Construction of Residential Property Price Indices Using the Hedonic Approach: An Application to the Residential Real Estate Market in Sri Lanka[1]

S R C L Gunawardhana[2]

Abstract

Monitoring real estate property prices is crucial for policy makers all over the world, particularly for central banks due to their interconnections with the monetary and financial system stability of an economy. As a result, compiling property price indices has increasingly gained attention from policymakers. However, compiling property price indices is believed to be difficult due to the highly heterogeneous nature of properties, requiring reliable data sources and a methodological approach that is different to those used in compiling other price indices. Against this backdrop, this paper attempts to compile price indices for residential properties in Sri Lanka with a view to supporting policymakers to monitor the price movements in the real estate sector. Considering the relatively less heterogenous nature of condominiums and stages of the buying and selling process, price indices are first developed for advertised condominiums and new condominiums using the Hedonic Regression based Rolling Window Time Dummy method, which is identified as the most suited property price index compilation method in the Sri Lankan context. Further, internationally accepted model specification improvement techniques and index smoothing techniques are also used in the study. Upon successful compilation of price indices for condominiums, the compilation process has been extended to cover the house and land markets of Sri Lanka.

Key Words*: Residential Property Price Index, Hedonic Regression, Asking Prices, Constant Quality Indices*

JEL Classification: *C2, C10, C43, C81, C82, C83, D12, E31, R21, R30, R31*

---

[1] The author is thankful to Mr. N O'Hanlon, Statistics Department of the International Monetary Fund (IMF) for providing technical assistance for this index compilation process and the Data for Decisions (D4D) fund of IMF for funding the technical assistance missions. The author is also thankful to Mr. A R K Wijesekara, Former Assistant Governor, Central Bank of Sri Lanka, Mr. G D P D Jayathilake, Former Director, Statistics Department of the Central Bank of Sri Lanka, Dr. (Mrs.) H K J Ekanayake, Director, Statistics Department of the Central Bank of Sri Lanka, Mrs. D G D I Ekanayake, Additional Director, Statistics Department of the Central Bank of Sri Lanka, Ms. R M M K Rathnayake, Deputy Director, Statistics Department of the Central Bank of Sri Lanka, Ms. P I Rupasinghe, Head of Economic Indicators and Statistical Investigations Division, Statistics Department of the Central Bank of Sri Lanka for their guidance and support extended throughout this study. In addition, the author would like to express her sincere thanks to all the staff members of the Economic Indicators and Statistical Investigations Division of Statistics Department of the Central Bank of Sri Lanka for contributing to the index compilation process in various stages. The views expressed in this paper are the author's own and do not necessarily reflect those of the Central Bank of Sri Lanka.

[2] Senior Assistant Director, Statistics Department, Central Bank of Sri Lanka (chamili@cbsl.lk)

1. Introduction

Developments in the real estate sector have an impact on the price and financial system stability due to its highly capital intensive nature. Thus, key interest rate changes can have an intermediate effect on the economy via real estate prices and might play a decisive role in the monetary policy transmission. In addition, affordability and price changes of properties have a direct impact on households, in terms of spending and indebtedness. Against this backdrop, real estate property price indices have gained importance, particularly for central banks, in making policy decisions. This is apparent in the case of Sri Lanka with the recent upswing in vertical developments in urban areas, especially condominiums. The unavailability of reliable indicators of real estate price movements in Sri Lanka was identified as a constraint to monitor developments in the real estate sector. To fill this gap, a condominium price index was initiated on experimental basis in 2016, and over time the methodology was improved while extending the compilation of price indices to house and land markets.

In some countries, administrative data collected by government agencies on property transactions provide suitable data sources, including information on individual property characteristics and prices, to underpin compilation of a price index. In others, however, the biggest challenge in compiling property price indices is the unavailability of required data. This is the case in Sri Lanka, where it is not possible to obtain timely granular data due to the absence of automated IT systems. Thus, surveys with property developers and real estate property advertisements published in property websites were identified as suitable alternative data sources to obtain sales information of properties. Accordingly, to extract advertised (asking) prices and other property characteristics published in property advertisements, web scraping techniques were implemented from January 2019 onwards on a monthly basis.

The expansion of property price indices compilation over time with various index compilation methods and different data sources are detailed in this paper. The data processing techniques and real estate property price index compilation methodologies applied in developing the indices were carried out based on internationally accepted best practices[3]. Accordingly, four residential property price index compilation methodologies, namely, the Hedonic Characteristic Laspeyres Price Index method, Hedonic Double Imputation Fisher Price Index method, Hedonic Imputation Fisher Price Index method, and the Hedonic Regression based Rolling Window Time Dummy method were used to experiment in developing the indices. It was observed that the Hedonic Regression based Rolling Window Time Dummy method is most suited for the thin real estate market in Sri Lanka. Further, it was identified that property price indices can be developed using prices available at different stages of its buying and selling process. Even though there are variations in price levels, the long term price trends are aligned to similar trends. In terms of price indices compiled for condominiums in Sri Lanka, the asking price index compiled using the advertised prices and the price index for new condominiums compiled using the prices at the point of the sale follows the same trend over time indicating a highly positive correlation. Thus, the outputs of this study fulfil the need of property price indices for Sri

[3] Indices were developed using the learnings gathered from three training sessions on developing Residential Property Price Indices (RPPIs) held at the Singapore Training Institute (STI) of the International Monetary Fund (IMF) and two technical assistance missions held at the Statistics Department of the Central Bank of Sri Lanka by an IMF expert. In addition, the methodologies introduced in the Handbook on RPPIs published by Eurostat of the European Commission and the relevant research papers published internationally were also referred.

Lanka to monitor the real estate sector prices and developments.

