Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP152021

Hedonic Models and House Price Index Numbers

Robert J. Hill and Alicia N. Rambaldi

Date: October 2021

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Hedonic Models and House Price Index Numbers*

Robert J. Hill,

Department of Economics, University of Graz, 8010 Graz, Austria

Alicia N. Rambaldi,

School of Economics, The University of Queensland, St Lucia, QLD 4072, Australia.

October 2021

Contents

*We would like to thank Chihiro Shimizu and D.S. Prasada Rao for comments on an earlier version. Rambaldi acknowledges funding support from the Australian Research Council, DP120102124 and LP160101518.

Abstract

We survey some recent developments in the literature on hedonic price indices for housing. The main classes of hedonic methods are presented along with some new methods that have become popular recently. A number of new approaches are then considered for controlling for location in hedonic models. Next we consider how hedonic models can be used to construct separate price indices for land and structures. Significant progress has been made recently in this field. The survey concludes with a discussion of ways of computing higher frequency (e.g., weekly) hedonic price indices, and ways of deriving house price indices for the whole housing stock, as opposed to just those properties that have traded recently.

Keywords: Adoption of hedonic indices; Controlling for location; Land and structure indices; Higher frequency indices

JEL codes: (JEL. C31; C43; E01; E31; E52; R31)

1 Introduction

This chapter provides a review of the literature on hedonic methods for constructing property price indices for residential housing. A number of methods for constructing price indices are based on the use and estimation of a hedonic regression, and thus are referred to in general as hedonic methods. However, there is a fundamental difference between methods that compute indices directly from the estimated parameters of the hedonic regression - the time-dummy method (Section 2.1.1) and the rolling time-dummy method (Section 2.1.4)- and those that compute indices from the imputed prices obtained from a hedonic regression model - the average-characteristics method (Section 2.1.2), the hedonic imputation method (Section 2.1.3), and the repricing method (Section 2.1.5).

Indices that are computed directly from parameter estimates have the advantage of readily providing standard errors; however, any biases due to model specification, such as omitted variables in the regression, are carried to the computed indices. Thus, careful model specification is required. Alternatively, indices that are computed from imputed prices require good prediction performance from the hedonic model. The concern is less on which individual variables to include in the model or whether there is collinearity. A well performing imputation model can include a number variables that cover key predictors of property prices, such as location (see Section 2.2), and controls for land and dwelling characteristics that jointly explain the movements and distribution of property prices. Typical sample sizes are large enough so that degrees of freedom are not a concern, and collinearity will not affect the computed index since it does not affect predictions (i.e. imputations). Hedonic imputation methods were signalled as the preferred alternative in the Handbook on Residential Property Price Indices (HRPPI) (Eurostat, 2013) [19], which has been the most comprehensive compendium to date on methodology to construct residential property price indices. However, the field has moved on since the HRPPI was published. Many National Statistical Institutes (NSIs) are now using hedonic methods in their official indices that were not discussed in the HRPPI.

In Section 2 we provide a comprehensive review of hedonic price index formulas building from earlier reviews and current practice at NSIs. This Section also provides a systematic review of parametric and non-parametric alternatives to controlling for the location of the property in hedonic models used to construct price indices.

The hedonic approach is appealing in its flexibility. In particular, it can be extended to address other questions beyond basic index construction. A notable example is considered in Section 3. This Section shows how hedonic methods can be used to construct separate price indices for land and structure. Section 4 reviews two other recent developments in the hedonic literature. The first is the computation of price indices at higher frequencies. The second is the construction of hedonic price indices for the whole housing stock rather than just for properties that have sold recently. Section 5 concludes.

2 Hedonic Methods for Constructing House Price Indices

This section is divided into two subsections. The first presents some of the hedonic methods that were covered in the HRPPI as they provide the basic framework to review a number of methods that have gained popularity since the HRPPI was published. In addition, one method – the repricing method – that was not discussed in the HRPPI is also considered. The second subsection considers ways of controlling for location effects in hedonic models.

2.1 Indices Covered in Chapter 5 of Eurostat's HRPPI

2.1.1 Time-dummy method

The time-dummy method estimates a single semi-log hedonic model as follows:

$$\ln p_h = \sum_{c=1}^{C} \beta_c z_{ch} + \sum_{s=b+1}^{t} \delta_s d_{sh} + \varepsilon_h, \tag{1}$$

where h indexes all the housing transactions between periods b and t, p_h is the transaction price of property h, c indexes the set of available characteristics of the transacted properties, and ε is an identically, independently distributed error term with mean zero. The characteristics of the properties are given by $z_{c,h}$, while $d_{s,h}$ are dummy variables that equal 1 when a property was traded in period s, and zero otherwise.

The price index for period t relative to the base period b is then calculated as follows:

$$\frac{P_t}{P_b} = \exp(\hat{\delta}_t) \tag{2}$$

where $\hat{\delta}$ denotes the least squares estimate of δ.

