Regional Science and Urban Economics 60 (2016) 300–315

Contents lists available at ScienceDirect

Regional Science and Urban Economics

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

Hedonic regression models for Tokyo condominium sales

W. Erwin Diewert [a], Chihiro Shimizu [b]

[a] School of Economics, University of British Columbia and the School of Economics, University of New South Wales, Canada
[b] Department of Sports Science, Nihon University and Institute of Real Estate Studies, National University of Singapore, Japan

ARTICLE INFO

Article history:
Received 7 January 2016
Received in revised form 6 July 2016
Accepted 9 August 2016
Available online 3 September 2016

Keywords:
Condominium property price indexes
System of National Accounts
Balance Sheets
Methods of depreciation
Land and structure price indexes
Hedonic regressions

JEL classification:
C2
C23
C43
E31
R21

ABSTRACT

The paper fits a hedonic regression model to the sales of condominium units in Tokyo over the period 2000–2015. The problem is complicated by the need to decompose the selling price of a unit into a component that can be attributed to the structure area of the unit and another component that can be attributed to the unit's share of land value. There is very little information on the value of condominium land and so this paper develops a methodology for reducing this knowledge gap. The paper extends the builder's model which was developed in Eurostat (2013). Characteristics which prove to be important in explaining condominium prices are: the floor space area of the unit, the total land area of the building, the number of units in the building, the total number of stories in the building, the height of the sold unit, the age of the structure and the amount of excess land. The paper also derives an estimate for the annual geometric structure depreciation rate for condominiums in Tokyo.

1. Introduction

The international System of National Accounts asks countries to provide estimates for the value of assets held by the various sectors in the economy. These estimates are supposed to appear in the Balance Sheet Accounts of the country. An important asset for the Household Sector is the stock of housing. For many modeling purposes, it is important to not only have estimates for the value of the housing stock but to decompose the overall value into (additive) land and structure components and then to further decompose these value aggregates into constant quality price and quantity components.[1] This is not an easy task. When a housing property is sold, the selling price values the sum of the structure and land components and so a structure-land decomposition must be obtained by a modeling exercise. The problem of obtaining constant quality price components for the land and structure components of a housing unit is further complicated by the fact that housing units are almost always unique assets. A dwelling unit is different from any other dwelling unit at the same period in time due to its location, which is unique (and as locations vary for the same physical structure, the price of the land plot for the unit will generally change due to locational amenities). The same dwelling unit compared over space will also be different due to depreciation and possible renovations to the structure.

Our task in the present paper is to present a modeling strategy to provide a decomposition of condominium sales into constant quality price and quantity components for the structure and land components of the condo sale.[2] We will follow roughly the same strategy as was outlined in Chapter 8 of Eurostat (2013) where a similar modeling strategy was applied to sales of detached dwellings. Our present task is much more difficult for two reasons:

- The value of a condominium unit is made up of a structure and a land component. But it is difficult to know exactly how to allocate the share of the total land value of the building plot to any particular unit. This problem does not arise for detached houses.

E-mail addresses: erwin.diewert@ubc.ca (W.E. Diewert), shimizu.chihiro@nihon-u.ac.jp (C. Shimizu).

[1] Governments in many countries impose separate tax rates on the land and structure components of residential properties. Thus if these taxes are to be based on market values, it is important to be able to determine the values of these land and structure components in a scientific way.

[2] There is a vast literature related to residential property price indexes. In Asian countries, Shimizu, Nishimura and Watanabe (2010) compared Repeat Sales Indexes to Hedonic Indexes for Tokyo, while Deng et al. (2012) proposed a matching method for the Singapore housing market. Wu et al. (2014) and Guo et al. (2014) constructed price indexes for newly built houses in China. However, there are very few papers that construct quality adjusted price indexes for condominium sales which is our focus, along with providing a method to decompose property sales into land and structure components.

http://dx.doi.org/10.1016/j.regsciurbeco.2016.08.002
0166-0462/

- There is much more heterogeneity in condo units than there is in detached dwelling units. With detached dwelling units, the suburb of the unit, its floor space area, the area of the land plot and the age of the unit can explain a great deal of the variation in detached houses. However, these variables are not sufficient to explain the variation in condo prices. As we shall see, other important explanatory variables are the height of the building, the height of the condo unit that is being sold, and the area of the land plot that is *not* being used to support the building.

Section 2 explains our quarterly data set which covers sales of condo units in Tokyo over the years 2000–2015.

Section 3 explains our basic regression model. We find that this preliminary regression model does not provide a reasonable decomposition of condo value into additive land and structure components, which is required for national income accounting purposes. Thus we construct an estimated imputed structure value for the condo unit and subtract this imputed value from the selling price of the condo unit to obtain an imputed land value that can be associated with the condo unit. In Sections 4–10, we use these imputed land values as the dependent variable in our regression models in an attempt to find characteristics which can explain the variation in these imputed land prices. In Section 11, we return to the actual selling prices for the condo units as the dependent variable in our regression model, using the land characteristics that we discovered were useful explanatory variables for the regressions in Sections 4–10. In Section 11, we now estimate the annual structure geometric depreciation rate instead of assuming it. Section 12 introduces a few additional characteristics into the regression model; these characteristics are thought to affect the structure value rather than the land value component of the total value of the condo unit.[3]

In Section 13, we group the 9 wards of Tokyo for which we have data into rich, medium and poorer wards and estimate ward time dummy variables for each type of ward. However, as will be shown in Section 14, the resulting ward specific land prices turned out to be too variable to be credible. The basic problem is that we do not have a large enough number of observations to support the model presented in Section 14. However, it is useful to show how our model can provide more detailed land prices by local area, if adequate data were available.