The rest of the paper is structured as follows: Section 2 contains the literature review on related past studies, Section 3 outlines the data collection, index compilation methodologies and the process used in the study, Section 4 consists of results and discussion while Section 5 concludes the study.

2. Literature Review

Securing access to a reliable data source and selecting a suitable index compilation methodology are crucial in compiling real estate property price indices. It is generally agreed that accurately reported transaction prices, collected at the completion of the transaction process, including granular information on individual property characteristics provide the most appropriate observations. However, property price information can be captured at different stages of the property buying and selling process, such as, at the point of advertising for sale, at the point of buyer-seller agreeing for a price, at the time of deed transformation and at the time of obtaining a loan if the buyer obtain a mortgage. In selecting a data source to compile the index, the availability of price determining characteristics information and the possibility of timely data collection are the most important factors to consider. In countries where official information systems are developed, administrative data sources can be used to obtain data for index compilation. But when such systems are underdeveloped, the compilers have to consider alternative data sources such as property agents and advertisements. In previous studies, it has been observed that the researchers have used different price types from different sources in compiling property price indices.

Shimizu *et al.* (2011) consider four types of property prices at different stages of its buying/selling process from different data sets. Accordingly, they use asking prices at which properties are initially listed for sale, prices when an offer is eventually made, contract prices reported by realtors, and finally the registry prices in the Greater Tokyo Area. Their study suggests that the prices collected at different stages of the buying/selling process exhibit substantial differences between price distributions but are still comparable since they follow similar trends. Therefore, different price types can be used in constructing house price indices, as long as they are quality adjusted in an appropriate manner. Further, the property characteristics that are listed in advertisements enable compilers to collect all required granular level data timely when advertisement information is used. Shimizu *et al.* (2010) use individual listings in a widely circulated real estate advertisement magazine in Tokyo, and Lyons (2019) examines the relationship between listed and transaction prices during Ireland's recent turbulent housing market cycle. The Irish study found that online listings represent a rich potential data source across economies and listed prices are an adequate substitute for transaction prices, even in relatively unstable market conditions.

In terms of index compilation methodologies, there are four standard methodologies that are discussed in the Handbook on Residential Property Price Indices (RPPIs) for real estate price index compilation. They are, (i) Stratification or Mixed Adjustment method, (ii) Hedonic Regression method, (iii) Repeat Sales method, and (iv) Appraisal-Based method. Among these, hedonic regressions are considered the most promising approach to control for changes in the quality of characteristics of properties transacted from period to period. Song and Wilhelsson (2010) also emphasise the necessity of using quality adjusted techniques in constructing house price indices due to their heterogeneous nature in conducting their study on compiling a price index for condominiums in Sweden.

Diewert *et al.* (2007) conducted a comparison between two main and quite distinct approaches to the measurement of hedonic price indices: time dummy hedonic indices and hedonic imputation indices. The study discusses that there may be reasons to prefer one against the other and accordingly, the hedonic imputation method is inherently more flexible since it constrains the parameters on the characteristic variables to be the same over the two periods under consideration. At the same time, it is likely to have less confidence in an index based on constraining the coefficients to be the same. In this sense the concern over the use of time dummy hedonic indices is warranted. Meanwhile, Hill *et al.* (2017) construct weekly hedonic house price indices using Rolling Time Dummy (RTD) method for Sydney and Tokyo. They note that RTD method tends to perform well with smaller data sets and computing high frequency indicators. Patrick (2017) also uses the rolling time dummy hedonic method for compilation of the official index for the republic of Ireland. De Haan (2004) discusses how the time dummy method fits into the matched model methodology in compiling these kinds of price indices. In addition, Diewert (2011) constructs house price indices for a small Dutch town following four alternative methods, namely, stratification method, time dummy hedonic regression method, hedonic regression imputation method, and additive hedonic regression method. He notes that the problem of change in historical results when new data become available, which is identified with many hedonic regression models, is addressed by the rolling window hedonic regression methodology. Shimizu *et al.* (2010) also discuss alternative hedonic housing price index compilation approach for condominiums in Tokyo Metropolitan Area and under its structurally restricted approach, they use the hedonic rolling window time dummy method for index compilation.

Thus, as per previous literature, the hedonic regression based rolling window time dummy method is recognised as an acceptable method in compiling residential property price indices. It not only corrects price changes for changes in the quality of items, but also allows the indices to incorporate matched and unmatched models. The hedonic regression based rolling window time dummy method has become popular with national statistical institutes in European countries such as Ireland, Luxembourg, Cyprus, and Malta in constructing the official House Price Index. Interestingly, each of these countries would be considered to have relatively thin housing markets.

Shimizu *et al.* (2010) discuss the different behaviours of house price indices depending on the estimation method. In their comparison between Hedonic and Repeat Sales Measures, it is found that there exists a substantial discrepancy in terms of turning points between hedonic and repeat sales indices, even though the hedonic index is adjusted for structural changes and the repeat sales index is also adjusted appropriately. The repeat sales measure is found to signal turning points later than the hedonic measure and this discrepancy cannot be fully removed even if the repeat sales index is adjusted for depreciation.

It is also important to look at the price determining characteristics considered in previous studies, in order to select suitable variables in compiling the property price indices. The explanatory variables used in the regression equations include price determining characteristics of properties in previous studies depending on the data availability. Accordingly, Diewert (2011) considers the age of the house, floor space area and land area in developing house price indices for a small Dutch town. In addition to these variables, Diewert and Shimizu (2014) consider the location of the property and number of bedrooms when compiling property price indices for Tokyo. Further, Shimizu *et al.* (2011) use the distance to the nearest station and travel time to the terminal station when observing prices of condominiums traded

in the Greater Tokyo Area. When compiling property price indices for Sydney, Hill *et al.* (2017) consider explanatory variables such as the property type (i.e., detached or semi), number of bedrooms, number of bathrooms, land area, postcode, and exact address (longitude and latitude). To simplify the computations, they merge the number of bathrooms and number of bedrooms to broader groups (one, two, and three or more bathrooms; one or two, three, four, five or more bedrooms). When compiling a price index for condominiums in Sweden, Song and Wilhelsson (2010) consider factors such as size, number of rooms, location, floor level, whether the property has a balcony or not and whether the property have an elevator or not. Further, Patrick (2017) uses the floor area, number of bedrooms, and dwelling type to control for constant quality when compiling the property price index for Ireland. Accordingly, in most cases property characteristics and location attributes are considered as price determining factors of properties.