The time-dummy method has three main attractions. First it is relatively simple to use. Second, given it uses the full dataset it does not need as much data per period as other hedonic methods. Third it provides standard errors on the estimated price indices.

The time-dummy method has two main weaknesses. First, the shadow prices can become stale, not reflecting the current state of the market when the hedonic model is estimated over many years. Second, whenever a new period is added to the dataset and the hedonic model re-estimated, all the price indices change.

2.1.2 Average characteristic method

The average characteristics method and the hedonic imputation method both begin by estimating the following semi-log hedonic model separately for each period. For example, for periods $t-1$ and t, the regression model takes the following forms:[1]

$$\ln p_{t-1,h} = \sum_{c=1}^{C} \beta_{t-1,c} z_{t-1,c,h} + \varepsilon_{t-1,h}, \tag{3}$$

$$\ln p_{t,h} = \sum_{c=1}^{C} \beta_{t,c} z_{t,c,h} + \varepsilon_{t,h}, \tag{4}$$

where h indexes the property transactions in period t, $p_{t,h}$ the transaction price, and $z_{t,h,c}$ is the level of characteristic c in dwelling h. No time dummies are included. The estimated shadow prices on the characteristics, $\beta_{t,c}$, are specific to period t and are updated every period.

[1] This section draws extensively on Hill et al., (2018) [30].

A reference period is selected and an average basket of characteristics constructed for this period. This average basket of characteristics can be interpreted as an average property. The hedonic price index simply measures the change in the imputed price of this average property over time. A price index between periods $t-1$ and t can now be calculated using the average property of period $t-1$ (denoted by $\bar{z}_{t-1}$) as the reference:

$$\frac{P_t}{P_{t-1}} = \frac{\exp(\sum_{c=1}^{C} \hat{\beta}_{t,c}\bar{z}_{t-1,c})}{\exp(\sum_{c=1}^{C} \hat{\beta}_{t-1,c}\bar{z}_{t-1,c})} = P^L_{t-1,t}, \tag{5}$$

where $P^L_{t-1,t}$ denotes a Laspeyres-type price index between periods $t-1$ and t.

Alternatively, the average property of period t could be used as the reference as follows:

$$\frac{P_t}{P_{t-1}} = \frac{\exp(\sum_{c=1}^{C} \hat{\beta}_{t,c}\bar{z}_{t,c})}{\exp(\sum_{c=1}^{C} \hat{\beta}_{t-1,c}\bar{z}_{t,c})} = P^P_{t-1,t}, \tag{6}$$

where $P^P_{t-1,t}$ denotes a Paasche-type price index. The terms $\bar{z}_{t-1,c}$ and $\bar{z}_{t,c}$ in (5) and (6) denote the average baskets of characteristics of periods $t-1$ and t.

$$\bar{z}_{t-1,c} = \frac{1}{H_{t-1}} \sum_{h=1}^{H_{t-1}} z_{t-1,h,c}, \qquad \bar{z}_{t,c} = \frac{1}{H_t} \sum_{h=1}^{H_t} z_{t,h,c}.$$

If one wants to treat both periods symmetrically, this can be done by taking the geometric mean of $P^L_{t-1,t}$ and $P^P_{t-1,t}$.

Each period the average characteristic basket is updated. Focusing on the Laspeyres case, relative to the base period b, the price index for period t is calculated as follows:

$$\frac{P_t}{P_b} = P^L_{b,b+1} \times P^L_{b+1,b+2} \times \cdots \times P^L_{t-1,t}.$$

The average characteristics method is still relatively simple and more market relevant than the time-dummy method in that the characteristic shadow prices are continually updated. However, estimating a separate hedonic model for each period can be problematic for smaller datasets. A second concern relates to the definition and interpretation of the average property. In particular, characteristics that take the form of dummy variables are probably best allocated fractionally to each category in proportion to the frequency in which they are observed.

2.1.3 Hedonic imputation method

The hedonic imputation method can be viewed as an extended version of the average characteristics method. Under certain conditions, as is shown below, the two methods are equivalent.

The underlying rationale of the hedonic imputation method is to use the hedonic model to impute missing prices so as then to allow standard price index formulas to be used.

Again we begin by estimating separate hedonic models for each period, as in (3) and (4). Geometric-Laspeyres and geometric-Paasche-type formulas can now be computed as follows:[2]

$$\text{Geometric Laspeyres (GL)}: \quad \frac{P_t}{P_{t-1}} = \left[\prod_{h=1}^{H_{t-1}} \frac{\hat{p}_{t,h}(z_{t-1,h})}{\hat{p}_{t-1,h}(z_{t-1,h})} \right]^{1/H_{t-1}}; \tag{7}$$

$$\text{Geometric Paasche (GP)}: \quad \frac{P_t}{P_{t-1}} = \left[\prod_{h=1}^{H_t} \frac{\hat{p}_{t,h}(z_{t,h})}{\hat{p}_{t-1),h}(z_{t,h})} \right]^{1/H_t}, \tag{8}$$

where $\hat{p}_{t,h}(z_{t-1,h})$ in (7) represents the predicted price in period t of a property with characteristic vector $z_{t-1,h}$ obtained from the hedonic model of period t, while $\hat{p}_{t-1,h}(z_{t-1,h})$ denotes the predicted price of the same property in period $t-1$ obtained from the hedonic model of period $t-1$. The terms in (8) have analogous interpretations.