Section 14 shows how the separate land prices generated by the models in Sections 12 and 13 can be combined with our structure prices to generate overall condo price indexes. The results presented in this section lead us to prefer the model presented in Section 12 over the model presented in Section 13.

Section 15 compares our preferred *overall* condo price index (generated by the model in Section 12) to four other indexes. The first alternative index is an approximate price index for the *stock* of condo units in our 9 wards of Tokyo as opposed to our Section 12 overall condo price index which is an index for the *sales* of condo units in the 61 quarters in our sample. However, we show that the two indexes are virtually identical. The next two alternative indexes are simple indexes based on the mean and median values of sales of condo units in the 61 quarters. These indexes perform poorly due to their variability and downward biases (due to their neglect of depreciation). Our final comparison index is based on a simple traditional time dummy hedonic regression. The resulting time dummy based index performs quite well in that it is close to our preferred indexes.[4]

Section 16 concludes.

[3] In the end, our hedonic regression model is similar in spirit to the models proposed by Gloudemans (2000, 2002). Longhofer and Redfearn (2009, p. 3) succinctly characterized the Gloudemans approach as follows: "Specifically, these papers model total property value as additive in its land and building components but multiplicative within the characteristics of each of these components. Because land and building values are separable in this model, it is possible to use the regression coefficients to separately estimate land and building values."

[4] However, the time dummy approach does not generate separate land and structure price components, which is the main purpose of our paper.

2. The Tokyo condominium data

Our basic data set is on sales of condominium units located in 9 Wards in the central area of Tokyo over the 61 quarters starting at the first quarter of 2000 and ending at the first quarter of 2015. In addition to the sales prices, various characteristics of the properties were obtained from the website, Suumo (Residential Information Website), provided by Recruit Co., Ltd., one of the largest vendors of residential listings information in Japan. This source provides time series of listed prices from the week when it is first posted until the week it is removed due to its sale.[5] We used the price in the final week because this can be safely regarded as sufficiently close to the contract price.[6]

There were a total of 3232 observations (after range deletions) in our sample of sales of condo units in Tokyo.[7] The definitions for the selling price and 11 characteristics of the units sold and their units of measurement are as follows:

V = the value of the sale of the condo unit in 10,000 Yen[8]

S = structure area (floor space area) of the condo in units of meters squared

TS = floor space area for the entire building

TL = lot area for the entire structure in units of meters squared

A = age of the structure in years

H = the story of the unit; i.e., the height of the unit that was sold

TH = the total number of stories in the building; i.e., the total height of the building

NB = number of bedrooms in the unit

TW = walking time in minutes to the nearest subway station

TT = subway running time in minutes to the Tokyo station from the nearest station during the day (not early morning or night)

SCR = reinforced concrete construction dummy variable (= 1 if reinforced; 0 otherwise)

$SOUTH$ = dummy variable (= 1 if the unit faces south; 0 otherwise)

After range trimming, the minimum and maximum values for the various variables are listed in Table 1. It can be seen that even after trimming, there is a considerable amount of variation left in the data.[9]

In addition to the above variables, we also have information on which Ward of Tokyo the sales took place. We used this information to create ward dummy variables, $D_{W,tn,j}$, which will be described more

[5] There are two reasons for the listing of a unit being removed from the magazine: a successful deal or a withdrawal (i.e. the seller gives up looking for a buyer and thus withdraws the listing). We were allowed access to information regarding which the two reasons applied for individual cases and we discarded those transactions where the seller withdrew the listing.

[6] Recruit Co., Ltd. provided us with information on contract prices for about 24 percent of all listings. Using this information, we were able to confirm that prices in the final week were almost always identical with the contract prices (Shimizu et al., 2016). White (2004) compared the distributions of prices collected at different stages of the house buying/selling process. The four stages are (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. The above research found 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 remained between the different house price distributions.

[7] It is risky to estimate hedonic regression models over wide ranges when observations are sparse at the beginning and end of the range of each variable. Moreover, real estate data usually contains many outliers and trimming the range of the independent variables will typically help eliminate some outliers.

[8] The variable V is V_{tn} where $t = 1, \ldots, 64$ indicates the quarter when the unit was sold and $n = 1\ldots.,N(t)$ indicates the nth condo sale in quarter t and $N(t)$ = the total number of condo sales in our sample during quarter t.

[9] Table 1 also reflects the results of range trimming for three synthetic variables: (i) $L_S \equiv (S/TS)TL$ (this is an imputation for the *share of the property's total land area TL* that can be attributed to the sold unit where the unit has floor space area S and the building has total floor space area TS); (ii) the *footprint ratio* F_R of the structure which is equal to the ratio of the land area occupied by the structure (TS/TH) to the total property land area TL so $F_R \equiv (TS/TH)/TL$ and (iii) an approximation to the *useable floor space ratio* of the building, $UFSR \equiv (N \times S)/TS$ where N is the number of units in the building, S is the floor space of the sold unit and TS is the total floor space of the building, including common space. We deleted observations that fell outside the following range limits: $7 \le L_S \le 60$; $0.1 \le F_R \le 0.8$ and $0.5 \le UFSR \le 1.5$.