3. Methodology

The price index compilation process was initiated for condominiums in the Colombo district in 2016 using manually collected advertisement data from property websites and newspapers. The index was compiled according to the Hedonic Characteristic Laspeyres Price Index method from August 2016 to January 2017. Thereafter, a Hedonic Double Imputation Fisher Price Index methodology was applied in 2017. Meanwhile, a survey was introduced with condominium property developers in September 2017 on a quarterly basis mainly with the objective of collecting actual transaction information for new condominiums. Using this information, a separate index was also compiled for new condominiums following the Hedonic Double Imputation Fisher Price Index method until 2019. However, both indices compiled were highly volatile and seemed not to reflect market sentiments.

Against this background, experimental work was carried out following three alternative index compilation methods[4], and the Rolling Window Time Dummy (RWTD) method was selected for compiling indices for condominiums since this method is more suitable for markets with less data points as per the global practices (Hill *et al.* (2017)). Meanwhile, in January 2019, collection of property advertisements through web scraping, instead of the manual data collection, was initiated. Since 2019, an asking price index for existing condominiums in the Colombo district has been compiled using web scraped data, rather than manually collected data. In 2020, data preparation and index compilation processes using the RWTD method for both new and advertised condominiums were improved continuously. Subsequently, following the RWTD methodology, property price index compilation was expanded to cover land and housing markets of the Colombo district using advertisement information collected through web scraping.

3.1 The different residential property price index compilation methodologies used

In constructing a suitable residential property price index for Sri Lanka, four index compilation methodologies, namely, Hedonic Characteristic Laspeyres Price Index, Hedonic Double Imputation Fisher Price Index, Hedonic Imputation Fisher Price Index and Hedonic Regression based Rolling Window Time Dummy methods were taken into consideration and these methodologies are detailed in this section.

[4] i.e., Hedonic Characteristic Laspeyres Price Index method, Hedonic Double Imputation Fisher Price Index method, Hedonic Imputation Fisher Price Index method and Hedonic Regression based Rolling Window Time Dummy method.

3.1.1 The Hedonic Characteristic Laspeyres Price Index

A regression equation is fitted for each t^{th} period considering price *(P)* as the dependent variable and *K* number of selected price determining factors as explanatory variables *(X),* where β represents their coefficients.

$$\boldsymbol{P^t = \sum_{K=0}^{K} \beta_K^t X_K^t + \varepsilon^t} \quad (1)$$

Then using the forward selection method, the regression equation with k (k<K) number of variables is identified as the best regression considering the statistical significance of variables. Coefficients derived for the explanatory variables and the characteristics values related to a selected average property are applied to the best fitted model to impute a price for each period.

$$P^t = \sum_{k=0}^{k} \beta_k^t x_k^t \quad (2)$$

Based on these imputed prices for both the based period and the t^{th} period, Laspeyres Price Index P^{0t} is calculated for period "t", compared with the base period "0".

$$P^{0t} = \frac{\sum_{k=0}^{k_2} \beta_k^t \, x_k}{\sum_{k=0}^{k_1} \beta_k^0 \, x_k} \quad (3)$$

3.1.2 The Hedonic Double Imputation Fisher Price Index

The Hedonic Double Imputation Fisher Price Index is computed using the Hedonic Double Imputation Laspeyres Price Index and Hedonic Double Imputation Paasche Price Index. A regression equation is fitted for each t^{th} period considering log value of price *(P)* as the dependent variable and k number of selected price determining factors as explanatory variables *(X),* where β represents their coefficients.

$$ln(P^t) = \sum_{k=0}^{k} \beta_k^t X_k^t + \varepsilon^t \quad (4)$$

Based on the coefficients, a price is imputed for each condominium unit due to the unavailability of matched prices to be considered in index compilation. For example, a property sold in period 0 does not have a matched sale price in period t. Therefore, a matched price for the period t is imputed using period 0 characteristics and the period t regression coefficients. Accordingly, a constant quality index can be compiled.

The equation used to impute log price for n^{th} unit in t^{th} period:

$$ln(P^t) = \sum_{k=0}^{k} \beta_k^t x_{nk}^t \quad (5)$$

To impute price for a property, β values and x values need to be assigned appropriately and the exponential value of the right side of the equation 5 needs to be calculated. Those price estimates are used for index compilation as follows.

The Hedonic Double Imputation Laspeyres (HDIL) Price Index

Under this method, prices are imputed for base period properties for both period t and the base period. HDIL price index indicates the effect on prices of base period properties at period t, compared to their base period prices.

Hedonic Double Imputation Laspeyres Price Index

$$(P_{HDIL}^{0t}) = \frac{Exp((\sum_{n=1}^{N_0} \sum_{k=0}^{k} \beta_k^t x_{nk}^0)/N_0)}{Exp((\sum_{n=1}^{N_0} \sum_{k=0}^{k} \beta_k^0 x_{nk}^0)/N_0)} \times 100 \quad (6)$$

The Hedonic Double Imputation Paasche (HDIP) Price Index

Under this method, prices are imputed for properties available/sold during period t for both period t and the base period. HDIP price index indicates the effect on prices of properties at period t, compared to their base period prices.