Here we consider only double imputation indices. This means that both the numerator and denominator in each price relative is imputed. By contrast, single imputation imputes only the numerator for GL and only the denominator for GP (see Silver and Heravi, 2001 [48]; Pakes, 2003 [38]; de Haan, 2004 [12]; and Hill and Melser, 2008 [27]). Double imputation is generally preferred since it partially controls for omitted variables in each price relative (see Hill and Melser, 2008 [27]).

[2] Again this section draws extensively on Hill et al. (2018) [30].

Taking the geometric mean of GL and GP we obtain a Törnqvist-type index:

$$\text{Törnqvist}: \quad \frac{P_t}{P_{t-1}} = \left\{ \left[\prod_{h=1}^{H_{t-1}} \frac{\hat{p}_{t,h}(z_{t-1,h})}{\hat{p}_{t-1,h}(z_{t-1,h})} \right]^{1/H_{t-1}} \left[\prod_{h=1}^{H_t} \frac{\hat{p}_{t,h}(z_{t,h})}{\hat{p}_{t-1,h}(z_{t,h})} \right]^{1/H_t} \right\}^{1/2}. \tag{9}$$

When the underlying hedonic model is semi-log, GL, GP and Törnqvist hedonic imputation indices can likewise be represented as average characteristic methods as follows (Hill and Melser, 2008 [27]):

$$\text{GL}: \quad \left[\prod_{h=1}^{H_{t-1}} \frac{\hat{p}_{t,h}(z_{t-1,h})}{\hat{p}_{t-1,h}(z_{t-1,h})} \right]^{1/H_{t-1}} = \frac{\exp(\sum_{c=1}^{C} \hat{\beta}_{t,c} \bar{z}_{t-1,c})}{\exp(\sum_{c=1}^{C} \hat{\beta}_{t-1,c} \bar{z}_{t-1,c})} = P^L_{t-1,t}; \tag{10}$$

$$\text{GP}: \quad \frac{P_t}{P_{t-1}} = \left[\prod_{h=1}^{H_t} \frac{\hat{p}_{t,h}(z_{t,h})}{\hat{p}_{t-1,h}(z_{t,h})} \right]^{1/H_t} = \frac{\exp(\sum_{c=1}^{C} \hat{\beta}_{t,c} \bar{z}_{t,c})}{\exp(\sum_{c=1}^{C} \hat{\beta}_{t-1,c} \bar{z}_{t,c})} = P^P_{t-1,t}; \tag{11}$$

$$\text{Törnqvist}: \quad \left\{ \left[\prod_{h=1}^{H_{t-1}} \frac{\hat{p}_{t,h}(z_{t-1,h})}{\hat{p}_{t-1,h}(z_{t-1,h})} \right]^{1/H_{t-1}} \left[\prod_{h=1}^{H_t} \frac{\hat{p}_{t,h}(z_{t,h})}{\hat{p}_{t-1,h}(z_{t,h})} \right]^{1/H_t} \right\}^{1/2}$$

$$= \left\{ \frac{\exp[\sum_{c=1}^{C} \hat{\beta}_{t,c}(\bar{z}_{t-1,c} + \bar{z}_{t,c})]}{\exp[\sum_{c=1}^{C} \hat{\beta}_{t-1,c}(\bar{z}_{t-1,c} + \bar{z}_{t,c})]} \right\}^{1/2} = \left(P^L_{t-1,t} \times P^P_{t-1,t} \right)^{1/2}. \tag{12}$$

This duality between the average characteristics and hedonic imputation methods breaks down when the functional form of the hedonic model is not semi-log.

2.1.4 Rolling time-dummy method

The rolling time dummy (RTD) method estimates a time-dummy hedonic model on a rolling window of time periods (see Shimizu et al., 2010 [47], and O'Hanlon, 2011 [34]). Each time a new period of data become available, the rolling window is moved forward one period and the hedonic model re-estimated.

Price indices are derived from the estimated coefficients on the time dummies in the same way as with the time-dummy method except that each time the hedonic model is estimated we are only interested in the coefficient on the last period in the rolling window.[3]

[3] More sophisticated versions of the RTD method are developed in Hill et al. (2021b) [31].

