages in each building at each time period whenever there are sales giving

$$\bar{y}_{t,p} = \beta_t + f(Z_p) + \gamma' \bar{X}_{t,p} + \mu_{z(p)} + \sum_{k=t_1(p)+1}^{t} u_{k,p} + \bar{h}_{t,p} + \bar{\epsilon}_{t,p}, \tag{5}$$

- Take the difference of model (5) at two time period t and $t'(t' > t)$

$$\begin{aligned} \bar{y}_{t',p} - \bar{y}_{t,p} &= \beta_{t'} - \beta_t + \gamma' \left(\bar{X}_{t',p} - \bar{X}_{t,p} \right) + \sum_{k=t+1}^{t'} u_{k,p} \quad (6) \\ &\quad + \bar{h}_{t',p} - \bar{h}_{t,p} + \bar{\epsilon}_{t',p} - \bar{\epsilon}_{t,p}. \end{aligned}$$

Estimation

- Run an OLS regression of model (6) to obtain initial estimates of β_t for all t and γ.
- Plug these initial estimates into (6) to calculate the regression residuals, denoted by $\widehat{e}_{t',p}$, which are fitted values of the composite component

$$e_{t',p} = \sum_{k=t+1}^{t'} u_{k,p} + \bar{h}_{t',p} - \bar{h}_{t,p} + \bar{\epsilon}_{t',p} - \bar{\epsilon}_{t,p}$$

and

$$Var\left(e_{t',p}\right) = (t' - t)\sigma_u^2 + \left(\frac{1}{n_{t',p}} + \frac{1}{n_{t,p}}\right)(\sigma_h^2 + \sigma_\varepsilon^2) \quad (7)$$

Estimation

- To calculate the weights to be used in GLS estimation, we run the following regression

$$\hat{e}^2_{t',p} = c + (t' - t)\sigma^2_u + \left(\frac{1}{n_{t',p}} + \frac{1}{n_{t,p}}\right)(\sigma^2_h + \sigma^2_\varepsilon) + v_{t',p},$$

  where $E(v_{t',p}) = 0$. Then the weights are the reciprocals of the fitted values from model (7). The diagonal matrix $\widehat{W}$ with weights appearing in the main diagonal is then the estimated weight matrix for GLS estimation.
- Using $\widehat{W}$ as the weight matrix, GLS regression of (6) gives the final estimates of β_t for all t and γ.

Empirical Analysis
Data

- Data Source: the Urban Redevelopment Authority (URA) in Singapore
- Quarterly data of private residential property sales from Q1 1995 to Q2 2014
- Total around 216,000 dwellings transacted. Among them, about 146,000 houses are single-sales and the remainder, about 70,000 houses, are repeat-sales. Total around 315,000 transactions recorded.
- Total number of buildings is 4820, which leads to 81,000 pairs at the building level.
- Extensive records of information: The sale price (both the total price and the price per square foot), the transaction period, the district, sector and postal code of every transacted property, floor and unit number, project number, size, sell type, property type, completion year, tenure length, and location type

Empirical Analysis

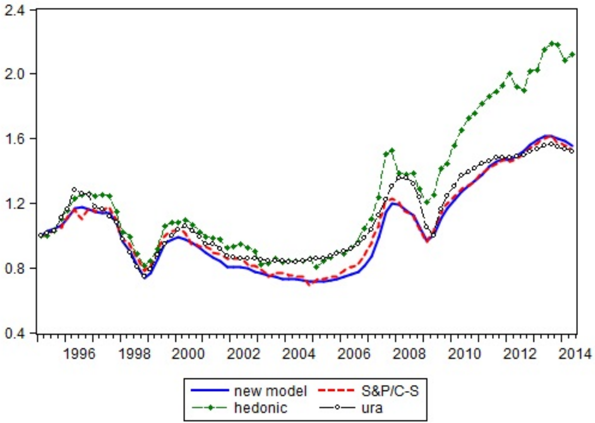

Figure 1: Real Estate Indices for Singapore: Q1 1995-Q2 2014

Comparison of out-of-sample predictive power

- Divide the observations into training and testing datasets by random sampling.
  - The testing set contains all the final sales of the houses sold three or more times in our sample period,
  - Among the houses sold twice, their second transactions are randomly put into the testing set,
  - We also randomly add single-sale houses into the testing set.
- The resulting testing set contains around 15% of sales in our sample, of which 50% are single-sale houses and the rest are repeat-sales.
- All the remaining houses are included in the training set.

