

## Accepted Manuscript

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

Liang Jiang, Peter C.B. Phillips, Jun Yu

PII: S0378-4266(15)00237-X

DOI: <http://dx.doi.org/10.1016/j.jbankfin.2015.08.026>

Reference: JBF 4809

To appear in: *Journal of Banking & Finance*

Received Date: 30 December 2014

Accepted Date: 18 August 2015

Please cite this article as: Jiang, L., Phillips, P.C.B., Yu, J., New Methodology for Constructing Real Estate Price Indices Applied to the Singapore Residential Market, *Journal of Banking & Finance* (2015), doi: <http://dx.doi.org/10.1016/j.jbankfin.2015.08.026>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



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

Liang Jiang

*Singapore Management University*

Peter C.B. Phillips

*Yale University, University of Auckland, University of Southampton,  
Singapore Management University*

Jun Yu

*Singapore Management University*

June 26, 2015

## Abstract

This paper develops a new methodology for constructing a real estate price index that utilizes all transaction price information, encompassing both single-sales and repeat-sales. The method is less susceptible to specification error than standard hedonic methods and is not subject to the sample selection bias involved in indexes that rely only on repeat sales. The methodology employs a model design that uses a sale pairing process based on the individual building level, rather than the individual house level as is used in the repeat-sales method. The approach extends ideas from repeat-sales methodology in a way that accommodates much wider datasets. In an empirical analysis of the methodology, we fit the model to the private residential property market in Singapore between Q1 1995 and

---

\*Liang Jiang, School of Economics, Singapore Management University, 90 Stamford Road, Singapore. Peter C. B. Phillips, Yale University, 30 Hillhouse Avenue, New Haven, CT, 06520, USA; Email: peter.phillips@yale.edu; Phillips acknowledges research support from the NSF under Grant No. SES 12-58258. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore. Email: yujun@smu.edu.sg. URL: <http://www.mysmu.edu/faculty/yujun/>. Yu thanks the Singapore Ministry of Education for Academic Research Fund under grant number MOE2011-T2-2-096. All authors thank the Singapore Ministry of Education for Academic Research Fund under grant number MOE2013-T3-1-017. An earlier version of this paper was presented in the first Conference on Recent Developments in Financial Econometrics and Applications held at Deakin University. We acknowledge helpful comments from Dong Hong, Paresh Narayan, Philip Ng, Susan Wachter, and especially two referees and the Editor.

Q2 2014, covering several periods of major price fluctuation and changes in government macroprudential policy. The index is found to perform much better in out-of-sample prediction exercises than either the S&P/Case-Shiller index or the index based on standard hedonic methods. In a further empirical application, the recursive dating method of Phillips, Shi and Yu (2015a, 2015b) is used to detect explosive behavior in the Singapore real estate market. Explosive behavior in the new index is found to arise two quarters earlier than in the other indices.

*JEL classification:* C58, R31

*Keywords:* Cooling measures, Explosive behavior, Hedonic models, Prediction, Real Estate Price Index, Repeat sales,

## 1 Introduction

Real estate prices are one of the key indicators of economic activity. Indices measuring changes in real estate prices help to inform households about their asset wealth and to make a wide variety of economic decisions that depend on wealth resources. Policy makers rely on the information imported by these indices in designing and formulating monetary and fiscal policies at the aggregate level as well as macro-prudential policies directed at the financial and banking sectors. Though real estate prices are widely accepted as highly important economic statistics,<sup>1</sup> the construction of a suitable index that will reflect movements in the price of a typical house in the economy presents many conceptual, practical, and theoretical challenges.

First, houses are distinctive, making it particularly difficult to characterize a “typical” house for the development of an index. Different houses have varying characteristics such as location, size, ownership, utilities and indoor/outdoor facilities. These differences imply that averaging all market transaction prices without controlling for house heterogeneity inevitably produces bias. Second, house transactions are infrequent and sales data are unbalanced for several reasons. Most houses on the market are single-sale houses. Houses that have been sold more than once account for a small portion of the whole market in a typical dataset. Also, houses sold in one period can be quite different from those sold in other periods. These factors unbalance the pricing data and complicate econometric construction of a price index due to problems of heterogeneous,

---

<sup>1</sup>The recent literature has witnessed an upsurge of interest in studying real estate markets from perspectives of banking, financial and macroprudential policy. See, for example, the study of the relationship between real estate prices and banking instability (Koetter and Poghosyan, 2010; Reinhart and Rogoff, 2013), the market linkage among different assets (Chan et al., 2011), the impact of macroprudential policy on housing prices (Shi et al., 2013; Mendicino and Punzi, 2014), the role of housing markets for macroeconomy (Iacoviello, 2005; Musso et al., 2011).

missing, and unequally spaced observations. Third, a typical presumption underlying construction of real estate price indices is that the average quality of properties in the market remains constant over time, whereas quality improvements in housing occurs continuously from advances in materials, design, utilities, and construction technologies. Meanwhile and in spite of ongoing maintenance, older dwellings age with the holding period, leading to some depreciation in house value. These countervailing effects can produce ambiguities regarding what movements in a real estate price index reflect: the underlying market situation or quality changes in the properties that happen to be sold. This problem is exacerbated in a fast growing real estate market where a substantial proportion of sales are new sales released directly from developers.

Two main approaches dominate the literature of real estate price indices: the hedonic regression method and repeat-sales method. The hedonic method assumes that house values can be decomposed into bundles of utility-bearing attributes that contribute to the observed heterogeneity in prices. Observed house prices may then be regarded as the composite sum of elements that reflect implicit structural and locational prices (Rosen, 1974). Hedonic methods of estimating a real estate price index employ regression techniques to control for various sources of heterogeneity in prices using observations on covariates and dummy variables that capture relevant characteristics. However, the choice of the covariates in such hedonic regressions is limited by data availability and involves subjective judgements by the researcher, which may lead to model *specification bias*. Moreover, Shiller (2008) argued that the hedonic approach can lead to spurious regression effects in which the irrelevant hedonic variables are significant. A further complication is that the precise relationship between hedonic information and sales prices is unknown, likely to be complex, and may well be house dependent.

Unlike the hedonic approach, which uses all transaction prices to create an index, the repeat-sales method uses only properties that are sold multiple times in the sample to track market trends. The technique was first introduced for building the real estate price index by Bailey, Muth, and Nourse (1963) and then extended to include time-dependent error variances in seminal and highly influential work by Case and Shiller (1987, 1989). The repeat-sales method seeks to avoid the problem of heterogeneity by looking at the difference in sale prices of the same house. No hedonic variables are needed, so the approach avoids the difficulties of choosing hedonic information and specifying functional forms. However, since the repeat-sales method confines the analysis only to houses that have been sold multiple times, it is natural to question whether repeat-sales are representative of the entire market and whether there exists

significant *sample selection bias*. Clapp et al. (1991) and Gatzlaff and Haurin (1997) argued that the properties that are sold more than once could not represent the whole real estate market and the index estimated by the repeat-sales method is most likely subject to some sample selection bias. Moreover, large numbers of observations must be discarded because repeat-sales typically comprise only a small subset of all sales. Not surprisingly, the repeat-sales method has been criticized by researchers (e.g., Case et al., 1991; Nagaraja et al., 2010) for discarding too much data. On the other hand, while repeat-sales themselves may not be representative of the entire market, price changes in repeat-sales may still be representative of the market. Moreover, as argued in Shiller (2008), “there are too many possible hedonic variables that might be included, and if there are  $n$  possible hedonic variables, then there are  $n!$  possible lists of independent variables in a hedonic regression, often a very large number. One could strategically vary the list of included variables until one found the results one wanted.” As a result, Shiller (2008) made the strong claim that “the repeat-sales method is the only way to go” and this assertion has been influential. In the U.S., for instance, indices produced by the repeat-sales method, such as the FHFA and S&P/Case-Shiller home price indices, are now routinely reported in official government and industry statistics and they regularly attract media attention.

A combined approach, called the hybrid model, has been introduced as an alternative method of constructing house price indices. In particular, Case and Quigley (1991) proposed a hybrid model and applied generalized least squares (GLS) to jointly estimate the hedonic and repeat-sales equations. In subsequent work, Quigley (1995) and Englund et al. (1998) proposed to model explicitly the structure of the error terms in their hybrid model to improve the estimated price index. Hill et al. (1997) instead employed an AR(1) process to model the error dynamics of the hybrid model. Nagaraja, Brown and Zhao (2011) also relied on an underlying AR(1) model to build the hybrid model. To answer the question why hybrid models are better, Ghysels et al. (2012) explained that improved estimation in the hybrid model is analogous to the better forecasts gained by forecast combinations. The hedonic model has less sample selection bias but potentially greater specification bias, whereas the repeat-sales model has less specification bias but more sample selection bias. Ideally, some combination of the two might lead to an improved procedure of delivering an index that reduces both sample selection and specification bias.

