A new approach for constructing home price indices: The pseudo repeat sales model and its application in China

Xiaoyang Guo [a,b,c], Siqi Zheng [a,*], David Geltner [b], Hongyu Liu [a]

[a] Department of Construction Management and Hang Lung Center for Real Estate, Tsinghua University, China
[b] Center for Real Estate, Massachusetts Institute of Technology, United States
[c] China State Construction Engineering Corporation, China

ARTICLE INFO

ABSTRACT

This paper develops a "pseudo repeat sale" estimation sample construction procedure (ps-RS) to construct more reliable and less biased quality-controlled price indices for newly-constructed homes. The method may be useful wherever new housing development is of sufficiently large scale and homogeneous. Such circumstances characterize many emerging market countries, and here we apply the technique in China. We match two very similar new sales within a defined matching space. Here we test three versions of matching spaces – complex, phase, and building. We then regress the within-pair price differentials onto time dummies and the differentials in unit-specific physical attributes. Locational and community variations, as well as many unobservable or difficult to measure physical attribute variations, are cancelled out in the model, and thereby controlled for. The building-version ps-RS index does the best job in this regard because its within-pair differential is the smallest. We further introduce a "hedonic value" distance metric criterion so that one can deal flexibly with the trade-off between the within-pair "similarity" and the sample size. We explicate and demonstrate formal signal-to-noise oriented metrics of index quality, which can be superior to traditional standard errors based metrics, and we use the new metrics to compare index construction methodologies. The ps-RS approach addresses the problem of lack of repeat-sales data in emerging markets and newly constructed properties and the omitted variables problem in the hedonic method. It also addresses the traditional problems with the classical same-property repeat-sales model in terms of small sample sizes and sample selection bias.

The present paper tests the ps-RS method using a large-scale micro transaction data set of new home sales from January 2006 to June 2011 (444,596 observations) in Chengdu, Sichuan Province, China. The resulting complex-based ps-RS index essentially parallels the hedonic index, suggesting that the hedonic index is not superior to that version of the ps-RS index in terms of systematic results. The phase-based ps-RS index has a lower growth trend and the building-based version lower still, indicating omitted variables relating to the physical quality of the units are not well controlled for in the hedonic, and suggesting that the building-based version of the ps-RS index provides the greatest control for such quality differences. Building-based ps-RS

* Corresponding author.
 E-mail address: zhengsiqi@tsinghua.edu.cn (S. Zheng).

indices with different distance metric thresholds are almost the same. Compared to the hedonic, the ps-RS provides a smoother index indicating less random estimation error (or "noise").

1. Introduction

In the world of transaction price indices used to track the dynamics in housing markets, the problem of controlling for heterogeneity in the homes transacting in different periods of time is perhaps the most crucial challenge. The simple mean or median values of all the sale prices per square meter each period will not produce good price indices because the location, size, quality, and components of the homes being sold keep changing over time. The two major methods in the academic literature for addressing this challenge are the hedonic and repeat sales regressions. Of these two, in the U.S., only the repeat-sales approach has seen widespread regular production and publication in official or industry statistics (for example, the FHFA and S&P/Case-Shiller home price indices).

In Chinese cities, as a representative case in this paper, we face two unique features in that country's urban residential market, features which also characterize development in many emerging market countries. First, new home sales account for an exceptionally large share of total sales in China (87% in 2010) due to a growth rate in the economy and urbanization that in the case of China has been truly unprecedented in world history. Thus, the classical repeat sales (RS) approach is of very limited usefulness because the typical housing unit in China has only appeared once on the market. Yet the hedonic method may face more than its usual challenges because the omitted variables problem may be more severe in Chinese cities due to very rapid evolution of urban spatial structure, infrastructure construction, and (most difficult to observe) the quality and features and amenities within the housing units themselves (such as apartment design, appliances, finishes, and HVAC) as household income rises at an extremely rapid rate. Secondly, housing development in many high-density cities, such as those in Mainland China, Taiwan, Singapore and many other Asian cities, occurs at a uniquely large scale and with a high degree of homogeneity in the units built within the typical residential "complex". In each complex, a number of buildings are constructed containing altogether hundreds or even thousands of units all having essentially the same location, architecture design, structure, appliances and finishes.

The proposal in this paper is to develop a new type of "repeat sales" model, which we dub "pseudo repeat sales" (ps-RS). Fundamentally similar to the matched-sample procedure recently proposed by McMillen (2010, 2012) in that the price observation pairs used in the regression are not actually repeat-sales of the same property, our proposal is a new matching criterion that we think is particularly appropriate for Chinese cities and other high-density cities where large-scale residential complexes dominate the urban housing development. We deal with the omitted variables issue by employing a within-building matching

criterion instead of the more stringent, same-unit criterion of the classical RS approach.[1] This approach not only addresses the problem of lack of repeat-sales data and problematical hedonic variables observation, but also addresses the traditional problems with the classical repeat-sales model of small sample sizes and sample selection bias in properties with repeated sales, as the ps-RS procedure, like a hedonic price index, uses all of the transactions data. More specifically, the proposed model is (in fact must be) a hybrid repeat sales/hedonic model of the type that has been demonstrated to have desirable features in the econometric literature, because the paired units in the ps-RS are not identical. The hybrid (hedonic) component of our model is small and relies only on variables for which good data can be easily obtained, because it only has to control for differences between units within the same building. We believe the ps-RS still retains essentially the characteristics of a "repeat sales" model. In this paper we present an argument and evidence that the ps-RS can produce a more reliable and accurate picture of home price appreciation in these very important markets.

The rest of this paper is organized as follows: Section 2 will present some relevant background and literature review. Section three describes the features of the new-home market in Chinese cities and how those features affect the choice of housing price index construction methodology. We describe in detail our approach for developing the ps-RS index in Section 4. After data description in Section 5, the index calculation results for our demonstration city of Chengdu are presented in Section 6, including a quantitative comparison of the ps-RS with the standard hedonic method (which is the only realistic alternative since classical same-property repeat sales is not possible for new housing). Section Seven concludes.

2. Background and literature review

The hedonic approach goes back to Kain and Quigley (1970), who decomposed the components of housing price dynamics using the hedonic model, from which a quality-controlled housing price index is generated by controlling for home transactions' physical and location attributes. Other pioneers of hedonic price modeling were Court (1939), Griliches and Adelman (1961), and Rosen (1974). Two alternative methods have been proposed to construct a hedonic housing price index. The first method assumes constant relative preferences for housing attributes over

---

[1] The matching criterion can also be applied to sales across different buildings but within the same phase (several buildings constructed at the same time), or within the same complex. However, as we will discuss below, larger matching spaces (across buildings) appear to be less effective in mitigating the problem of omitted variables and controlling for quality differences. Our empirical results indicate that the within-building criterion is the best choice in our study market, the metropolitan area of Chengdu, Sichuan.