Here we denote the first period in the window as period t. A semi-log hedonic model is estimated with a $k+1$ period window as follows:[4]

$$\ln p_h = \sum_{c=1}^{C} \beta_c z_{c,h} + \sum_{s=t+1}^{t+k} \delta_s d_{s,h} + \varepsilon_h, \tag{13}$$

where h indexes the housing transactions that occur within the rolling window. The set of available characteristics is indexed by c. The transaction price of property h is denoted by p_h, the property characteristics by $z_{c,h}$, and time dummy variables capturing the period in which property h is sold by $d_{s,h}$. Finally, ε is a random error term with mean zero.

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

$$\frac{P_{t+k}}{P_{t+k-1}} = \frac{\exp(\hat{\delta}^{t}_{t+k})}{\exp(\hat{\delta}^{t}_{t+k-1})}, \tag{14}$$

where $\hat{\delta}$ denotes the least-squares estimate of δ. The superscript t indicates that the estimated δ coefficient was obtained from the hedonic model with period t as the base (i.e., $P_t = 1$). This hedonic model is used only to compute the change in house prices from period $t+k-1$ to $t+k$. The window is then rolled forward one period and the hedonic model re-estimated. The price index comparing periods $t+k$ and $t+k+1$ is now computed as follows:

$$\frac{P_{t+k+1}}{P_{t+k}} = \frac{\exp(\hat{\delta}^{t+1}_{t+k+1})}{\exp(\hat{\delta}^{t+1}_{t+k})}. \tag{15}$$

In (15) the base period in the hedonic model is now $t+1$. Over multiple periods the price index is computed by chaining as follows:

$$\frac{P_{t+k+1}}{P_t} = \left[\frac{\exp(\hat{\delta}^{t-k}_{t+1})}{\exp(\hat{\delta}^{t-k}_{t})}\right] \left[\frac{\exp(\hat{\delta}^{t-k+1}_{t+2})}{\exp(\hat{\delta}^{t-k+1}_{t+1})}\right] \times \cdots \times \left[\frac{\exp(\hat{\delta}^{t+1}_{t+k+1})}{\exp(\hat{\delta}^{t+1}_{t+k})}\right]. \tag{16}$$

Unlike the time-dummy method, RTD price indices are never revised as new periods of data become available.

[4] This section draws extensively on Hill et al., (2021b) [31].

The RTD method is used by some European countries (Croatia, Cyprus, France, Ireland and Portugal) to compute their official house price indices (Hill et al, 2018 ([30]). In addition, Japan's Official Property Price Index since 2012 uses RTD (Real Estate and Construction Economy Bureau, 2020 [45]). Japan has recently decided likewise to compute its commercial property price indices using the RTD method (see Shimizu and Diewert, 2019 [46]). Brunei Darussalam (see https://www.ambd.gov.bn/Site%20Assets%20%20News/RPPI-Technical-Notes.pdf), Peru and Thailand (see https://www.bot.or.th/App/BTWS_STAT/statistics/DownloadFile.aspx?file=EC_EI_008_S2_ENG.PDF) also use RTD, and Indonesia is about to start using it (see Rachman, 2019 [39]).

Given its popularity, it is perhaps surprising that RTD was not discussed in chapter 5 of the HRPPI (on hedonic regression methods), except for the special case of a two-quarter rolling window – sometimes also referred to as the adjacent period method (see for example Triplett, 2004 [50]). The RTD method was, however, discussed in chapters 8 and 12, where it is recommended. The main reason RTD was excluded from chapter 5 was probably because in 2013 it was still quite new. Also, RTD is a variant on the time-dummy method.

The reason RTD is now widely used is due to its attractive features. The method allows the index provider to choose the window length. This involves a trade-off. A longer window allows more data to be used each time the hedonic model is estimated, increasing the efficiency of the parameter estimates. By contrast, a shorter window increases the market relevance of the estimated shadow prices. Hence in general bigger countries should choose shorter windows than smaller countries. The official house price index for France, for example, has a two-quarter rolling window, while the house price indices of Croatia and Cyprus have four-quarter rolling windows. This flexibility, combined with its simplicity and non-revisability, explains why the RTD method is becoming increasingly popular in recent years.

2.1.5 Repricing method

The repricing method is used by Austria, Finland, Hungary, Italy, Latvia, Luxembourg, Norway, and Slovenia to compute their official house price indices.[5] The repricing method, which is related to the average characteristics method, estimates a semi-log hedonic model using only the data of the base year b. The hedonic model can be written as follows:

$$\ln p_{b,h} = \sum_{c=1}^{C} \beta_{b,c} z_{b,h,c} + \varepsilon_{b,h}, \tag{17}$$

where h denotes a property sold in year b, $c = 1, \ldots, C$ indexes the characteristics of properties available in the dataset (such as floor area or number of bedrooms), and ε is a random error term.