With this goal in mind, the present paper proposes 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. This novel design offers four main advantages. First, the method makes use of all the real estate information in the sample, including both single-sale and repeat-sale homes. This approach contrasts with the use of just a small fraction of the sample that occurs in repeat-sales methods, thereby reducing both sample selection bias and information loss. With this design, the new real estate price index offers robustness against sample selection bias and gains in efficiency. Second, unlike standard hedonic models, 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. Both these features make the new index less susceptible to specification error than standard hedonic models. Third, 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. Consequently, pairs in our approach are typically more informative about price changes than those in repeat-sales methods. Finally, our model involves a simple and convenient GLS estimation procedure that is easy to implement and computationally efficient.

In triadic comparisons of out-of-sample predictions, the new index is found to give superior performance in predicting both repeat-sale home prices and single-sale home prices relative to the S&P/Case-Shiller index and the index constructed from a standard hedonic model. In dyadic comparisons, we find that the S&P/Case-Shiller index performs much better than the index from the hedonic model. These findings indicate that the specification bias in the standard hedonic method has more serious implications than the sample selection bias inherent in the S&P/Case-Shiller index, at least as far as the Singapore residential property market is concerned. When we test for explosive behavior in the three indices, we find evidence of earlier explosive behavior in our index than in the other indices. This finding has some important implications for macroprudential policy that are discussed in the paper.

The remainder of the paper is organized as follows. Section 2 develops the model and the estimation method. In Section 3, the method is applied to build a real estate price index for Singapore and out-of-sample performance of the alternative indices is compared. In Section 4 we test for explosive behavior in the index and the alternative indices using the recursive method of bubble detection developed recently in Phillips, Shi and Yu (2015a, 2015b). The results are discussed in the context of policy measures conducted by the Singapore government to cool the local real estate market. The Appendix

provides details of these policy cooling measures. Section 5 concludes. Throughout the paper we use the terminology ‘house’ to refer to an independent dwelling (apartment, flat, condominium, terraced, duplex, or free-standing) located within a specific building.

## 2 Model and Estimation

Let the log price per square foot for the  $j$ th sale of the  $i$ th house in building  $p$  be  $y_{i,j,p}$  and  $t(i, j, p)$  be the time when the  $i$ th house in building  $p$  is sold for the  $j$ th time. The model design given below in (1) seeks to explain  $y_{i,j,p}$  in terms of constituent components. In particular, we assume that the log price can be modeled as the sum of a log price index component, an unknown function of building level hedonic covariates, a location effect, an individual house effect, other individual house hedonic covariates, plus a partial sum of intervening building specific shocks, and a time-dependent error term. The log price index component is described by the parameter  $\beta_{t(i,j,p)}$ , which captures the time specific effect of house prices and is the primary parameter of interest. The building level hedonic information (whether observed or not) is denoted as  $Z_p$ ; and an unknown function  $f(Z_p)$  relates this building level information to the individual house price, capturing both observed and unobserved building level effects on price. The location effect is captured by a location variable  $\mu_p$ , which is assumed to be a fixed effect with respect to the location of the building  $p$ , which may well be correlated with covariates. The individual house effect is captured by  $h_{i,p}$ , which is assumed to be independent over  $i$  with mean zero and variance  $\sigma_h^2$ . The building specific shocks at time  $t$  are described by the random variables  $u_{t,p}$  which have mean zero and variance  $\sigma_u^2$ , and are assumed to be independent of each other across all buildings and for all time periods.

Suppose the total number of time periods (in quarters, say) is  $T$ . Then,  $t(i, j, p)$  belongs to the set  $\{1, \dots, T\}$ . When there is no confusion, we simply write  $t(i, j, p)$  as  $t$ . Let  $L$  be the total number of buildings. Then the model is formulated as

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

where  $X_{i,p}$  is the vector of covariates for the  $i$ th house in building  $p$ ,  $f$  is a nonparametric function of  $Z_p$ , and  $\epsilon_{i,j,p}$  are idiosyncratic shocks that are assumed to be  $iid(0, \sigma_\epsilon^2)$ . The covariates  $X_{i,p}$  capture the available house level hedonic information in the data.

The standard hedonic model (Ghysels et al., 2012) can be written as:

$$y_{i,j,z} = \mu_z + \beta_{t(i,j,z)} + \gamma' X_{i,z} + \epsilon_{i,j,z}, \quad (2)$$

where  $y_{i,j,z}$  is the log price per square foot for the  $j$ th sale of the  $i$ th house in area  $z$  and  $t(i, j, z)$  is the time when the  $i$ th house in area  $z$  is sold for the  $j$ th time. There are a few important differences between our model and the standard hedonic model which we now discuss.

There are still two restrictions implicit in model (2). First, a parametric form must be imposed to relate the observed building level covariates to the price. In model (2), a linear specification is adopted. However, any parametric specification is potentially invalid. Second, unobserved building level information cannot be accommodated in model (2). In the new model (1), building level hedonic information ( $Z_p$ ) is included nonparametrically (whether observed or not). Furthermore, individual house fixed effects are not included in the standard hedonic model as they cannot be consistently estimated. In the new model, individual house fixed effects,  $h_{i,p}$ , are included.

Since (1) contains more detailed building-level information than (2) as well as a semiparametric specification, the new model is less susceptible to specification bias. To see this, note that housing heterogeneity arises both at the individual building level and the individual house level. To capture heterogeneity at the building level, it is necessary to include all the relevant hedonic information in (2). Inevitably some covariates will be omitted in (2) due to data unavailability and latent variable effects. These covariates are generally correlated with the observed covariates and are absorbed into the error term,  $\epsilon_{i,j,z}$ , in (2). As a result,  $\epsilon_{i,j,z}$  is correlated with  $X_{i,z}$  in (2). Whereas, in the new model,  $f$  is left unspecified and  $Z_p$  can include all relevant building level information, observed or unobserved, that is related to the house price. Hence, (2) suffers potential specification bias from missing heterogeneity at the building level and from the use of a particular functional form.

Focusing on houses that have sold more than once, the repeat-sales method of Case and Shiller (1987, 1989) is based on the following model

$$y_{i,j,z} - y_{i,j-1,z} = \beta_{t(i,j,z)} - \beta_{t(i,j-1,z)} + \sum_{k=t(i,j-1,z)+1}^{t(i,j,z)} u_{i,z}(k) + \epsilon_{i,j,z} - \epsilon_{i,j-1,z}. \quad (3)$$

where  $u_{i,z}(k) \sim_{iid} N(0, \sigma_u^2)$  is the interval error at time  $t(i, j-1, z) + k$  for house  $i$  in area  $z$ . So the partial sum  $\sum_{k=t(i,j-1,z)+1}^{t(i,j,z)} u_{i,z}(k)$  is a Gaussian random walk and is used to model the concatenation of pricing shocks to this house between its  $j-1$ th and  $j$ th sale. Model ((3) may be motivated from the specification

$$y_{i,j,z} = \beta_{t(i,j,z)} + f(X_{i,z}) + \mu_z + \sum_{k=0}^{a_{t(i,j,z)}} u_{i,z}(k) + \epsilon_{i,j,z}, \quad (4)$$



where  $a_{t(i,j,z)}$  is house age at time  $t(i, j, z)$  for the  $i$ th house in area  $z$ . In this model, the functional form that captures the impact of hedonic information (whether it is observed or not) is  $f$ , which is left unspecified. For houses that have been sold multiple times in the sample, taking the difference of model (4) at two time stamps gives model (3) as both the hedonic covariates (both observed and unobserved) and the location effect are eliminated by differencing. Only houses that have been sold multiple times in the sample are retained in model (3). The model was estimated by Case and Shiller (1987, 1989) using a multi-stage method and led to the construction of the S&P/Case-Shiller real estate price index (S&P/Case-Shiller methodology report, 2009).

To facilitate estimation of our model, we take the average of equation (1) for all sales in the same building at each time period whenever there are sales. This yields

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

where  $\bar{y}_{t,p}$  is the average price of all transaction prices in building  $p$  at time  $t$  and  $t_1(p)$  is the time when the first sale in building  $p$  occurred. Similar to the Case-Shiller method, if there is another time period  $t' (> t)$  when the most recent transactions occur in the same building  $p$ , we have model (5) at time  $t'$ . Taking the difference of model (5) at these two time periods, we obtain

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

It is clear from Equation (6) that we create “pairs” at the building level at periods  $t$  and  $t'$ , and then match the average building price at  $t'$  against that at  $t$ , after taking account of the hedonic information at the individual house level and a building specific random walk effect.

There are three advantages in our method relative to the repeat-sales method. First, since the repeat-sales method only uses data on repeat-sales, it is assumed that price change in repeat-sales are representative of the whole market. In our model, the full sample is used to construct the index, including both single-sales and repeat-sales. As a result, the approach does not suffer from sample selection bias. Second, given that the full sample has been used, there are consequential efficiency gains compared with the use of a subsample of data, as in the repeat-sales model. Third, the time gap between  $t$  and  $t'$  in our approach is calculated on a building basis whereas the time gap in the repeat-sales method is based on houses. As a result, the time gaps that appear in our

approach are never bigger than and often much smaller than those in the repeat-sales method. Indeed, for a high percentage of cases,  $t' - t = 1$ , as in the empirical application considered later in the paper. Since both methods put more weights on pairs whose time gap is smaller, the pairs in our method turn out to be more informative than those in the repeat-sales methods.