Comparison of out-of-sample predictive power

- First estimate the indices on the training set by the different modeling methods and then examine their out-of-sample forecasts on the testing set.
- To calculate the estimated price for the repeat-sales homes in the testing set,
  - our new method

$$\hat{Y}_{t',i,p} = \frac{\widehat{I}^{bb}_{t'}}{\widehat{I}^{bb}_{t}} \bar{Y}_{t,p}, \tag{8}$$

  - the S&P / Case-Shiller

$$\hat{Y}_{t',i} = \frac{I_{t'}}{I_t} Y_{t,i} \tag{9}$$

  - the hedonice model

$$\widehat{y}_{i,j,z} = \widehat{\mu}_z + \widehat{\beta}_{t(i,j,z)} + \widehat{\gamma}' X_{i,z} \tag{10}$$

Comparison of out-of-sample predictive power

Table 1: Testing set (with only repeat sales houses included): RMSE & MAE for the Indices (SG dollars)

Loss Function	new model	S&P/C-S	hedonic
RMSE	141	175	291
MAE	92	122	220

Time Gap of Sales in the Same Building

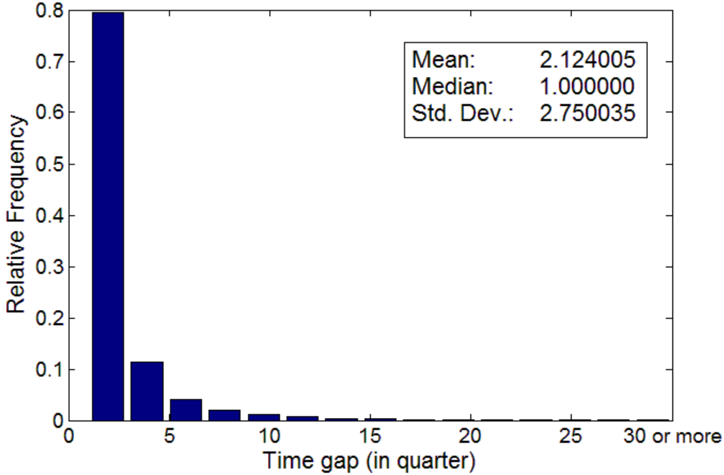

Figure 2: Histogram, mean, median and standard deviation of the time gap of sales in the same building.

Time Gap of Sales in the Same House

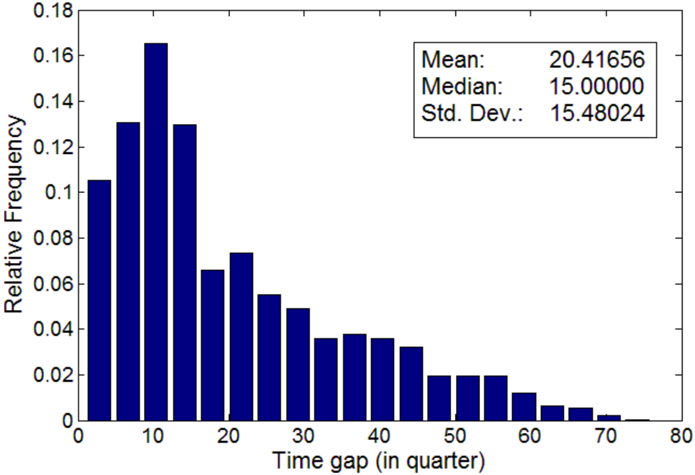

Figure 3: Histogram, mean, median and standard deviation of the time gap of sales in the same house.

Comparison of out-of-sample predictive power

Table 2: Testing set (only single sale houses included) RMSE & MAE for the Indices (SG dollars)

Loss Function	new model	hedonic
RMSE	156	297
MAE	86	188

Table 3: Testing set (all houses included) RMSE & MAE for the Indices (SG dollars)

Loss Function	new model	hedonic
RMSE	149	294
MAE	89	204

Indices (Q1/08-Q2/14) and Cooling Measures

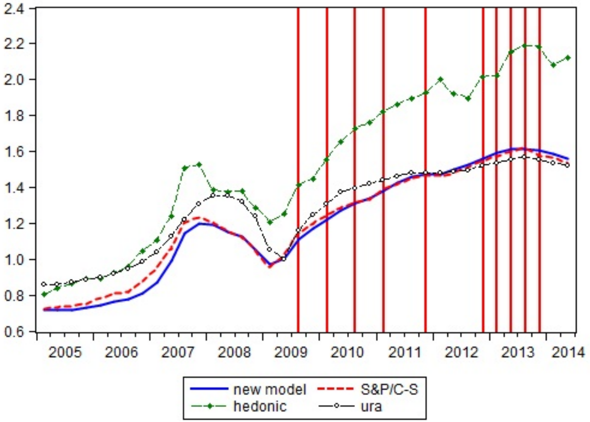

Figure 4: Four real estate price indices and the dates of ten rounds of successive macroprudential cooling measures.

Explosive behavior

- Phillips, Shi and Yu (2015a, 2015b): Method for testing for multiple bubbles in real time.
- Idea: find explosive behavior in housing prices, but not in rental in real time, by comparing the test statistic and the critical value.