The repricing price index formula divides a quality-unadjusted price index (QUPI) by a quality-adjustment factor (QAF). The QUPI is the ratio of the geometric mean prices in both periods $t-1$ and t, computed as follows:

$$QUPI_{t-1,t} = \frac{\tilde{p}_t}{\tilde{p}_{t-1}}, \tag{18}$$

where $\tilde{p}_{t-1}$ and $\tilde{p}_t$ denote the geometric mean price of properties sold in periods $t-1$ and t, respectively.

$$\tilde{p}_{t-1} = \prod_{h}^{H_{t-1}} \left(p_{t-1,h}\right)^{1/H_{t-1}}, \quad \tilde{p}_t = \prod_{h}^{H_t} \left(p_{t,h}\right)^{1/H_t}. \tag{19}$$

In (19), H_{t-1} and H_t denote the number of properties sold in periods $t-1$ and t.

The quality adjustment factor (QAF) uses the characteristic shadow prices $\hat{\beta}_b$ of year b to compare the cost of buying the average properties of periods $t-1$ and t as follows:

$$QAF_{t-1,t} = \frac{\exp(\sum_{c=1}^{C} \hat{\beta}_{b,c} \bar{z}_{t,c})}{\exp(\sum_{c=1}^{C} \hat{\beta}_{b,c} \bar{z}_{t-1,c})}. \tag{20}$$

In (20),

$$\bar{z}_{t-1,c} = \frac{1}{H_{t-1}} \sum_{h=1}^{H_{t-1}} z_{t-1,h,c}, \quad \bar{z}_{t,c} = \frac{1}{H_t} \sum_{h=1}^{H_t} z_{t,h,c},$$

denote the average basket of characteristics of periods $t-1$ and t.

[5] This section draws extensively on Hill et al. (2018) [30].

The repricing price index is obtained by dividing the quality-unadjusted index (QUPI) in (18) by the quality adjustment factor (QAF) in (20) as follows:

$$\frac{P_t}{P_{(t-1}} = \frac{QUPI_{t-1,t}}{QAF_{t-1,t}} = \frac{\tilde{p}_t}{\tilde{p}_{t-1}} \bigg/ \frac{\exp(\sum_{c=1}^{C} \hat{\beta}_{b,c} \bar{z}_{t,c})}{\exp(\sum_{c=1}^{C} \hat{\beta}_{b,c} \bar{z}_{(t-1),c})} \, , \tag{21}$$

One attractive feature of the repricing method is that the characteristic shadow prices can be calculated using a full year of data, even when the index itself is being calculated on a quarterly basis. By contrast a quarterly average characteristics index computes the shadow prices each time using only one quarter's data, which can be problematic for smaller countries with less transactions per quarter.

Another interesting feature of the repricing method is that it only requires one set of shadow prices. However, failure to update the base period shadow prices can cause drift in the index.

The repricing method was not discussed in the HRPPI and yet it is the most widely used method in Europe. The reason is because the repricing method was not well known before it received a strong recommendation in an early version of a Eurostat report on the treatment of owner-occupied housing (OOH) in the Harmonized Index of Consumer Prices (HICP). This report has been through a number of drafts – for example version 4 was available online from 2015 (Eurostat, 2015 [20]). The chapter on house price indices in this and earlier drafts was written completely independently from the HRPPI. A later draft from 2017 (also available online) includes the RTD method and does not endorse the repricing method.

National Statistical Institutes (NSIs) at the time of the earlier drafts were under pressure from Eurostat to start computing official house price indices (if they were not already doing so). A number of NSIs turned to the OOH Manual rather than the HRPPI for guidance and hence decided to use the repricing method. As was noted above, the repricing method is fine as long as the reference hedonic model is updated every year (as it is in Italy and Luxembourg). However, some European countries using the repricing method have not been updating their reference hedonic model as often as maybe they should.

2.2 Controlling for Location

Constructing a price index is intrinsically about the temporal dimension of property prices. However, these prices also show complex spatio-temporal relationships (for a recent review see Teye and Ahelegbey(2017) [49] and their many references). The spatial dimension of this relationship is directly related to the physical location of properties. A property is an asset bundle composed of land and structure. Prices are determined by the interaction of the characteristics of these two assets. Location is a characteristic of the land component. The structure can be demolished and replaced, but the land will stay in the same physical location (see Section 3). In this section we concentrate our attention on how location is measured and incorporated when constructing property price indices.

It is evident from the extensive review in the HRPPI [19] and Hill (2013) [26], that location has been a key consideration in the price index literature for a long time. All methodologies used to construct price indices for residential housing have attempted to incorporate or control for location in some form. Indices using stratification or mixed adjustment approaches constructed by the Australian Bureau of Statistics (ABS,2006 [1]) use clustering of locations within cities. The standard repeat sales method compares pairs of sales of the same address which provides a micro location control. More recent extensions of the hybrid repeat-sales/hedonic model (Case and Quigley (1991) [7], Hill et al., (1997) [25]) have been proposed which include a nearest-neighbour estimator to control for location (Gunternmann et al., (2016) [23]).