Hedonic Double Imputation Paasche Price Index

$$(P_{HDIP}^{0t}) = \frac{Exp((\sum_{n=1}^{N_t} \sum_{k=0}^{k} \beta_k^t x_{nk}^t)/N_t)}{Exp((\sum_{n=1}^{N_t} \sum_{k=0}^{k} \beta_k^0 x_{nk}^t)/N_t)} \times 100 \quad (7)$$

The Hedonic Double Imputation Fisher (HDIF) Price Index

HDIF price index is calculated by taking the geometric mean of HDIL price index and HDIF price index as follows.

$$(P_{HDIF}^{0t}) = [P_{HDIL}^{0t} . P_{HDIP}^{0t}]^{1/2} \quad (8)$$

3.1.3 The Hedonic Imputation Fisher Price Index

The Hedonic Imputation Fisher Price Index is computed using the Hedonic Imputation Laspeyres Price Index and Hedonic Imputation Paasche Price Index. A regression equation is fitted for each t^{th} period considering log value of price *(P)* as the dependent variable and k number of selected price determining factors as explanatory variables *(X)*, where β represents their coefficients, similar to the previous method. Under this approach, only the matched price is imputed.

The Hedonic Imputation Laspeyres (HIL) Index

The index is calculated based on the imputed prices of base period properties at period t, compared to their base period actual prices.

$$\text{Hedonic Imputation Laspeyres Index} = \frac{Exp(\sum_{n=1}^{N_0} \sum_{k=0}^{k} \beta_k^t x_{nk}^0)}{\sum_{n=1}^{N_0} p_n^0} \times 100 \quad (9)$$

The Hedonic Imputation Paasche (HIP) Index

The index is calculated based on the actual prices of t[th] period properties at period t, compared to their imputed prices for base period.

$$\text{Hedonic Imputation Paasche Index} = \frac{\sum_{n=1}^{N_t} p_n^t}{Exp(\sum_{n=1}^{N_t} \sum_{k=0}^{k} \beta_k^0 x_{nk}^t)} \times 100 \quad (10)$$

The Hedonic Imputation Fisher (HIF) Index

The price index is calculated by taking the geometric mean of HIL Index and HIP Index.

$$(P_{HIF}^{0t}) = [P_{HIL}^{0t}.\ P_{HIP}^{0t}]^{1/2} \quad (11)$$

3.1.4 The Rolling Window Time Dummy (RWTD) Method

Considering the standard version of the RWTD method with a window length of k+1 periods and the 1[st] period in the window is period t, the 1[st] step to estimate a semi-log hedonic model is as follows:

$$\ln p_h = \sum_{c=0}^{C} \beta_c Z_{hc} + \sum_{s=t+1}^{t+k} \delta_s D_{hs} \quad (12)$$

Where h indexes the property sales/advertised information data set, p_h the price and c indexes the set of selected characteristics of the property sold/advertised (eg: floor area, floor level, bedrooms, bathrooms, etc.). The explanatory variables are given by the Z_{hc} matrix, while D_{hs} is a matrix of dummy variables that equals 1 when s is the period in which the property is sold/advertised, and 0 otherwise.

The change in the price index from period t+k-1 to period t+k is then calculated as follows:

$$\frac{P_{t+k}}{P_{t+k-1}} = \frac{\exp(\delta_{t+k})}{\exp(\delta_{t+k-1})} \times 100 \quad (13)$$

Initial price index at period T, where T=t+k:

$$P_T = \frac{\exp(\delta_T)}{\exp(\delta_{T-1})} \times P_{T-1} \times 100 \quad (14)$$

3-month moving average index at period T:

$$P_{T'} = \frac{P_T + P_{T-1} + P_{T-2}}{3} \quad (15)$$

Final index series at period T, where the months/quarters of the base period are denoted by n and the number of months or quarters for the year are denoted by m:

$$P_{T''} = \frac{P_{T'} \times \mathrm{m} \times 100}{\sum_{n=1}^{m} P_{Base\ year\ n'}} \quad (16)$$

3.2 The index compilation process

This section discusses the index compilation process followed in developing the price indices. The data collected by scraping contains some erroneous or implausible data and need to be trimmed to form acceptable data sets. Therefore, data trimming was carried out monthly considering standard deviations of data distributions to remove extremely small or large properties from the data set and to retain the data which would reflect the commonly used types of real estate properties in the Colombo District.

Thereafter, variable transformations were identified in experimental work and bedroom/bathroom categories and location groups were assigned. In order to use the RWTD method, data pooling was done such that 12 months data were used for monthly series and 4 quarters data were used for quarterly series. In estimating regression models, the semi-log regression models were run in two stages: the first stage to identify outlier observations using Cook's Distance and the second stage to rerun the models without outliers. The coefficients obtained from the second regression output were used for compiling the RWTD Index. Accordingly, the initial index was compiled as per equation 14 and smoothing the initial index by taking the three months moving average to minimise the volatility of the monthly indices was done as per equation 15. Finally, the base period of the indices was set by converting the index such that the periodical average values for the selected base year equals to 100 as shown in equation 16.

4. Results and Discussion

Improvements made to the residential property price indices for Sri Lanka overtime, the process followed in compiling the current indices which adhere to international best practices and the movements of the price indices are discussed in this section.

4.1 Evaluation of different index compilation methodologies

Different index compilation methodologies were followed since 2016 in compiling real estate property price indices. Initially, the Hedonic Characteristic Laspeyres Price Index method was employed for condominium sales advertisements collected manually on a monthly basis from January 2016 to March 2017. Under its methodology, prices were imputed for a hypothetical unit which represent a property with average characteristics. However, it was challenging to select the perfect average property to track over time considering the heterogeneous nature of the market. This reduces the representative nature of wider market conditions in the index.