The specification used in our approach based on model (6) is more detailed and complex than that of the repeat-sales model (3). But estimation of the new model is accomplished in the same manner as the method of Case and Shiller (1987, 1989) and is therefore a simple procedure to implement. The details of the required calculations are as follows.

1. Run an OLS regression of model (6) to obtain initial estimates of  $\beta_t$  for all  $t$  and  $\gamma$ .
2. Plug these initial estimates into (6) to calculate the regression residuals, denoted by  $\hat{e}_{t',p}$ , which are fitted values of the composite component  $\sum_{k=t+1}^{t'} u_{k,p} + \bar{h}_{t',p} - \bar{h}_{t,p} + \bar{e}_{t',p} - \bar{e}_{t,p}$ . Note that

$$E \left( \sum_{k=t+1}^{t'} u_{k,p} + \bar{h}_{t',p} - \bar{h}_{t,p} + \bar{e}_{t',p} - \bar{e}_{t,p} \right) = 0,$$

and

$$Var \left( \sum_{k=t+1}^{t'} u_{k,p} + \bar{h}_{t',p} - \bar{h}_{t,p} + \bar{e}_{t',p} - \bar{e}_{t,p} \right) = (t' - t)\sigma_u^2 + \left( \frac{1}{n_{t',p}} + \frac{1}{n_{t,p}} \right) (\sigma_h^2 + \sigma_\varepsilon^2), \quad (7)$$

because the building specific shocks, individual house effects and error terms are all independent of each other. In (7)  $n_{t,p}$  refers to the number of house sales transacted at time  $t$  in building  $p$ .

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

$$\hat{e}_{t',p}^2 = c + (t' - t)\sigma_u^2 + \left( \frac{1}{n_{t',p}} + \frac{1}{n_{t,p}} \right) (\sigma_h^2 + \sigma_\varepsilon^2) + v_{t',p}, \quad (8)$$

where  $E(v_{t',p}) = 0$ . Then the weights are the reciprocals of the fitted values from model (8). The diagonal matrix  $\widehat{W}$  with weights appearing in the main diagonal is then the estimated weight matrix for GLS estimation.

4. Using  $\widehat{W}$  as the weight matrix, GLS regression of (6) gives the final estimates of  $\beta_t$  for all  $t$  and  $\gamma$ . To be specific, we stack equation (6) into matrix form as

$$Y = Q\theta + e, \quad (9)$$

where  $\theta = [\beta' \ \gamma']'$ ,  $\beta$  is a  $T$ -dimensional coefficient vector with elements  $\beta_t$ ,  $Y$  is an  $N$ -dimensional vector with elements  $\bar{y}_{t',p} - \bar{y}_{t,p}$ ,  $N$  is the number of pairs in the building level, and  $Q = [D \ X]$ , where  $D$  is a selection matrix designed to capture the differential components  $\beta_{t'} - \beta_t$  in the model. The matrix  $D$  is constructed so that its  $n$ th row and  $t$ th column element has value  $-1$ , corresponding to the house price average in the previous period in the building level (viz.,  $\beta_t$ ) used at time  $t$ , and value 1 for the house price average in the current period in the building level (viz.,  $\beta_{t'}$ ) used at time  $t'$ , and value 0 otherwise. In the partition of  $Q$ ,  $X$  is a matrix with each row corresponding to elements of the form  $\bar{X}_{t',p} - \bar{X}_{t,p}$ . GLS applied to (9) gives the estimate

$$\hat{\theta} = (\hat{\beta}', \hat{\gamma}')' = (Q'\widehat{W}Q)^{-1}(Q'\widehat{W}Y),$$

whose components are used to extract the house price index.

### 3 Empirical Analysis

In this section, we apply the proposed model and the repeat-sales method to real estate price data involving quarterly transactions of private non-landed residential property sales in Singapore from Q1 1995 to Q2 2014. The period is of substantial interest given the fluctuations and growth in property prices in Singapore over this period and because of the extensive policy measures introduced by the government to cool the real estate market whose effectiveness can be gauged by empirical analysis of the real estate price indices.

There are mainly two residential property markets in Singapore: a private residential market and the public residential market that is managed by the Housing and Development Board (HDB). HDB is the statutory board of the Ministry of National Development and HDB flats are heavily subsidized by the Singapore government. Not surprisingly, the HDB market is largely segmented from the private residential market. Given its special nature and strong differentiation from the private market, we have excluded HDB transactions in the construction of the property market price index. The sample used for analysis therefore refers only to the private non-landed property

Table 1: Summary Statistics of Single-Sale Houses in Singapore

| Property Type    | No. Houses | Mean | Sd  | Min | Max  |
|------------------|------------|------|-----|-----|------|
| Apartments       | 40,097     | 1177 | 620 | 154 | 5146 |
| Condominiums     | 106,073    | 947  | 459 | 156 | 6393 |
| 99 years tenure  | 81,086     | 939  | 446 | 154 | 5000 |
| 999 years tenure | 6864       | 884  | 375 | 233 | 2695 |
| Freehold         | 58,220     | 1125 | 600 | 202 | 6393 |
| All              | 146,170    | 1010 | 519 | 154 | 6393 |

Table 2: Summary Statistics of Repeat-Sales Houses in Singapore

| Property Type    | No. Houses | Mean | Sd  | Min | Max  |
|------------------|------------|------|-----|-----|------|
| Apartments       | 20,618     | 901  | 455 | 137 | 4700 |
| Condominiums     | 49,715     | 850  | 404 | 94  | 4820 |
| 99 years tenure  | 33,554     | 864  | 366 | 94  | 4700 |
| 999 years tenure | 4674       | 864  | 317 | 197 | 2491 |
| Freehold         | 32,105     | 985  | 454 | 183 | 4820 |
| All              | 70,333     | 865  | 420 | 94  | 4820 |

market.<sup>2</sup>

The data source for private house information is the Urban Redevelopment Authority (URA),<sup>3</sup> which is Singapore's urban planning and management authority. The URA property market dataset provides extensive records of information for all transactions in the property market. The sale price (both the total price and the price per square foot) and the transaction period are reported. The district, sector and postal code of every transacted property are also recorded. Other characteristics include floor and unit number, project number, size, sell type, property type, completion year, tenure length, and location type.

During the sample period our data include some 315,000 transactions and the number of the dwellings involved is around 216,000.<sup>4</sup> Among these, about 146,000 houses are single-sales and the remainder, about 70,000 houses, are ones that sold more than once. The number of pairs for repeat-sales is around 97,000. So single-sales dominate repeat-sales in the sample in terms of the number of houses. In addition, the total number of buildings  $L$  is 4820<sup>5</sup>, which leads to around 81,000 pairs at the building

<sup>2</sup>Non-landed residential property is the largest and most popular housing form in Singapore, constituting more than 75% of private residential units in the market by Q2 2014.

<sup>3</sup><http://www.ura.gov.sg/>

<sup>4</sup>We delete houses with incomplete information on characteristics. Sales that occur less than a quarter after the previous sale of the same house are also excluded.

<sup>5</sup>We delete buildings in which only one transaction occurs during the whole sample period. The number of buildings deleted is around 300, which implies only 300 single-sale houses are deleted. The

level.

There are two types of private non-landed residential properties in the Singapore real estate market: apartments and condominiums. The main difference between them is that condominiums are equipped with full facility but apartments may not be (Sing, 2001). The total number of condominium houses in our sample is around 155,000 and apartments account for some 60,000. In addition, in terms of ownership type, there are freehold, 999-year leasehold and 99-year leasehold. Most private residential properties transacted in the sample are either freehold or 99-year leasehold. Freehold houses are more expensive than 99-year leasehold houses. We have postal district information in our database which is used to identify house location and zipcode information which is used to identify individual buildings.<sup>6</sup> Table 1 and Table 2 provide summary statistic information on the sample.

The dataset is well-suited to compare our new method with the standard hedonic method and the S&P/Case-Shiller repeat-sales method for index construction. First, we have the complete record of all transactions and the sample size of total sales is large, enabling us to estimate the proposed model accurately. With estimation error being small, attention can focus on comparing the indices constructed by different methods. Second, the hedonic information in the data is extensive so that many variables and alternative specifications can be included on the right hand side of models (2) and (1). Third, there are a very large number of repeat-sales in the data, so that model (3) can also be estimated accurately. Consequently, we can ignore estimation errors and focus on comparing the out-of-sample performance of different methods. By doing so, we can evaluate the relative magnitude of the price implications of implicit specification bias and sample selection bias in the three methods.

It is worth noting that single-sale properties display different summary statistics from repeat-sales properties. The mean price and the standard deviation for repeat-sale houses is lower than single-sale houses across all categories. This observation seems to support the argument that repeat-sale houses are not a representative random sample of the entire market and may carry a sample selection bias. Furthermore, in spite of the long sample period, about 68% of houses in the sample that have changed hands are single-sale houses. So the repeat-sale method is based on only about 32% of the houses in the sample.

---

loss of information is negligible given that we have around 146,000 single-sale houses in the dataset.