Hedonic regression based methods can control for location in a number of ways. Location can be assumed to explain the behaviour of the mean of (log) price, the variance of (log) price or both[6].

Approaches that assume location explains the mean of (log) price add parametric or non-parametric terms to the regression function. These approaches are discussed in Section 2.2.1, and include post(zip)code dummy variables, distances to points of interest, a spatial price lag term, and non-parametric approaches. The spatial error model on the other hand controls for location as explaining the variance of (log) prices (Section 2.2.2). It is also possible to control for the effect of

[6] This discussion applies to Appraisal-Based Methods as they use a hedonic imputation model (see Chapter 7 of the HRPPI [19])

location on prices via both the mean and the variance of the hedonic model - presented in Section 2.2.3.

2.2.1 Controlling for location dependence in the hedonic log-price function

Including post(zip)code (neighbourhood) dummy intercepts This is the traditional control for location that has been used in the price index literature. It is a parametric approach which consists of including a set of intercept dummy shifts into the regression model. To control for location using post(zip)code or neighbourhood dummies, a term is added to the hedonic regressions. For instance, the term (22) is added to models such as to those presented in equations (1), (3), (4), (13), or (17),

$$\sum_{l=1}^{PC} \lambda_l d_{lh} \tag{22}$$

where l indexes the postcodes of the market area from where the sample is drawn, d_{lh} is one if property h is in postcode l, and λ_l is a parameter that provides the size of the regression function shift associated with that post(zip)code location, and PC is the number of postcodes in the study area. This approach has been used extensively, and has served to provide a base specification for comparing alternative approaches to controlling for location (Hill and Scholz, 2018 [29]; Diewert and Shimizu, (*forthcoming*) [18]; Hill et al., 2021a [28]).

Using parametric and non-parametric functions of the coordinates

Geographical information systems (GIS) software allows for the expression of property addresses using coordinates (latitude, longitude). Currently, standard software packages such as R can work with coordinates and functions of coordinates embedded in statistical models. These advances provide a number of options for modelling. Eucledian distances between a given property and a point of interest (POI) or its nearest neighbours can be easily computed and used in the estimation of the hedonic model. Next parametric and non-parametric alternatives are presented.

Parametric options:

a) Distances to points of interest In this case a set, L, of the hedonic characteristics in z_{ch} are regressors that control for location. This set is represented by the term (23) which is added to equations (1), (3), (4), (13) or (17). These regressors are either direct distances to major POI such us city centre, hospitals, shopping centres, schools, etc., or functions of these distances (e.g. inverse function).

$$\sum_{pi=1}^{L} \lambda_{pi} l_{pi,h} \tag{23}$$

Here $l_{pi,h}$ gives the distance in km(miles) to POI pi, and λ_{pi} is a parameter associated with the shadow price of that particular POI.

The use of these types of regressors is a common approach in the real estate and urban economics literature where the aim might be to estimate the willingness to pay associated with specific POIs (see for example the review in D'Acci (2013) [11]), in addition to using location to improve the prediction of (log) prices.

The price index literature has adopted the use of these type of regressors (Rambaldi and Fletcher (2014) [40]; Diewert and Shimizu, 2015 [16]).

b) Spatial Lag model The spatial lag model (SLM) is a type of spatially dynamic model. It is an autoregressive model but on the spatial instead of the time dimension. A time-dummy hedonic regression, (1), written in a spatially dynamic form is given by (24). A hedonic regression to construct hedonic imputated indices (such as (4)) would be given by a specfication such as that in (25),

$$\ln p_h = \rho_1 \sum_{i=1}^{N} w_{hi} \ln p_i + \sum_{c=1}^{C} \beta_c z_{ch} + \sum_{s=b+1}^{t} \delta_s d_{sh} + \varepsilon_h, \tag{24}$$

$$\ln p_{t,h} = \rho_1 \sum_{i=1}^{N} w_{hi} \ln p_i + \sum_{c=1}^{C} \beta_{t,c} z_{t,h,c} + \varepsilon_{t,h}, \tag{25}$$

where,

$|\rho_1| < 1$ is a spatial autoregressive parameter;

$w_{hh} = 0$, since unit h is not its own neighbour;

$\sum_{i=1}^{N} w_{hi} = 1$ for all i, the weighting matrix is row normalised; and

$\varepsilon_{t,h}$ is assumed mean zero and uncorrelated.

The weights w_{hi} are based on the geographic distance between property h and property i, d_{hi}. Then,

$$\begin{cases} w_{hi} = \dfrac{d_{hi}}{\sum_i d_{hi}} \text{ if } h \text{ and } i \text{ are neighbours} \\ w_{hi} = 0 \text{ if } h \text{ and } i \text{ are not neighbours} \end{cases} \tag{26}$$

There are some alternatives in how neighbours are defined. For example, "neighbours" can be defined either as a fixed number or as all of the first nearest-neighbours (see Chapter 2, Kelejian and Piras (2017) [32]).