Secondly, the Hedonic Double Imputation Fisher Price Index method was employed during 2017 and 2018 using manually collected condominium advertisements. Although this satisfies the constant quality requirement, at certain times, the Hedonic Double Imputation Laspeyres Price Index and Hedonic Double Imputation Paasche Price Index moved in different directions resulting in the final index to be incompatible with the market conditions. Accordingly, it was observed that the Hedonic Double Imputation Fisher Price Index method was also not performing well with the manually collected data set. Further it was noted that the inadequate number of observations used was a constraint in following this methodology.

Since the results obtained using above methodologies for the manually collected data set failed to produce market representative residential property price indices in the Sri Lankan context, web scraping techniques were used since 2019 to expedite the data collection process while expanding the coverage.

Further, applying more statistical techniques for data pre-processing, experiments were carried out to compile the price indices following the Hedonic Double Imputation Fisher Price Index method, Hedonic Imputation Fisher Price Index method, and Hedonic Regression based Rolling Window Time Dummy method. Accordingly, property sales advertisement information was collected via web scraping monthly and this enabled collection of sufficiently large data sets required for index compilation. Thereafter, data sets were trimmed using trimming thresholds determined based on data distributions and summary statistics. Then, a descriptive analysis was conducted to identify the variables with suitable transformations which have an impact on the property prices considering their adequacy and significance, in order to be included in the regression model used for index compilation. Further, outliers were removed using Cook's Distance and the three months moving average technique was applied for smoothing purposes to minimise the statistical noise in the index, as followed by Patrick (2017). With these improvements in data pre-processing, testing was carried out for the three alternative index compilation methods mentioned above and the Hedonic Regression based Rolling Window Time Dummy method performed as the most suited method in compiling price indices for condominiums in Sri Lanka. Therefore, the index compilation process was expanded to cover the lands and housing markets as well, using the same methodology.

In the index compilation process following the Hedonic Regression based Rolling Window Time Dummy method, data manipulation is not carried out manually since the pre-defined methodologies are introduced for each stage. If any changes/instability is observed in coefficients which require a variable transformation different to the existing one, the change can be made without adjusting the past index series since the index follows a chain based method. If a new variable is identified as suitable to be included in the hedonic regression equation, that can be included once data are available for a time span of a window length, without changing the past index series.

4.2 Index development following the Hedonic Regression based Rolling Window Time Dummy Method

After selecting the most appropriate index compilation methodology, the steps followed in developing the price indices using the Hedonic Regression Based Rolling Window Time Dummy Method and the outcomes would be discussed in this section. A preliminary analysis was carried out for each index below to select the price determining characteristics to be used in the hedonic regression equations as explanatory variables. The variables or their transformations were selected based on their statistical significance and improvements to model specifications.

4.2.1 The Development of Asking Price Index for Condominiums

Data collection and transformation

Condominium sales advertisements published were collected through web scraping since January 2019 monthly using python codes developed to be run in Google Colab and Jupyter Notebook. The location of the property, number of bedrooms and bathrooms, floor area, availability of air conditioning, swimming pool, furniture, sea view, and price were the information that was collected. In addition, the distance to the city of Colombo from each property was found using Google Maps. In terms of data transformations, condominiums were classified into six predefined groups based on their location. The six location groups were determined with market experts' opinion considering the proximities and

market conditions. When the size of the condominium was not given in square feet but in a different measure, formulas were used to convert them. Further, both bedrooms and bathrooms were classified based on their number. A descriptive analysis was conducted to identify the variables which impact the prices of condominiums, in order to be included in the regression model used for index compilation.

Data trimming

The distribution of the condominium size (sq ft) was positively skewed for each month. Therefore, the data set was trimmed by removing condominiums lying beyond three standard deviations from both the lower and upper ends of the distribution monthly. Further, condominiums with more than six bedrooms/bathrooms and below 100 sq ft in size were removed considering the nature of commonly used condominiums in the Colombo District. However, more than 97 per cent of the advertisements were retained in the data set for each month after trimming.

Index compilation

The index was calculated according to the RWTD method using pooled data sets of 12 months. Initially, a semi-log least square regression was fitted considering the log of price of the condominium as the dependent variable and the price determining factors as explanatory variables. In terms of using the number of bedrooms and bathrooms in the regression equation, some experimental work was carried out and it was identified that it was better to consider as a categorical variable by grouping bedrooms and bathrooms, rather than a numerical variable. Among the alternative grouping combinations that were compared, it was decided to consider as two groups based on containing one or more bedrooms and bathrooms. Therefore, floor area, bedroom category, bathroom category, availability of a swimming pool, distance to Colombo city and location group and month of the advertisement were retained in the regression equation as explanatory variables. Among the property characteristics collected, the availability of air conditioning, furniture, and sea view were excluded from the equation due to their lack of statistical significance towards the price based on the results of the preliminary analysis.

Thereafter, outliers of the data set were removed using regression based Cook's distance method and another regression was fitted for the selected variables. More than 95 per cent of the data was retained in each 12-month window after removing the outliers. The exponential value of the coefficients for months of this regression was used for initial index compilation. After the initial index was compiled, a 3-month moving average index was calculated, and then the final index series was computed considering the base period as 2019, by setting the average of 3 months moving average series for 2019 equals to 100 and converting the monthly index numbers accordingly.

4.2.2 The Development of a Price Index for New Condominiums

Data collection and transformation

Sales information of new condominiums were collected quarterly through a survey of condominium developers since the third quarter of 2017. The information obtained includes the location of the condominium, number of bedrooms and bathrooms, floor area, floor level, whether it is in a mixed development project, availability of sea view and furniture, and the price. In addition, the distance to the city of Colombo from each condominium property was recorded manually using Google Maps. Following the similar techniques used in the asking price index for condominiums in terms of variable grouping, new condominiums were then classified into six

predefined groups based on their location and into two groups based on containing one or more bedrooms and bathrooms.