<sup>6</sup>There are 27 postal districts and 69 postal sectors in the sample. In Singapore each building is assigned a unique zip code. This location and zipcode information is directly retrievable from the database.

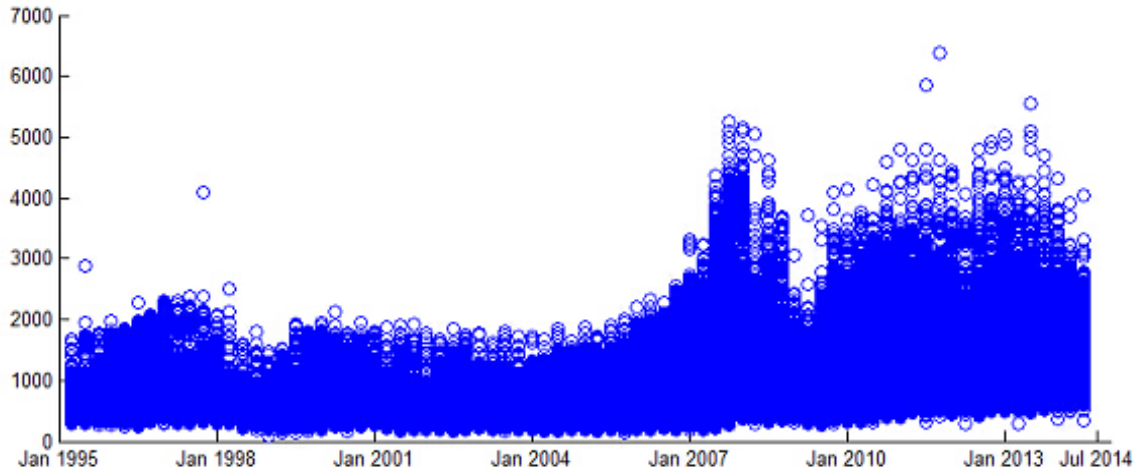


Figure 1: Scatter plots of house prices per square foot over January 1995 - June 2014

The scatter plot of all house prices per square foot over time is given in Figure 1. It is difficult to discern price trends from this scatter plot, especially for houses at the low-end of the market because of the density of the data points. For high-end houses, at least, prices seem to be more stable between 2000 and 2006 than during other periods.

To fit the model in equation (6), we take account of the following two property characteristics: building zipcode and transaction period. Zipcode information in our database is used to identify buildings. The real estate price index is given by the parametric sequence  $\{\beta_t\}$ , which delivers the quarterly index from Q1 1995 to Q2 2014 (78 quarters in total). To keep our model as parsimonious as possible in this application, we do not use other hedonic covariates in our empirical analysis and hence the model has the form

$$\bar{y}_{t',p} - \bar{y}_{t,p} = \beta_{t'} - \beta_t + \sum_{k=t+1}^{t'} u_{k,p} + \bar{h}_{t',p} - \bar{h}_{t,p} + \bar{\epsilon}_{t',p} - \bar{\epsilon}_{t,p}. \quad (10)$$

The model can be easily expanded to include additional hedonic information as covariates. We have experimented with other covariates in our dataset and the main empirical findings reported here are qualitatively unchanged. So, for simplicity, we only report results obtained from the above specification.

We follow the estimation procedure described in Section 2 to obtain  $\{\hat{\beta}_t\}$ . Since our purpose is to construct the house price index itself, rather than its logarithm, it is convenient to use the parameterization in Nagaraja, Brown and Zhao (2011) and

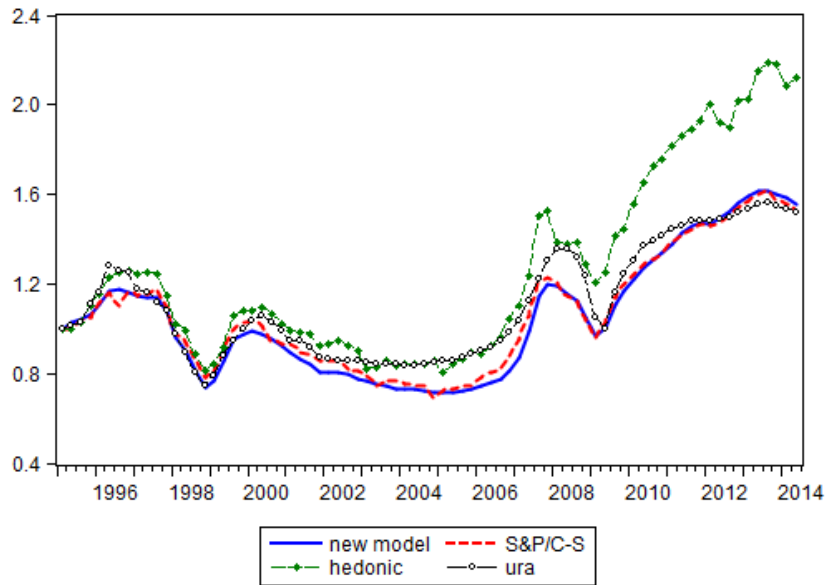


Figure 2: Four Real Estate Price Indices for Singapore: Q1 1995 – Q2 2014

calculate  $\widehat{I}_t = \exp(\widehat{\beta}_t)$ .<sup>7</sup> Finally, we take the first quarter in our sample as the reference point for which the price index is set to unity.

For comparison, we apply the hedonic method to all transaction prices and the S&P/Case-Shiller method to repeat-sales prices to build the indices.<sup>8</sup> We plot the proposed index, the S&P/Case-Shiller index, the standard hedonic index and the URA private non-landed residential property price index created by the Urban Redevelopment Authority (URA) in Figure 2.<sup>9</sup> As is apparent in the figure, there are some substantial discrepancies among the four indices. In particular, the standard hedonic index is more elevated and appears more volatile than the other indices and seems to diverge from the other indices towards the end of the sample period. This discrepancy may be due to the index's greater susceptibility to specification bias, a possibility that becomes clearer in the out-of-sample analysis below. Also, the URA index has different

<sup>7</sup>Although  $\widehat{I}_t$  is biased downward for  $I_t$ , the biased corrected estimator leads to virtually no change in our results since the estimation error (and hence the variance estimate that appears in the bias calculation) is small.

<sup>8</sup>We employ the following four property characteristics in the hedonic model: location, transaction periods, property type, and ownership type to construct the hedonic index which is displayed in Figure 2. We have experimented with other covariates in our dataset and the main empirical findings reported here are qualitatively unchanged when additional covariates are included.

<sup>9</sup>Since the exact methodology of URA is not sufficiently clear for reproduction, we cannot include the URA index in our out-of-sample exercise.

turning points from the other three indices. For example, over the period of the global financial crisis, the turning point in the middle 2008 suggested by the URA index is two quarters later than that implied by the other three indices; and the turning point at the beginning of 2009 suggested by the URA index is one quarter later than that implied by the other three indices. Interestingly, our new index and the S&P/Case-Shiller index are very close to each other although our index suggests a longer trough in prices following the outbreak of SARS.

To compare the new index, the standard hedonic index and the S&P/Case-Shiller index and to examine the price implications of the specification bias and sample selection bias, we investigate the out-of-sample predictive power of the three indices.<sup>10</sup> To do so, we divide the observations into training and testing datasets. 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 placed into the testing set with probability 0.04. We also randomly add the single-sale houses into our testing set with probability 0.24, so that the testing set contains the same number of single-sale houses and repeat-sale houses.<sup>11</sup> All the remaining houses are included in the training 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.

We first estimate all models based on the training set and then examine their out-of-sample predictive power on the testing set. Before analyzing the findings, we first explain how price predictions of the repeat-sale homes are obtained using the alternative indices. To calculate the predicted prices of the repeat-sale homes using the new method, we use

$$\hat{Y}_{t',i,p} = \frac{\hat{I}_{t'}^{bb}}{\hat{I}_t^{bb}} \bar{Y}_{t,p}, \quad (11)$$

where  $\hat{Y}_{t',i,p}$  is the price per square foot for house  $i$  in building  $p$  at time  $t'$ ,  $\hat{I}_t^{bb}$  is the estimated index from the new model at time  $t$ ,  $t$  is the time period of the previous sale in building  $p$ , and  $\bar{Y}_{t,p}$  is the average price per square foot for building  $p$  at time  $t$  in the training set.

For the S&P/Case-Shiller model, given that all single-sales are deleted, we use

<sup>10</sup>We evaluate the indices by their out-of-sample predictive power rather than their in-sample fitting because out-of-sample performance is more important in the context of specification testing. It is also well-known that that good in-sample fits often translate into poor out-of-sample predictions (for a recent discussion, see e.g. Hansen, 2010).

<sup>11</sup>To compare the out-of-sample predictive power of three indices on single sale houses, the test set does not include the single sale houses which are transacted as the first sales in their building. So the single sale houses, which are sold in the same period as the first sale in the building, are automatically included in the training set.



$$\hat{Y}_{t',i} = \frac{\hat{I}_{t'}^{cs}}{\hat{I}_t^{cs}} Y_{t,i}, \quad (12)$$

where  $Y_{t,i}$  is the price per square foot for house  $i$  at time  $t$ ,  $t' > t$  and  $\hat{I}_t^{cs}$  is the estimated S&P/Case-Shiller index at time  $t$ , and  $t$  is the time period of the previous sale for house  $i$  (which is typically much smaller than  $t$  in equation (11)).