To see how a **time-dummy index** can be computed from the SLM in (24), we use its reduced form in (27):

$$\ln p_h = \sum_{c=1}^{C} \beta_{c,w} z_{ch} + \sum_{s=b+1}^{t} \delta_{s,w} d_{sh} + \varepsilon_{h,w}, \tag{27}$$

where,

$\beta_{c,w} = \frac{\beta_c}{(1-\rho \sum_{i=1}^{N} w_{hi})}$

$\delta_{s,w} = \frac{\delta_s}{(1-\rho \sum_{i=1}^{N} w_{hi})}$

The price index for period t relative to the base period b is then calculated as follows:

$$\frac{P_t}{P_b} = \exp(\hat{\delta}_{t,w}) \tag{28}$$

where $\hat{\delta}_{t,w}$ denotes the estimate of $\delta_{t,w}$ in (27).

To compute a **hedonic imputed price index** for period t, as in (9), the SLM in (25) can be written in its reduced form , (29), and the required predictions, $\hat{p}_{t,h}(z_{t-1,h}), \hat{p}_{t-1,h}(z_{t-1,h}); \hat{p}_{t,h}(z_{t,h})$, and $\hat{p}_{t-1,h}(z_{t,h})$, obtained to compute the index.

$$\ln p_{t,h} = \sum_{c=1}^{C} \beta_{t,c,w} z_{t,h,c} + \varepsilon_{t,h,w}, \tag{29}$$

where,

$\beta_{t,c,w} = \frac{\beta_{t,c}}{(1-\rho \sum_{i=1}^{N} w_{hi})}$.

For further details on predictions from SLM see Chapter 4 of Kelejian and Piras (2017) [32].

Non-parametric options:

a) Splines and Spatial Coordinates

The use of spatial coordinates to model location effects non-parametrically in the price index literature has been adopted in Hill and Scholz (2018) [29], Diewert and Shimizu, (*forthcoming*) [18], and Hill et. al (2021a) [28].

Hill and Scholz (2018) [29] proposed to use the semi-parametric model (30) to obtain the predictions required to compute a hedonic imputed index of the form in (9),

$$\ln p_{t,h} = \sum_{c=1}^{C} \beta_{t,c} z_{t,h,c} + g(lat_h, long_h)_t + \varepsilon_{t,h}, \tag{30}$$

This model is estimated using a penalized least squares approach. Hill and Scholz (2018) [29] find the index computed from the predictions of this model at the quarterly frequency do not differ substantially from the index obtained using postcodes in place of the $g(.)$ function as in (22).

In a recent paper, Diewert and Shimizu, (*forthcoming*) [18] argue that the use of penalized least squares results in a smoothing method that fails the "smoothing invariance test" which implies the smooth series produced will change if a second round of smoothing is applied to the smoothed series originally obtained. They propose to use a modification of Colwell's (1998) [10] spatial interpolation method. The modification can be viewed as a general non-parametric method for estimating a function of two variables. Their paper is concerned with constructing indices of the value of land and not of property prices (see Section 3).

b) Geographically Weighted Regressions Geographically weighted regressions (GWR) are also in the family of modelling approaches that use functions of the latitude and longitude coordinates of the hth property, and they are non-parametric spatial models.

The GRW model is as follows,

$$\ln p_{t,h} = \sum_{c=1}^{C} \beta_{t,c,(lat_h,long_h)} z_{t,h,c} + \varepsilon_{t,h}, \tag{31}$$

The parameters $\beta_{t,c,(lat_h,long_h)}$ are estimated using a Gaussian spatial kernel and the geographical distribution of the estimates are based on the Euclidean distance between observations. As other kernel estimation techniques, there is a need to choose the bandwidth.

The observant reader would have noticed that the GRW model (31) is a non-parametric version of the SLM in (29). The use of a Gaussian kernel in this case would lead to the weight, w_{hi} associated with property i at location h being defined as:

$$w_{hi} = exp[-1/2(\frac{d_{hi}}{b})^2] \tag{32}$$

where, b is the bandwidth, and d_{hi} is defined as in (26). That is, the geographical distance between h and i.

Bidanset and Lombard (2014) [3] compare the SML and the GRW in the context of mass appraisals for tax assessments using as a comparison metric the coefficient of dispersion (COD). Both models can provide geographically disaggregated estimates. Both dominate a geographically unaware model; however, neither is found to be the dominant over the other uniformly.

