Regional Science and Urban Economics 59 (2016) 37–53

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

House prices at different stages of the buying/selling process☆

Chihiro Shimizu [a,*], Kiyohiko G. Nishimura [b], Tsutomu Watanabe [b]

[a] Institute of Real Estate Studies, National University of Singapore, 21 Heng Mui Keng Terrace, #04-02, 119613 Singapore
[b] Faculty of Economics, The University of Tokyo

ARTICLE INFO

Article history:
Received 1 December 2015
Received in revised form 21 March 2016
Accepted 7 April 2016
Available online 25 April 2016

JEL classification:
R21
R31
C10

Keywords:
House price index
Quantile regressions
Hedonic regressions
Quality adjustment
Goodness-of-fit tests

ABSTRACT

In constructing a housing price index, one has to make at least two important choices. The first is the choice among alternative estimation methods. The second is the choice among different data sources of house prices. The choice of the dataset has been regarded as critically important from a practical viewpoint, but has not been discussed much in the literature. This study seeks to fill this gap by comparing the distributions of prices collected at different stages of the house buying/selling process, including (1) asking prices at which properties are initially listed in a magazine, (2) asking prices when an offer for a property is eventually made and the listing is removed from the magazine, (3) contract prices reported by realtors after mortgage approval, and (4) registry prices. These four prices are collected by different parties and recorded in different datasets. We find that there exist substantial differences between the distributions of the four prices, as well as between the distributions of house attributes. However, once quality differences are controlled for, only small differences remain between the different house price distributions. This suggests that prices collected at different stages of the house buying/selling process are still comparable, and therefore useful in constructing a house price index, as long as they are quality adjusted in an appropriate manner.

1. Introduction

Many of the property price bubbles experienced during the 20th century were triggered by steep increases and sharp decreases in property prices. More recently, the Global Financial Crisis has several causes but a main factor was a housing bubble in the U.S. which led banks to make mortgage loans that were based on the assumption that house prices were not rising unusually quick. If it had been that accurate regional house price indexes for the U.S. were widely available to the public, it is unlikely that so many bad housing loans would have been made; i.e., the housing bubble would have been immediately visible and both lenders and borrowers could have foreseen that U.S. house price inflation rates over the period 2000–2007 were not sustainable.

Part of the problem is that national statistical agencies have not constructed official house price indexes.

Why have national statistical agencies not constructed appropriate property price indexes?

There are two reasons:

(1) It is very difficult to construct accurate property price indexes and as a result statistical agencies have been reluctant to allocate their scarce resources to the construction of indexes where there has not been international agreement on how exactly to construct such an index;
(2) Property prices by themselves do not occupy an important position in the major statistics that countries construct; i.e., house price indexes do not appear directly in either the Consumer Price Index or in the main components of GDP.

The European statistical agency, Eurostat, is about to publish a Residential Property Price Index (RPPI) Handbook in collaboration with UN, OECD, IMF, BIS and World Bank. This book describes some of the problems associated with constructing price indexes for residential prices and gives advice on methods that could be used in order to construct RPPI. The RPPI Handbook is an attempt by the international statistical community to encourage countries to construct house price indexes according to the suggested best practices that are laid out in the Handbook. For official statistics such as CPIs or National Accounts, it is common to describe the creation method in detail, including the data sources, to enable comparisons between countries.

☆ This paper is based on the outcome of a project at the Ministry of Land, Infrastructure, Transport and Tourism ("Study Group on the Development of Real Estate Price Indexes"), as well as on the outcome of a project at the Real Estate Information Network for East Japan ("Working Team on Research-Oriented Use of the REINS Database"). We would like to thank Erwin Diewert, David Fenwick, Sadao Sakamoto, and Hiwon Yoon for their helpful discussions and comments. This research forms part of the project on "Comparable Studies on Property Price Indexes in Japan and US" funded by Nomura Research Foundation.

* Corresponding author.
E-mail address: cshimizu@nus.edu.sg (C. Shimizu).

http://dx.doi.org/10.1016/j.regsciurbeco.2016.04.001
0166-0462/

In constructing a housing price index, one has to make several nontrivial choices. One of them is the choice among alternative estimation methods, such as repeat-sales regression, hedonic regression, and so on. There are numerous papers on this issue, both theoretical and empirical (Case et al., 1991; Dorsey et al., 2010; Shimizu et al., 2010).[1] However, there is another important issue which has not been discussed much in the literature, but has been regarded as critically important from a practical viewpoint: the choice among different data sources for housing prices. There are several types of datasets for housing prices: datasets collected by real estate agencies and associations; datasets provided by mortgage lenders; datasets provided by government departments or institutions; and datasets gathered and provided by newspapers, magazines, and websites.[2] Needless to say, different datasets contain different types of prices, including sellers' asking prices, transaction prices, valuation prices, and so on.

With multiple datasets available, one may ask several questions. Are these prices different? If so, how do they differ from each other? Given the specific purpose of the housing price index one seeks to construct, which dataset is the most suitable? Alternatively, with only one dataset available in a particular country, one may ask whether this is suitable for the purpose of the index one seeks to construct. This paper is a first attempt to address some of these questions.[3]

Specifically, in order to do so, we will conduct a statistical comparison of different house prices collected at different stages of the house buying/selling process. To conduct this exercise, we focus on four different types of prices: (1) asking prices at which properties are initially listed in a magazine, (2) asking prices when an offer for a property is eventually made and the listing is removed from the magazine, (3) contract prices reported by realtors after mortgage approval, and (4) registry prices. We prepare datasets of these four prices for condominiums traded in the Greater Tokyo Area from September 2005 to December 2009. The four prices are collected by different institutions and therefore recorded in different datasets: (1) and (2) are collected by a real estate advertisement magazine publisher; (3) is collected by an association of real estate agents; and (4) is collected jointly by the Land Registry and the Ministry of Land, Infrastructure, Transport and Tourism.