Phillips, Shi and Yu (2015a, 2015b)

- Propose a right sided unit root tests (against explosive alternatives) by using a fitted model for data of the following form

$$\Delta X_t = \hat{\alpha} + \hat{\beta} X_{t-1} + \sum_{i=1}^{K} \hat{\beta}_i \Delta X_{i-1} + \hat{e}_t \qquad (11)$$

- The origination and termination dates of an explosive period are then determined from the crossing times

$$\hat{r}^e = \inf_{r \in [r_0, 1]} \{r : BSADF_r > cv\} \text{ and } \hat{r}^f = \inf_{r \in [\hat{r}^e, 1]} \{r : BSADF_r < cv\}$$

where $BSADF_r = \sup_{r_1 \in [0, r_2 - r_0], r_2 = r} \{ADF_{r_1}^{r_2}\}$, a recursive statistic.

Explosive behavior in the Case-Shiller Index

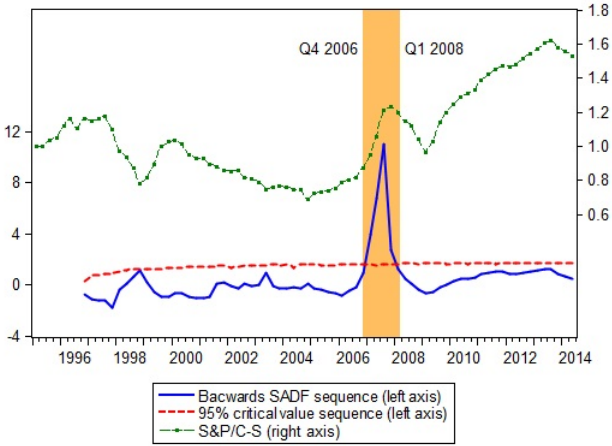

Figure 5: The S&P/Case-Shiller index, the BSADF statistic of PSY and the critical values.

Explosive behavior in the hedonic index

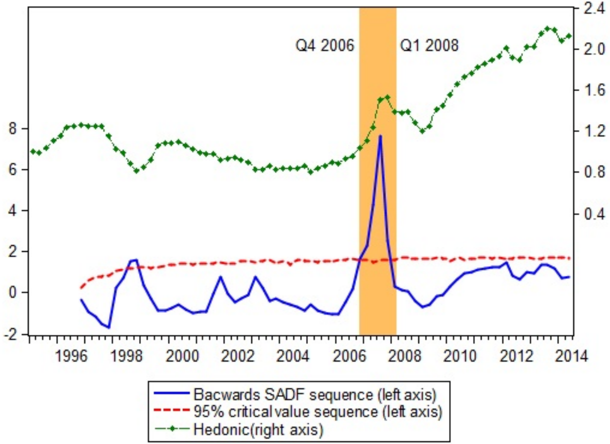

Figure 6: The index from the hedonic model, the BSADF statistic of PSY and the critical values

Explosive behavior in our new index

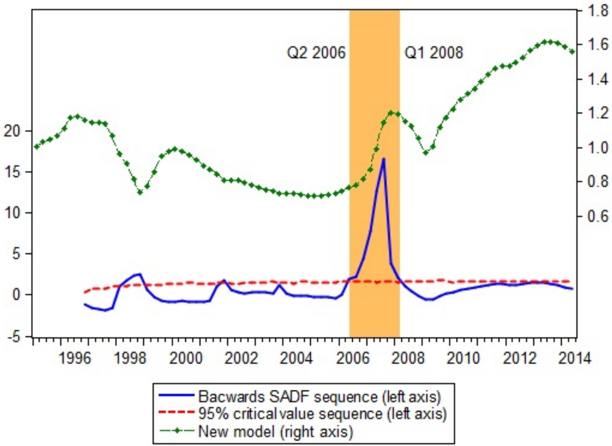

Figure 7: The new index, the BSADF statistic of PSY and the critical values.

Explosive behavior

- While the same conclusion date for the explosive period is found for the three indices, our new index suggests that explosive behavior commenced two quarters earlier.
- If the government had been alerted to an explosive market condition in real time, some cooling measures could have been implemented.
- No explosiveness between 2009 and 2013.
- This curvature must have been a result of many rounds of the cooling measures by the government
- Government's policy intervention to cool down the housing market is effective

Conclusion

- This paper provides a new methodology for the estimation of real estate price indices, which is less prone to specification bias than the standard hedonic model.
- Moreover, it exploits all available information in real estate markets, therefore has no sample selection bias involved.
- The model is also easy to estimate.
- Our empirical assessment shows that, compared with alternative methodologies, our method has superior performance out-of-sample.
- We also apply the recurisve detection method to each index to locate episodes of real estate price exuberance in Singapore and to assess the cooling measures implemented by Singapore government.

Thank You