Data Trimming

The distribution of the condominium size (sq ft) was positively skewed for each quarter. Therefore, in order to remove the extremely small or large condominiums from the data set, condominiums lying beyond three standard deviations from both the lower and upper ends of the distribution on quarterly basis were removed by trimming. Further, condominiums with more than six bedrooms/bathrooms and below 100 sq ft in size were removed considering the nature of commonly used condominiums in the Colombo District. However, more than 97 per cent of the data were retained for each quarter after trimming.

Index compilation

The price index for new condominiums was calculated according to the Hedonic Regression based RWTD method using pooled data sets of four quarters. Initially, a semi-log least square regression was fitted considering the log of price of the condominium as the dependent variable and floor area, floor level, bedroom category, bathroom category, whether it is in a mixed development project, distance to the city of Colombo and location group and quarter in which the condominium was sold, as explanatory variables. Among the available property characteristics, availability of furniture and sea view were excluded from the equation since they become statistically insignificant based on the results of the preliminary analysis. Thereafter, outliers of the data set were removed using regression based Cook's distance method and another regression was fitted for the same variables. More than 92 per cent of the data were retained in each 4-quarter window after removing the outliers. The exponential value of the coefficients for quarters of this regression were used for initial index compilation. After the initial index was compiled, the final index series was computed considering the base period as 2019 by setting the average of initial index values of the four quarters of 2019 equals to 100 and converting the quarterly index numbers accordingly.

4.2.3 The Development of Asking Price Index for Lands

Data collection and transformation

Land sales advertisements published were collected through web scraping using python codes developed to be run in Google Colab and Jupyter notebook since January 2019 monthly. The location of the land, number of perches to be sold, and price per perch were the information that was collected. The total price of the land was used in the index compilation process as the dependent variable. Therefore, when the price was not displayed in total and when the land size was not displayed in perches in the advertisement, formulas were used to convert them into the required term. Further, lands were also classified into six predefined groups based on their location.

Data trimming

The distribution of the land size was highly positively skewed for each month. Therefore, in order to remove the extremely small or large lands from the data set, lands lying beyond three standard deviations from the lower end and 0.5 standard deviations from the upper end based on the land size (perches) distribution on a monthly basis and below four perches, were removed considering the nature of residential lands in the Colombo District. However, more than 99 per cent of the advertisements were retained in the data set for each month after trimming.

Index compilation

The asking price index for lands was calculated according to RWTD method using pooled data sets of 12 months. Initially, a semi-log least square regression was fitted considering the log of total price of the land as the dependent variable and the number of perches, location group and month of the advertisement as explanatory variables. Thereafter, outliers of the data set were removed using regression based Cook's distance method and another regression was fitted for the same variables. More than 95 per cent of the data was retained in each 12-month window after removing the outliers. The exponential value of the coefficients for months of this regression was used for initial index compilation. After the initial index was compiled, a 3-month moving average index was calculated, and the final index series was computed considering the base period as 2019, by setting the average of 3-month moving average series for 2019 equals to 100 and converting the monthly index numbers accordingly.

4.2.4 The Development of Asking Price Index for Houses

Data collection and transformation

House sales advertisements published were collected since October 2019 monthly using python codes developed to be run in Google Colab and Jupyter Notebook. The location of the house, land plot size (perches), house size (sq ft), number of bedrooms and bathrooms, and price of the property were the information that was collected. When the house size was not displayed in sq ft and land size was not displayed in perches in the advertisement, formulas were used to convert them. Further, houses were classified into six predefined groups based on their location as per market experts' opinion. In terms of bedroom and bathroom grouping, they were classified into three groups based on having one or two, three or four, or more than four bedrooms and bathrooms, based on experimental level analysis.

Data trimming

The distributions of the land size and house size were highly positively skewed for each month. Therefore, in order to remove the extremely small or large landed houses from the data set, properties lying beyond three standard deviations from the lower end and two standard deviations from the upper end based on the land size (perches) distribution first and then on house size (sq ft) distribution on a monthly basis were removed by trimming. Further, houses below 100 sq ft in size and built on lands below four perches were removed considering the nature of commonly used houses in the Colombo District. However, more than 95 per cent of the advertisements were retained in the data set for each month after trimming.

Index compilation

The index was calculated applying the RWTD method on pooled data sets of 12 months. Initially, a semi-log least square regression was fitted considering the log of price of the property as the dependent variable and the size of the land, size of the house, bedroom category, bathroom category, location group, and month of the advertisement as explanatory variables. Thereafter, outliers were removed using Cook's distance method and another regression was fitted for the same variables. More than 91 per cent of the data was retained in each 12-month window after removing the outliers. The exponential value of the coefficients for months of this regression were used for initial index compilation and then a 3-month moving average index was calculated. The final index series was computed considering the base period as 2019, by setting the average of 3-month moving average series for 2019 equals to 100 and converting the monthly index numbers accordingly.

Details of the four types of real estate property price indices compiled covering the Colombo District of Sri Lanka are summarised in the tables below.