An important advantage of prices at earlier stages of the house buying/selling process, such as initial asking prices in a magazine, is that they are likely to be available earlier, so that house price indexes based on these prices become available in a timely manner. The issue of timeliness is important given that it takes more than 30 weeks before registry prices become available. On the other hand, it is often said that prices at different stages of the buying/selling process behave quite differently. For example, it is said that when the housing market is, say, in a downturn, prices at earlier stages of the buying/selling process, such as initial asking prices, will tend to be higher than prices at later stages. Also, it is said that, for various reasons, prices at earlier stages contain non-negligible amounts of "noise" (Allen and Dare, 2004; Haurin et al., 2010; Knight et al., 1998). For instance, prices can be renegotiated extensively before a deal is finalized, and not all of the prices appearing at earlier stages end in transactions, for example, because a potential buyer's mortgage application is not approved (Genesove and Mayer, 1997, 2001; Engelhardt, 2003).

The main question of this paper is whether the four prices differ from each other, and if so, by how much. We will focus on the entire cross-sectional distribution for each of the four prices to make a judgment on whether the four prices are different or not.[4] The cross-sectional distributions for the four prices may differ from each other simply because the datasets in which they are recorded contain houses with different characteristics. For example, the dataset from the magazine may contain more houses with a small floor space than the registry dataset, which may give rise to different price distributions. Therefore, the key to our exercise is how to eliminate quality differences before comparing price distributions.

We will conduct quality adjustments using two different methods. The first is to only use the intersection of two different datasets, that is, observations that appear in two datasets. For example, when testing whether initial asking prices in the magazine have a similar distribution as registry prices, we first identify houses that appear in both the magazine dataset and the registry dataset and then compare the price distributions for those houses in both datasets. In this way, we ensure that the two price distributions will not be affected by differences in house attributes between the two datasets. This idea is quite similar to the one adopted in the repeat sales method, which is extensively used in constructing quality-adjusted house price indexes. As is often pointed out, however, repeat sales samples may not necessarily be representative because houses that are traded multiple times may have certain characteristics that make them different from other houses.[5] A similar type of sample selection bias may arise even in our intersection approach. Houses in the intersection of the magazine dataset and the registry dataset are cases which successfully ended in a transaction. Put differently, houses whose initial asking prices were listed in the magazine but which failed to get an offer from buyers, or where potential buyers failed to get approval for a mortgage, are not included in the intersection.

The second method is based on hedonic regressions. This is again widely used in constructing quality-adjusted house price indexes. The hedonic regression we will employ in this paper differs from those extensively used in previous studies, which are based on the assumption that the hedonic coefficient on, say, the size of a house is identical for high-priced and low-priced houses. This restriction on hedonic coefficients may not be problematic as long as one is interested in the mean or the median of a price distribution, but it is a serious problem when one is interested in the shape of the entire price distribution. In this paper, we will use quantile hedonic regression in which hedonic coefficients are allowed to differ for high-priced and low-priced houses.

The main findings of this paper are as follows. We find that the four prices have substantially different distributions. However, these differences mainly come from differences in the attributes of houses contained in the different datasets. By looking at the intersections of the datasets and by employing quantile regressions, we show that once quality differences are eliminated, there remain only small

[1] Recently, Deng et al. (2012), McMillen (2012) and McMillen and Thorsnes (2006) focused on the distribution of housing prices and proposed new index estimation method, matching model.

[2] Eurostat (2011) provides a summary of the sources of price information in various countries. For example, in Bulgaria, Canada, the Czech Republic, Estonia, Ireland, Spain, France, Latvia, Luxembourg, Poland and the USA price data collected by statistical institutes or ministries is used. In Denmark, Lithuania, the Netherlands, Norway, Finland, Hong Kong, Slovenia, Sweden and the UK information gathered for registration or taxation purposes is used. In Belgium, Germany, Greece, France, Italy, Portugal and Slovakia data from real estate agents and associations, research institutes or property consultancies are used. Finally, in Malta, Hungary, Austria and Romania data from newspapers or websites is used.

[3] There are several papers that focused on data sources for house price indexes (Gatzlaff and Haurin, 1998; Genesove and Mayer, 2001; Goetzmann and Peng, 2006). However, they don't compare multiple data sources.

[4] An alternative approach would be to compare the four prices in terms of their average prices or in terms of their median prices. However, these statistics capture only one aspect of cross-sectional price distributions.

[5] Shimizu et al. (2010) construct five different house price indexes, including hedonic and repeat sales indexes, using Japanese data for 1986 to 2008. They find that there exists a substantial discrepancy in terms of turning points between hedonic and repeat sales indexes. Specifically, the repeat sales measure signal turning points later than the hedonic measure: for example, the hedonic measure of condominium prices bottomed out at the beginning of 2002, while the corresponding repeat sales measure exhibits a reversal only in the spring of 2004.