It should be pointed out that the predictive equations (11) and (12) are implied by models (10) and (3), respectively. From model (10), the predictive value of the average log price for building  $p$  at time  $t'$  can be represented as

$$\hat{y}_{t',p} = \bar{y}_{t,p} + \hat{\beta}_{t'} - \hat{\beta}_t.$$

When converting the log price to price, the predictive value of the average price for building  $p$  at time  $t'$  is

$$\hat{Y}_{t',p} = \exp \{ \hat{y}_{t',p} \} = \exp \{ \bar{y}_{t,p} + \hat{\beta}_{t'} - \hat{\beta}_t \} = \frac{\exp \{ \hat{\beta}_{t'} \}}{\exp \{ \hat{\beta}_t \}} \exp \{ \bar{y}_{t,p} \} = \frac{\hat{I}_{t'}^{bb}}{\hat{I}_t^{bb}} \bar{Y}_{t,p}$$

where  $\bar{Y}_{t,p}$  is the geometric mean price per square foot for building  $p$  at time  $t$  in the training set. We take this predictive value  $\hat{Y}_{t',p}$  as the predictive value for house  $i$  in building  $p$  at time  $t'$ , that is  $\hat{Y}_{t',i,p}$ . In a similar way, we can derive equation (12) from (3).

For the standard hedonic model, we plug the estimated parameters into model (2) to obtain

$$\hat{y}_{i,j,z} = \hat{\mu}_z + \hat{\beta}_{t(i,j,z)} + \hat{\gamma}' X_{i,z} \quad (13)$$

where  $\hat{y}_{i,j,z}$  is the predicted log price for the  $j$ th sale of house  $i$  in area  $z$  and  $\hat{\mu}_z$  is the estimated location dummy variable coefficient for area  $z$ . We then follow Nagaraja, Brown and Zhao (2011) to convert the log price into price by means of the transform

$$\hat{Y}_{i,j,z} = \exp \left\{ \hat{y}_{i,j,z} + \frac{MSR}{2} \right\} \quad (14)$$

where  $MSR = \frac{1}{M} \sum_{i=1}^M (y_{i,j,z} - \hat{y}_{i,j,z})^2$ , the mean square residuals and  $M$  is the total number of transactions to fit the model.

All three predictive prices are matched against the actual prices observed in the testing set. The root mean squared error (RMSE) and the mean absolute error (MAE)

are reported in Table 3. Several important findings emerge. First, the S&P/Case-Shiller index performs much better than the standard hedonic index. In particular, compared with the standard hedonic method, the S&P/Case-Shiller method reduces the RMSE by around 40% and reduces the MAE by about 45%. In economic terms, the reduction in the MAE means that the repeat-sales method leads to a reduction of nearly \$100 (per square foot) in pricing error. This number compares to, as reported in Table 2, the mean price of all repeat-sales homes of \$865. Clearly the improvement is economically highly significant. These findings suggest that the sample selection bias present in the repeat-sales method is much less serious than the specification bias in the standard hedonic method, at least as far as house price prediction is concerned. Although we reported evidence that repeat-sales houses are not a representative random sample of the entire market in Singapore, the good out-of-sample performance of the S&P/Case-Shiller index suggests that perhaps the price changes in repeat-sales homes reflects well the changes that occur in single-sale homes.

Second and more importantly, the new model clearly has the best predictive power for repeat-sale homes. In particular, compared with the S&P/Case-Shiller, our model reduces the RMSE by around 19% and reduces the MAE by about 25%. Compared with the standard hedonic model, our model reduces the RMSE by around 52% and reduces the MAE by about 58%. In the economic terms, these reductions in the MAE imply that the the new method leads to a pricing error reduction of \$30 (per square foot) relative to the repeat-sale method and \$130 (per square foot) relative to the standard hedonic method. All these reductions are substantial. At first glance, it may be surprising that the new model outperforms the repeat-sale method for predicting repeat-sales homes because the two indices are close to each other as shown in Figure 2. The superiority of the new method can be explained as follows. When we predict prices of repeat-sale homes, based on the specification of the new model, the average price of the most recent sales of all homes in the same building are used. However, based on the specification of the S&P/Case-Shiller model, one can only use the most recent sale price of the same home, which because of time lags may not reflect the present market as well. Indeed, the time gap in the latter case is usually much larger than the former case, making the most recent sale price of the same home far less relevant for prediction than the average price of the most recent sales of all homes in the same building.

Figures 3 and 4 plot the histograms for these two sets of time gaps and report the mean, median and standard deviation of the gap time in each case. Apparently, in the new method with probability of around 80% the gap time is 1 or 2 periods with median of 1 period and standard deviation of 2.75. In the repeat-sale method, the

Table 3: 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     |

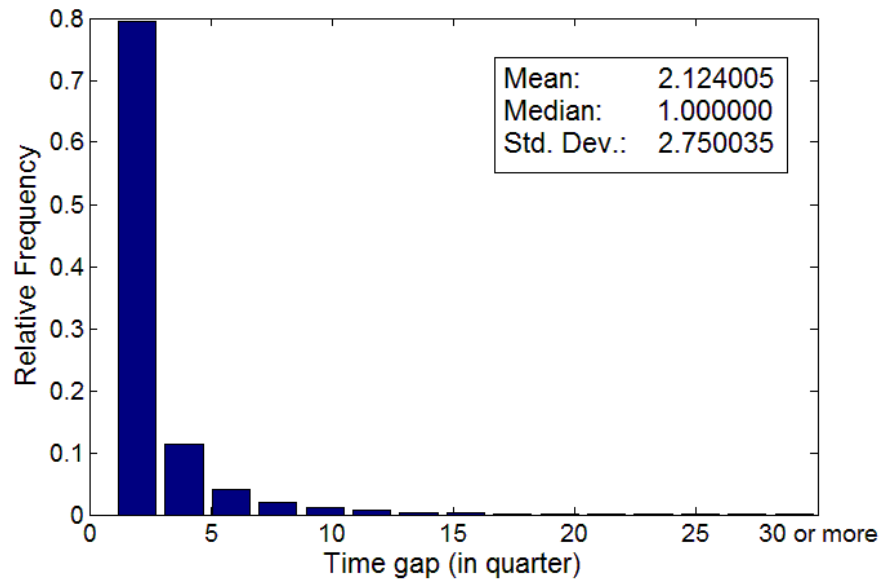


Figure 3: Histogram, mean, median and standard deviation of the time gap of sales in the same building.

distribution of the gap time is much more dispersed with median of 15 periods and standard deviation of 15.48. The average price of all sale prices in the same buildings last quarter can be expected to be far more informative in predicting prices in the current period than the price of the same house 15 periods ago.

Next we discuss how to predict prices of single-sale homes using the alternative indices. Since the S&P/Case-Shiller method discards all single-sale information, we cannot use this method to predict the price of single-sale homes. We therefore compare the predictive power of the new model with the standard hedonic model in this case. As before, we use equation (11) in our model and equation (13) and (14) in the hedonic model. The RMSE and MAE are shown in Table 4. Again, the new model performs much better in predicting prices of the single-sale homes than the standard hedonic model. Our model reduces the RMSE by around 48% and reduces the MAE by about 54%.

We can also compare the out-of-sample performance of our new model and the

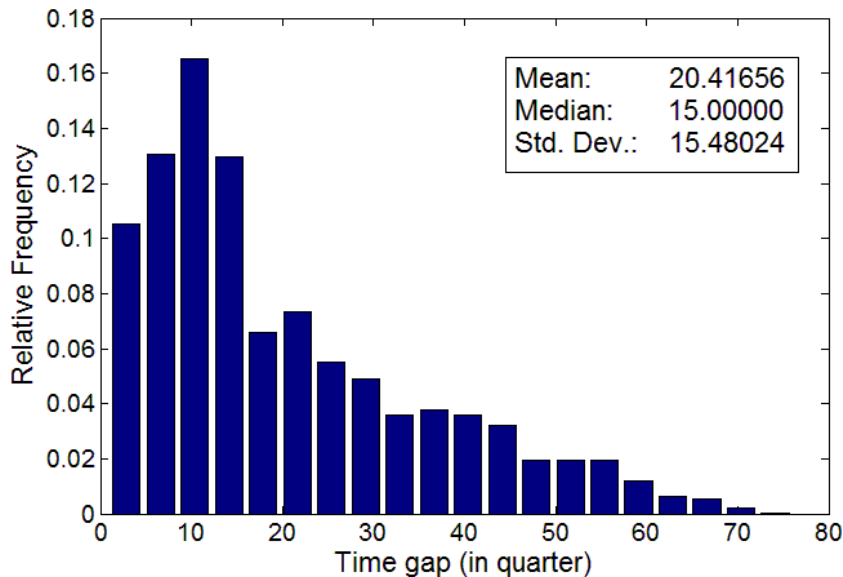


Figure 4: Histogram, mean, median and standard deviation of the time gap of sales of the same house.

standard hedonic model on all houses in the testing set. The RMSE and the MAE are shown in Table 5. Our model reduces the RMSE by around 50% and reduces the MAE by about 56%.