Table 1. Summary of the Real Estate Property Price Indices

	Condominiums		Lands	Houses
Index Name	Price Index for New Condominiums	Asking Price Index for Condominiums	Asking Price Index for Lands	Asking Price Index for Houses
Frequency	Quarterly	Monthly	Monthly	Monthly
Data Source	Condominium Market Survey	Sales Advertisements	Sales Advertisements	Sales Advertisements
Coverage	Colombo District	Colombo District	Colombo District	Colombo District
Methodology	Hedonic Regression based Rolling Window Time Dummy	Hedonic Regression based Rolling Window Time Dummy	Hedonic Regression based Rolling Window Time Dummy	Hedonic Regression based Rolling Window Time Dummy
Initiation	3rd Quarter 2017	January 2019	January 2019	October 2019
Base Year	2019	2019	2019	2019
Window Length	4 Quarters	12 Months	12 Months	12 Months
Variables	Log (price), floor area, floor level, bedroom and bathroom category, whether it is in a mixed development project, distance to the Colombo city and location group	Log (price), floor area, bedroom and bathroom category, availability of a swimming pool, distance to the Colombo city and location group	Log (price), the number of perches and location group	Log (price), land plot size, house size, bedroom and bathroom category and location group
Data Trimming	Condominiums within 3 standard deviations from both lower and upper ends of the floor area distribution on monthly basis and those with less than 6 bedrooms/ bathrooms, and above 100 sq.ft. were considered	Condominiums within 3 standard deviations from both lower and upper ends of the floor area distribution on monthly basis and those with less than 6 bedrooms/ bathrooms, and above 100 sq.ft. were considered	Lands within 3 standard deviations from the lower end and 0.5 standard deviations from the upper end based on the land size (perches) distribution on monthly basis and those with a land size of more than 4 perches were considered	Houses between 3 standard deviations from the lower end and 2 standard deviations from the upper end based on the land size (perches) and house size (sq.ft.) distributions, and those above 100 sq.ft. built on lands more than 4 perches were considered

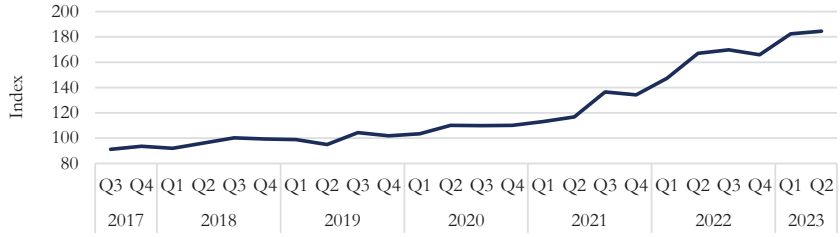
Table 2. Details of Data Used in Index Compilation

	Price Index for New Condominiums	*Asking Price Indices for*		
		Condominiums	Lands	Houses
Average % of Data Retained after Trimming	97.9	97.9	99.6	95.4
Average % of Data Retained after Outlier Removal	92.3	95.5	95.0	91.1
Average no. of Data Points in a Window	981	7,215	30,576	38,500
Average R^2 of Hedonic Regression Models	0.95	0.74	0.71	0.78

4.3 Movements of real estate property price ındices compiled for Sri Lanka

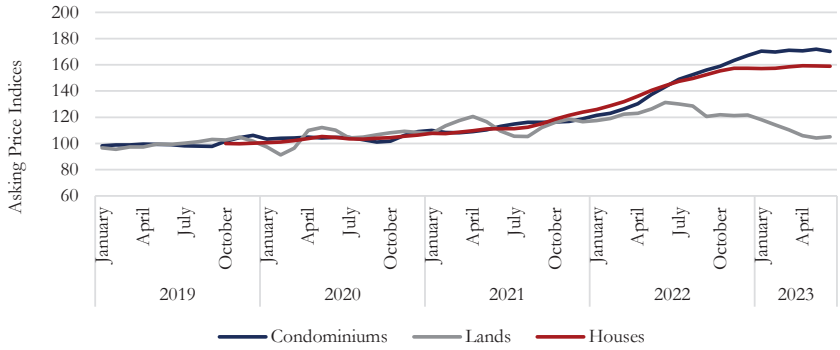
The latest available real estate property price indices compiled using the hedonic regression based RWTD method covering the Colombo Distrcit of Sri Lanka are presented below.

Figure 1. Price Index for New Condominiums (2019=100)



The index is compiled using sales information of new condominiums, which are obtained quarterly through the survey and thus reflects the primary condominium market.

Figure 2. Asking Price Indices for Condominiums, Lands and Houses (2019=100)



The asking indices for condominiums, lands, and houses are compiled using sales advertisements since 2019. Thus, the asking price indices for condominiums and houses represent both primary and secondary markets.

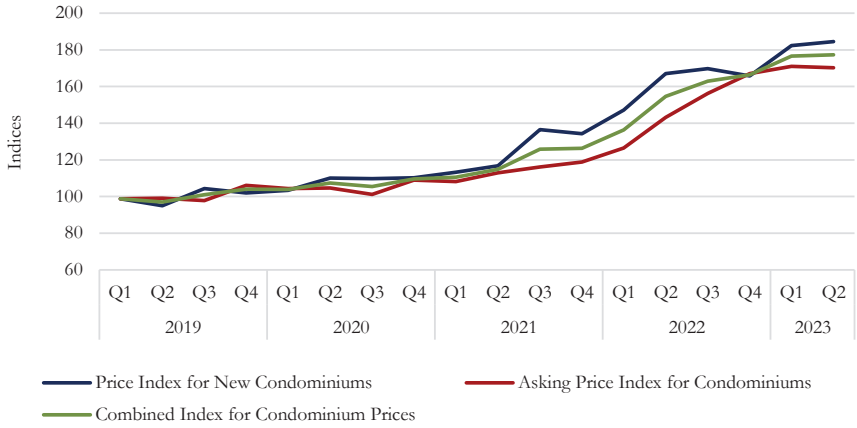
The four price indices reflect the trends observed in the data and the indices are more appropriate to observe long term trends, even if there are certain ups and downs in the short term. The regression equations used for index compilation explain more than 70 per cent[5] of the relevant data sets, which implies the level of precision of the indices. As per the index movements, new condominium prices are rapidly increasing while the asking prices of both condominiums and houses indicate stagnation at high price levels. Further, it is observed that the asking price indices for condominiums and houses follow similar trends with a strong positive correlation of 0.98. In the meantime, land prices have shown a declining trend during the last year.