Table 4: 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     |

Based on this out-of-sample analysis, it is clear that the standard hedonic model suffers from serious specification bias. Two sources of specification bias are expected. First, the attributes of houses or the factors that influence the house price are too many to be recorded in the data, leading to the problem of omission of relevant variables. Second, when covariates are observed, their exact relationship with the house price is almost always unknown and the use of a parametric form is potentially misspecified.

Moreover, the out-of-sample analysis also tells us that discarding single-sale houses from the analysis leads to a significant loss of information for prediction. This is because past prices of single-sale houses in the same building carry useful information. That explains why our new model increases the predictive power considerably relative to the S&P/Case-Shiller even though the two indices appear not to differ so much. To further

illustrate this point, we consider a hypothetical (and infeasible) exercise, in which the single-sale houses are not eliminated from the prediction exercise and we predict the price in the testing set with our method and the repeat-sales method. With the repeat-sales method, we use the following fabricated equation (15) to calculate the predictive price

$$\hat{Y}_{t',i,p} = \frac{\hat{I}_{t'}^{cs}}{\hat{I}_t^{cs}} \bar{Y}_{t,p}, \quad (15)$$

where  $\hat{Y}_{t',i,p}$  is the price per square foot for house  $i$  in building  $p$  at time  $t'$ ,  $\hat{I}_t^{cs}$  is the estimated S&P/Case-Shiller index at time  $t$  and  $t$  is the time period of the previous sale in building  $p$ , and  $\bar{Y}_{t,p}$  is the average price per square foot in building  $p$  at time  $t$  in the training set. There are two main differences between equation (15) and (12). The first difference is that  $\bar{Y}_{t,p}$  is used to estimate  $\hat{Y}_{t',i,p}$  in (15) instead of  $Y_{t,i}$  in (12). This allows us to predict prices of all houses in the testing set. Whereas (12) is only applicable to the repeat-sales houses. Secondly,  $t$  in (15) is the time period of the previous sale in building  $p$  whereas  $t$  in (12) is the time period of the previous sale of house  $i$ . As a result, for the same house  $i$ , the time period of the previous sale in building  $p$  is potentially much closer to  $t'$  than that of the previous sale of house  $i$ , even for repeat-sales homes. In our new model and the Case-Shiller model, more recent sales are informative due to the random walk component. Equation (15) is infeasible for prediction in the Case-Shiller model because the single-sale data have been removed by the S&P/Case-Shiller method. We do this hypothetical comparison only to explain the usefulness of the most recent sales in the same building for prediction.

The RMSE and the MAE from the two models are reported in Table 6 when we only predict prices of single-sale houses in the testing set. Tables 7, 8 give the results when only repeat-sale houses are predicted and all houses are predicted, respectively. By incorporating the information of the most recent sale prices in the same building, both the RMSE and MAE generated by the S&P/Case-Shiller index are substantially reduced. Consequently, although the predictive power of our new model is still slightly better than the S&P/Case-Shiller model, the outperformance in this case (here evident in MAE) is only marginal because of the use of additional information (infeasibly) in the S&P/Case-Shiller index.

The out-of-sample analysis suggests that our new model captures the overall housing market situation in Singapore better than both the standard hedonic method and the repeat-sales method. As demonstrated before, our new method utilizes all the information, is robust to specification bias, and performs best in out-of-sample analysis.

Table 5: Testing set (all houses included) RMSE &amp; MAE for the Indices (SG dollars)

| Loss Function | new model | hedonic |
|---------------|-----------|---------|
| RMSE          | 149       | 294     |
| MAE           | 89        | 204     |

Table 6: The hypothetical exercise – Testing set (only single sale houses included) RMSE &amp; MAE for the Indices (SG dollars)

| Loss Function | new model | S&P/Case-Shiller |
|---------------|-----------|------------------|
| RMSE          | 156       | 156              |
| MAE           | 86        | 87               |

Table 7: The hypothetical exercise – Testing set (only repeat sales houses included) RMSE &amp; MAE for the Indices (SG dollars)

| Loss Function | new model | S&P/Case-Shiller |
|---------------|-----------|------------------|
| RMSE          | 141       | 141              |
| MAE           | 92        | 93               |

Table 8: The hypothetical exercise – Testing set (all houses included) RMSE &amp; MAE for the Indices (SG dollars)

| Loss Function | new model | S&P/Case-Shiller |
|---------------|-----------|------------------|
| RMSE          | 149       | 149              |
| MAE           | 89        | 90               |

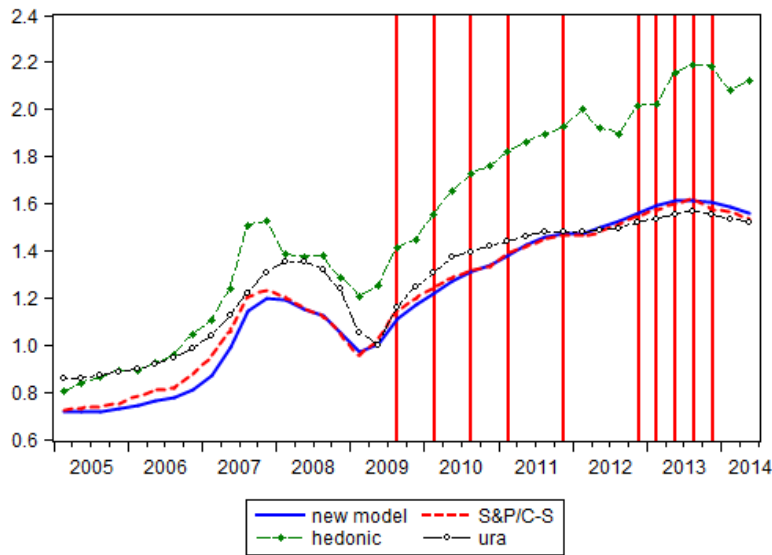


Figure 5: Four real estate price indices and the dates of ten rounds of successive macro-prudential cooling measures (indicated by vertical lines).

Moreover, the procedure is very convenient to implement in practical work.

## 4 Cooling Measures and Explosive Behavior

Housing is a highly important sector of the economy and provides the largest form of savings of household wealth in Singapore. Property prices play an important role in consumer price inflation and can therefore have a serious impact on public policy. The private housing sector, property prices and rents also impact measures of Singapore's competitiveness in the world economy. For these and other reasons, the Singapore government has closely watched movements in housing prices over the last decade and particularly since the house price bubble in the USA. Recently, Singapore implemented ten successive rounds of macro-prudential measures intended to cool down the housing market. These measures were undertaken between September 2009 and December 2013, the first eight of which were targeted directly at the private residential market.

The Appendix summarizes the dates and the nature of these macro-prudential measures. As is evident, a variety of macro-prudential policies have been used by the Singapore government. These include introducing a Seller's Stamp Duty (SSD), lowering the Loan-to-Value (LTV) limit, introducing an Additional Buyer's Stamp Duty (ABSD), and reducing the Total Debt Servicing Ratio (TDSR) and the Mortgage Ser-

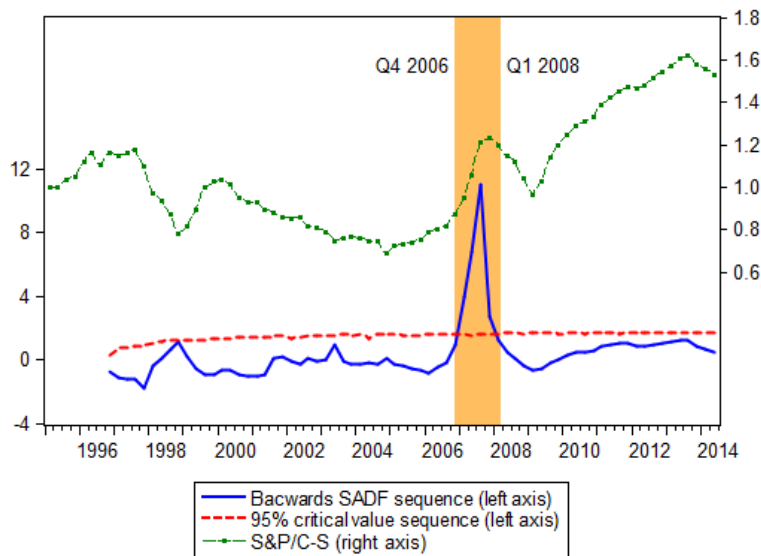


Figure 6: Testing for Bubbles in Singapore Real Estate Prices: using the S&P/Case-Shiller index, the BSADF statistic of PSY and the critical values.

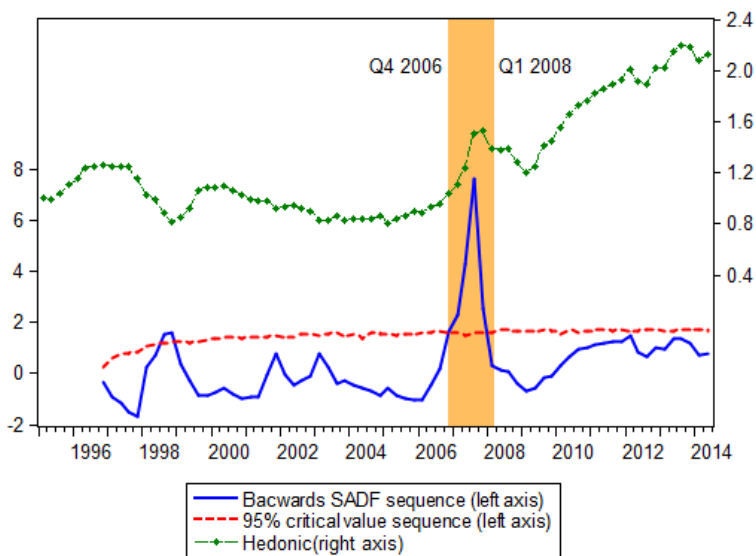


Figure 7: Testing for Bubbles in Singapore Real Estate Prices: using the index from the hedonic model, the BSADF statistic of PSY and the critical values.



ving Ratio (MSR). To visualize the impact of these cooling measures on the dynamics of real estate price movements, Figure 5 plots the four price indices for the period between Q1 2008 and Q2 2014, superimposed by vertical lines indicating the introduction of these ten cooling measures.

The primary goal of the macro-prudential policies is to reduce or eliminate emergent price bubbles in the real estate market and bring prices closer in line with fundamental values. Shi et al. (2013) and Mendicino and Punzi (2014) examined the impact of macro-prudential policies on real estate prices. Using the present value model, Diba and Grossman (1988) showed the presence of a rational bubble solution that implies that an explosive behavior in the observed price. If fundamental values are not explosive, then the explosive behavior in prices is a sufficient condition for the presence of bubble. Phillips, Wu and Yu (2011) and Phillips, Shi and Yu (2014a, 2014b, PSY hereafter) introduced recursive and rolling window econometric methods to test for the presence of mildly explosive behavior or market exuberance in financial asset prices. These methods also facilitated estimation of the origination and termination dates of explosive bubble behavior. The method of Phillips, Wu and Yu (2011) is particularly effective when there is a single explosive episode in the data while the method of PSY can identify multiple explosive episodes. In the absence of prior knowledge concerning the number of explosive episodes, in what follows we use the PSY method to assess evidence of bubbles in real estate prices.

Bubble behavior and market exuberance and collapse are subsample phenomena. So, PSY proposed the use of rolling window recursive application of right sided unit root tests (against explosive alternatives) using a fitted model for data  $\{X_t\}_{t=1}^n$  of the following form

$$\Delta X_t = \hat{\alpha} + \hat{\beta} X_{t-1} + \sum_{i=1}^K \hat{\beta}_i \Delta X_{t-i} + \hat{e}_t. \quad (16)$$

Details of the procedure and its asymptotic properties are given in PSY. We provide a synopsis here and refer readers to PSY for further information about the specifics of implementation and the procedure properties. Briefly, the unit root test recursion involves a sequence of moving windows of data in the overall sample that expands backward from each observation  $t = \lfloor nr \rfloor$  of interest, where  $n$  is the sample size and  $\lfloor nr \rfloor$  denotes the integer part of  $nr$  for  $r \in [0, 1]$ . Let  $r_1$  and  $r_2$  denote the start and end point fractions of the subsample regression. The resulting sequence of calculated unit root test statistics are denoted as  $\{ADF_{r_1}^{r_2}\}_{r_1 \in [0, r_2 - r_0]}$  where  $r_0$  is the minimum window size used in the recursion. and  $t = \lfloor Tr \rfloor$  is the point in time for which we intend to test for normal market behavior against exuberance. PSY define the recursive statistic

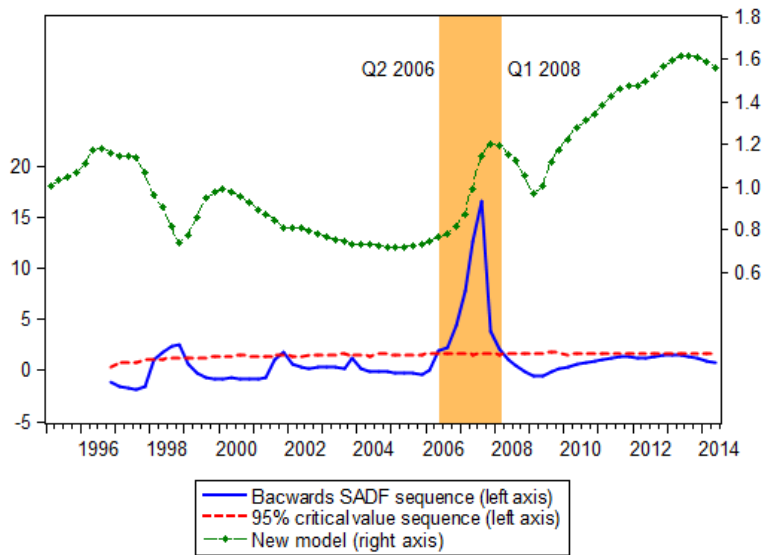


Figure 8: Testing for Bubbles in Singapore Real Estate Prices: using the new index, the BSADF statistic of PSY and the critical values.

$BSADF_r = \sup_{r_1 \in [0, r_2 - r_0], r_2 = r} \{ADF_{r_1}^{r_2}\}$ . The origination and termination dates of an explosive period are then determined from the crossing times

$$\hat{r}^e = \inf_{r \in [r_0, 1]} \{r : BSADF_r > cv\} \text{ and } \hat{r}^f = \inf_{r \in [\hat{r}^e, 1]} \{r : BSADF_r < cv\}, \quad (17)$$

where the recursive statistic  $BSADF$  crosses its critical value  $cv$ . The quantity  $\hat{r}^e$  estimates the origination date of an explosive period and  $\hat{r}^f$  estimates the termination date of an explosive period. After the first explosive period is identified, the same method may be used to identify origination and termination dates of subsequent explosive episodes in the data.

To assess evidence for potential bubbles in the private real estate market in Singapore, we applied the PSY method first to both the S&P/Case-Shiller index and the index built from the hedonic model with minimum rolling window size  $r_0 = 8$ , corresponding to two years. Figures 6 and 7 report the two indices, the corresponding  $BSADF$  statistics and the 5% critical values, respectively. The (orange) shaded area corresponds to the explosive period where the  $BSADF$  statistic exceeds the critical value. The PSY method identifies an explosive period, namely Q4 2006 to Q1 2008, in both the S&P/Case-Shiller index and the index built from the hedonic model.

We also applied the PSY method to our new index with minimum rolling window size  $r_0 = 8$ . Figure 8 reports the index, the test recursion, and the test 5% critical

values. PSY identifies an explosive period in the private real estate market over Q2 2006 to Q1 2008. 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, a finding that can have important practical implications for policy.

During the period 2006 - 2008, no cooling measures were introduced by the government. If the government had been alerted to the existence of exuberant market conditions in real time during this period, the opportunity would have been available for the implementation of cooling measures to affect the market. If the Case-Shiller index had been used, the government may have been stimulated to introduce cooling measures in Q4 2006, whereas if the new index were available and acted upon, the government may have introduced cooling measures earlier in Q2 2006. Moreover, although all three indices suggest that there were upward movements in price following 2008, between 2009 and 2013, these movements are not determined to be explosive and the PSY detector indicates little or no evidence of explosive behavior after 2009. This tapering in real estate market exuberance coincides with the period September 2009 through December 2013 during which macro-prudential cooling measures were actually implemented by the government and therefore appear to have been effective.

## 5 Conclusion

In order to exploit all available information in real estate markets, this paper provides a new methodology for the estimation of real estate price indices. The proposed new model has some of the advantages of the standard hedonic method as it uses both single-sales and repeat-sales data but it is less prone to specification bias than the standard hedonic model. Moreover, it generalizes the attractive feature of the repeat-sales method by creating sale pairs from within the individual building level, thereby increasing the number of observations used in the index. The model is also easy to estimate. Unlike the maximum likelihood methods of Hill, Knight and Sirmans (1997) and Nagaraja, Brown and Zhao (2011), this approach uses GLS estimation and is computationally efficient with large datasets. Other methods have been suggested to construct sale pairs in the literature – see, for example, McMillen (2012), and Guo et al (2014). Our matching rule is simpler to implement and has the advantage of a semiparametric nature.

We apply our estimation procedure to the real estate market for private residential dwellings in Singapore and examine the model's out-of-sample predictive performance in comparison with indices produced using the repeat-sales methodology of Case and

Shiller (1987, 1989) and the standard hedonic method. The findings reveal that, compared with these alternative methodologies, our method has superior performance out-of-sample. We expect our method is well suited to build real estate indices for high density cities where houses are mainly project-based. Each project contains a number of buildings with many units. These units share essentially the same location, facility, design, developer ownership, and utilities, among other common features. In theory, our method can also be applied for single-family homes as long as we can define suitable groups (such as estates) for single-family homes and create sale pairs from the group level. Another useful idea is to use other simple criteria to choose pairs – see Baltagi and Li (2015), for instance, for the use of housing projects. These ideas will be investigated in the future work.