In terms of the of the real estate market conditions in Sri Lanka, high construction costs that prevailed during the recent years owing to a depreciated currency, shortages of construction material due to

[5] Average R^2 of Hedonic Regression Models are shown in Table 2.

import restrictions, and high interest rates mainly led to an increase in prices of houses and condominiums. In addition, taxes imposed on condominium sales such as the 15 per cent value added tax and the 2.5 per cent social security contribution levy contributed further towards the price increases in condominiums. These factors created adverse market conditions which lowered the demand. Even against this background, property developers follow a wait and see approach and are not ready to reduce the prices to attract buyers due to their high-cost burden. As per real estate property experts', the real estate market is complex, and supply and demand are just one of many factors that impact property prices. Accordingly, the current market circumstances in Sri Lanka seems to be a cost push price increase rather than a demand push price increase. These market conditions are reflected in the price index for new condominiums. The price effect on new properties in the primary market is eventually reflected in the secondary market where prices of existing properties also remain stagnant at high levels. Further, the uncertainties in economic and political conditions affect the decisions of both buyers and sellers in purchasing properties for personal use and as an investment option. Consequently, the land market remains less active compared to the other types of properties as per market participants' views. Thus, the asking prices for condominiums, houses, and lands reflect the prevailing market conditions even at the recent turbulent environment.

Figure 3. Price Indices for Condominiums (2019=100)



In order to scrutinise the price indices for condominiums, a combined index is calculated by taking the geometric mean of the price index for new condominiums and the asking price index for condominiums[6]. As observed in Figure 3, the asking price index compiled based on the advertised prices and the price index compiled based on the actual selling prices follow similar trends over time and indicate a strong positive correlation of 0.97.

Similar studies are carried out by Shimizu *et al.* (2010) and Lyons (2019), demonstrating that the trends of prices collected at different stages of condominium buying and selling process align with each other even

[6] Since the asking price index is a monthly series which consists of three-month moving averages, the last month value of a quarter can be considered to match with the quarterly series of price index for new condominiums.

though there are differences in the price levels. Further, results of Lyons suggest that even in extreme market conditions, advertisement based price indices are a good measure of ultimate sale prices. Therefore, based on previous literature and the outcomes of this study, it is understood that the different price types collected at different stages of the property transaction process can be used in constructing property price indices.

5. Conclusion

In developing a property price index, a reliable data source, and a suitable price index compilation methodology are the principal components. The data source to be used for real estate property price indices can be collected at different stages of its buying and selling process and advertised prices can be used to develop asking price indices which performs as a good indicator to observe market price trends in the absence of official transaction data. Among the alternative residential property price index compilation methods, the Hedonic Regression based Rolling Window Time Dummy Method is identified as the best fitted method for thin real estate markets like Sri Lanka since it uses a pooled set of data for each window. The model specification improvement techniques, the index compilation method and smoothing techniques used are internationally accepted procedures followed by many countries. Thus, these indices are useful to observe trends in real estate property prices of Sri Lanka where no other reliable indices are available to monitor the sector.

References

De Haan, J. (2004). "Direct and Indirect Time Dummy Approaches to Hedonic Price Measurement". *Journal of Economic and Social Measurement*, vol. 29, no. 4, pp. 427-443.

Diewert, W.E. (2011). "Alternative Approaches to Measuring House Price Inflation". *Discussion Paper 10-10, Department of Economics, The University of British Columbia, Vancouver, Canada*.

Diewert, W.E., Heravi, S. and Silver, M. (2007). "Hedonic Imputation versus Time Dummy Hedonic Indexes". *IMF Working Paper. WP/07/234*.

Diewert, W.E. and Shimizu, C. (2014). "Residential Property Price Indexes for Tokyo". *Discussion Paper 13-07, School of Economics, The University of British Columbia, Vancouver, Canada*.

Eurostat, European Commition. (2013). *Handbook on Residential Property Price Indices (RPPIs).*

Hill, R. J., Scholz, M. and Shimizu, C. (2017). "Weekly Hedonic House Price Indices and the Rolling Time Dummy Method: An Application to Sydney and Tokyo". *Conference Proceedings at Hitotsubashi-RIETI International Workshop on Real Estate and the Macro Economy in Tokyo.*

Lyons, R. C. (2019). "Can list prices accurately capture housing price trends? Insights from extreme markets conditions". *Finance Research Letters,* vol. 30, pp. 228-232.

Patrick, G. (2017). "Redeveloping Ireland's Residential Property Price Index (RPPI)". *Conference Proceedings at Fifteenth Meeting of the Ottawa Group 2017 – International Working Group on Price Statistics.*

Shimizu, C., Nishimura, K. and Watanabe, T. (2010). "Housing Prices in Tokyo: A Comparison of Hedonic and Repeat Sales Measures". *Jahrbücher für Nationalökonomie und Statistik,* vol. 230, no.6, pp. 792-813.

Shimizu, C., Takatsuji, H., Ono, H. and Nishimura, K.G. (2010). "Structural and Temporal Changes in the Housing Market and Hedonic Housing Price Indices". *International Journal of Housing Markets and Analysis,* vol. 3, no. 4, pp. 351-368.

Shimizu, C., Nishimura, K.G. and Watanabe, T. (2011). "House Prices from Magazines, Realtors, and the Land Registry". *Property Market and Financial Stability, BIS Papers No.64, Bank of International Settlements,* pp. 29-38.

Song, H. S. and Wilhelmsson, M. (2010). "Improved price index for condominiums". *Journal of Property Research,* vol. 27, no.1, pp. 39-60.