The recursive detection method of Phillips, Shi and Yu (2015a, 2015b) is applied to each of the indices to locate episodes of real estate price exuberance in Singapore. While for all three indices PSY identifies the same bubble, the bubble origination date in the new index comes two quarters earlier than that in the other two indices. Although all three indices grew during 2009 - 2013, the expansion is not explosive, indicating that the ten recent rounds of cooling measure intervention in the real estate market conducted by the Singapore government have been successful in controlling prices.

## Appendix

Dates and the content of recent real estate market cooling measures introduced in Singapore.

1. 2009/9/14

- Reinstatement of the confirmed list for the 1st half 2010 government land sales programme
- Removal of the interest absorption scheme and interest-only housing loans
- Non-extension of the January 2009 budget assistance measures for the property market

2. 2010/2/20

- Introduction of a Seller's Stamp Duty (SSD) on all residential properties and lands sold within one year of purchase

- Loan-to-Value (LTV) limit lowered from 90% to 80% for all housing loans

3. 2010/8/30

- Holding period for imposition of SSD increased from one year to three years
- Minimum cash payment increased from 5% to 10% and the LTV limit decreased to 70% for buyers with one or more outstanding housing loans
- The extended SSD does not affect HDB lessees as the required Minimum Occupation Period for HDB flats is at least 3 years

4. 2011/1/14

- Increase the holding period for imposition of SSD from three years to four years
- Raise SSD rates to 16%, 12%, 8% and 4% for residential properties sold in the first, second, third and fourth year of purchase respectively
- Lower the LTV limit to 50% on housing loans for property purchasers who are not individuals
- Lower the LTV limit on housing loans from 70% to 60% for second property

5. 2011/12/8

- Introduction of an Additional Buyer's Stamp Duty (ABSD)
- Developers purchasing more than four residential units and following through on intention to develop residential properties for sale would be waived ABSD

6. 2012/10/6

- Mortgage tenures capped at a maximum of 35 years
- For loans longer than 30 years or for loans that extend beyond retirement age of 65 years: LTV lowered to 60% for first mortgage and to 40% for second and subsequent mortgages
- LTV for non-individuals lowered to 40%

7. 2013/1/12

- Higher ABSD rates

- Decrease the LTV limit for second/third loan to 50/40% from 60%; non-individuals' LTV to 20% from 40%
  - Mortgage Servicing Ratio (MSR) for HDB loans now capped at 35% of gross monthly income (from 40%); MSR for loans from financial institutions capped at 30%
8. 2013/6/28: Introduction of Total Debt Servicing Ratio (TDSR). The total monthly repayments of debt obligations should not exceed 60% of gross monthly income.
9. 2013/8/27
- Singapore Permanent Resident (SPR) Households need to wait three years, before they can buy a resale HDB flat
  - Maximum tenure for HDB housing loans is reduced to 25 years. The MSR limit is reduced to 30% of the borrower's gross monthly income
  - Maximum tenure of new housing loans and re-financing facilities for the purchase of HDB flats is reduced to 30 years. New loans with tenure exceeding 25 years and up to 30 years will be subject to tighter LTV limits
10. 2013/12/9
- Reduction of cancellation fees From 20% to 5% for executive condominiums
  - Resale levy for second-timer applicants
  - Revision of mortgage loan terms. Decrease MSR from 60% to 30% of a borrower's gross monthly income

## References

- [1] Bailey, Martin J., Richard F. M., and H. O. Nourse. "A regression method for real estate price index construction." *Journal of the American Statistical Association* 58.304 (1963): 933-942.
- [2] Baltagi, B. and Li, J., "Cointegration of Matched Home Purchases and Rental Price Indexes – Evidence from Singapore". Working Paper, Singapore Management University. (2015)
- [3] Case, B., H. O. Pollakowski, and S. M. Wachter. "On choosing among house price index methodologies." *Real Estate Economics* 19.3 (1991): 286-307.

- [4] Case, B., and J. M. Quigley. "The dynamics of real estate prices." *The Review of Economics and Statistics* (1991): 50-58.
- [5] Case, K. E., and R. Shiller. "Prices of single-family homes since 1970: New indexes for four cities." *New England Economic Review* (1987): 45-56.
- [6] Case, K. E., and R. J. Shiller. "The Efficiency of the Market for Single-Family Homes." *The American Economic Review* (1989): 125-137.
- [7] Chan, K. F., S. Treepongkaruna, R. Brooks, and S. Gray. "Asset market linkages: Evidence from financial, commodity and real estate assets." *Journal of Banking and Finance* 35.6 (2011): 1415-1426.
- [8] Clapp, J. M., C. Giaccotto, and D. Tirtiroglu. "Housing price indices based on all transactions compared to repeat subsamples." *Real Estate Economics* 19.3 (1991): 270-285.
- [9] Diba, Behzad T., and Herschel I. Grossman. "Explosive rational bubbles in stock prices?." *The American Economic Review* (1988): 520-530.
- [10] Englund, P., J. M. Quigley, and C. L. Redfearn. "Improved price indexes for real estate: measuring the course of Swedish housing prices." *Journal of Urban Economics* 44.2 (1998): 171-196.
- [11] Hansen, P. R. "A winner's curse for econometric models: on the joint distribution of in-sample fit and out-of-sample fit and its implications for model selection." Manuscript (2010) Department of Economics, Stanford University.
- [12] Hill, R. Carter, John R. Knight, and C. F. Sirmans. "Estimating capital asset price indexes." *Review of Economics and Statistics* 79.2 (1997): 226-233.
- [13] Iacoviello, Matteo. "House prices, borrowing constraints, and monetary policy in the business cycle." *American Economic Review* (2005): 739-764.
- [14] Koetter, Michael, and Tigran Poghosyan. "Real estate prices and bank stability." *Journal of Banking and Finance* 34.6 (2010): 1129-1138.
- [15] Gatzlaff, D. H., and D. R. Haurin. "Sample selection bias and repeat-sales index estimates." *The Journal of Real Estate Finance and Economics* 14.1-2 (1997): 33-50

- [16] Ghysels, E., A. Plazzi, W. N. Torous, and R. Valkanov. “Forecasting real estate prices.” *Handbook of Economic Forecasting* 2 (2012).
- [17] Guo, X., Zheng, S., Geltner, D., and Liu, H. “A new approach for constructing home price indices: The pseudo repeat sales model and its application in China.” *Journal of Housing Economics* 24 (2014): 20-38.
- [18] McMillen, D.P.. “Repeat sales as a matching estimator.” *Real Estate Economics* (2012): 745–772 .
- [19] Mendicino, Caterina, and Maria Teresa Punzi. “House prices, capital inflows and macroprudential policy.” *Journal of Banking and Finance* 49 (2014): 337–355.
- [20] Musso, Alberto, Stefano Neri, and Livio Stracca. “Housing, consumption and monetary policy: How different are the US and the euro area?” *Journal of Banking and Finance* 35.11 (2011): 3019-3041.
- [21] Nagaraja, C. H., L. D. Brown, and S. M. Wachter. “House price index methodology.” Working Paper, University of Pennsylvania, (2010).
- [22] Nagaraja, C. H., L. D. Brown, and L. H. Zhao. “An autoregressive approach to house price modeling.” *The Annals of Applied Statistics* 5.1 (2011): 124-149.
- [23] Phillips, P. C. B., S. Shi, and J. Yu. “Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500.” *International Economic Review*, forthcoming (2014a).
- [24] Phillips, P. C. B., S. Shi, and J. Yu. “Testing for Multiple Bubbles: Limit Theory of Real Time Detector,” *International Economic Review*, forthcoming (2014b)
- [25] Phillips, P. C. B., Wu, Y., and J. Yu, “Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?” *International Economic Review*, 52 (2011):201–226.
- [26] Quigley, J. M. “A simple hybrid model for estimating real estate price indexes.” *Journal of Housing Economics* 4.1 (1995): 1-12.
- [27] Shiller, R. J. “Derivatives markets for home prices.” No. w13962. *National Bureau of Economic Research* (2008).
- [28] Reinhart, Carmen M., and Kenneth S. Rogoff. “Banking crises: An equal opportunity menace.” *Journal of Banking and Finance* 37.11 (2013): 4557-4573.



- [29] Rosen, S.. “Hedonic prices and implicit markets: Product differentiation in pure competition.” *Journal of Political Economy* 82.1 (1974): 34-55.
- [30] Shi, Song, Jyh-Bang Jou, and David Tripe. “Can interest rates really control house prices? Effectiveness and Implications for Macroprudential Policy.” *Journal of Banking and Finance* 47 (2014): 15–28.
- [31] Sing, T.. “Dynamics of the condominium market in Singapore.” *International Real Estate Review* 4.1 (2001): 135-158.
- [32] S&P/Case-Shiller Home Price Indices-Index Methodology (Nov. 2009)  
<http://www.standardandpoors.com/>.