Constructing **hedonic imputed price indices** from the predictions of (31) follows the standard procedure of producing the four predictions required for the computation of the price index in (9). This is the same procedure as that stated for the SLM in (29). To obtain a **time-dummy**

hedonic index would require a semi-parametric alternative where the hedonic regressors enter the regressions in a non-parametric form, while the time-dummy term is parametric. Bárcena et al. (2014) [2] used a geographically weighted regression to study the distribution of prices, but then proposed to construct a price index from a semi-parametric model where the hedonic characteristics enter parametrically and a cubic spline function of time is then normalised to compute a non-parametric version of a time-dummy price index for the whole geographical area under study.

2.2.2 Controlling for location dependence in the variance of the log-price function

In this case prices are assumed to be indirectly interrelated via spatially interrelated errors. This specification is then assuming that the covariance of prices is spatially dependent. A model for the computation of the time-dummy price index is given by (33), and one for the computation of hedonic imputed indices is given by (34),

$$\begin{aligned} \ln p_h &= \sum_{c=1}^{C} \beta_c z_{ch} + \sum_{s=b+1}^{t} \delta_s d_{sh} + u_h \\ u_h &= \rho_2 \sum_{i=1}^{N} w_{hi} \ln u_i + \varepsilon_h, \end{aligned} \tag{33}$$

$$\begin{aligned} \ln p_{t,h} &= \sum_{c=1}^{C} \beta_c z_{t,c,h} + u_{t,h} \\ u_{t,h} &= \rho_2 \sum_{i=1}^{N} w_{hi} \ln u_{t,i} + \varepsilon_{t,h}, \end{aligned} \tag{34}$$

Note that both models can be written alternatively as (35) and (36), respectively,

$$\ln p_h = \sum_{c=1}^{C} \beta_c z_{ch} + \sum_{s=b+1}^{t} \delta_s d_{sh} + (1 - \rho_2 \sum_{i=1}^{N} w_{hi})^{-1} \varepsilon_h \tag{35}$$

$$\ln p_h = \sum_{c=1}^{C} \beta_c z_{ch} + u_h + (1 - \rho_2 \sum_{i=1}^{N} w_{hi})^{-1} \varepsilon_h \tag{36}$$

which shows why the error term in the hedonic model is not uncorrelated, and thus from first principles it follows that while OLS is a consistent estimator of the parameters of the model, the OLS computed standard errors will be biased. This model can be estimated by maximum likelihood (details are provided in Kelejian and Piras (2017) [32]).

The computation of hedonic indices follows the standard approach, time-dummy indices from the estimated δ_s's and imputed indices from the predictions required for the computation of the price index in (9). The autoregressive spatial error structure of this model leads to at least two possible predictors, one is the standard predictor $\hat{p}_{t,h} = \exp\left[\sum_{c=1}^{C} \hat{\beta}_{t,c} z_{t,c,h}\right]$, the other is one that adds a correction due to the correlation induced from the spatial error lag in $u_{t,h}$ (the interested reader is directed to Chapter 4 of Kelejian and Piras (2017) [32]).

2.2.3 Controlling for location dependence in the mean and the variance of log-prices

A general parametric model can be specified which includes both a spatial lag in the prices as well as in the error. Model (37) shows the specification to compute time-dummy hedonic indices, while model (38) shows the specification to compute hedonic imputed type indices.

$$\begin{aligned} \ln p_h &= \rho_1 \sum_{i=1}^{N} w_{hi} \ln p_i + \sum_{c=1}^{C} \beta_c z_{ch} + \sum_{s=b+1}^{t} \delta_s d_{sh} + u_h \\ u_h &= \rho_2 \sum_{i=1}^{N} w_{hi} \ln u_i + \varepsilon_h, \end{aligned} \tag{37}$$

$$\ln p_{t,h} = \rho_1 \sum_{i=1}^{N} w_{hi} \ln p_i + \sum_{c=1}^{C} \beta_c z_{t,c,h} + u_{t,h}$$
$$u_{t,h} = \rho_2 \sum_{i=1}^{N} w_{hi} \ln u_{t,i} + \varepsilon_{t,h}, \tag{38}$$

The estimation of these models is by maximum likelihood (details are provided in Kelejian and Piras (2017) [32]).

Time-dummy indices can be computed from the estimated δ_s's and hedonic imputed price indices from the predictions following the standard procedure of producing the four predictions required for the computation of the price index in (9).

2.3 Empirical Feasibility

The previous subsections have provided a taxonomy of modelling approaches to compute both time-dummy and hedonic imputed price indices for residential housing that control for the dependence of prices on location. It is shown that all of the alternative specifications can be used to construct hedonic imputed property price indices. All alternatives are also easily implementable to construct time-dummy hedonic indices, except perhaps for the GWR model. Some authors have recently proposed to combine alternative models to improve price prediction (Oust et al., 2020 [36]), which as stated are inputs to hedonic imputed price indices. Importantly, there are packages in R, a toolbox in Matlab, and STATA routines that can estimate all or most of the above presented models making them feasible to practitioners everywhere.

3 Constructing Separate Price Indices for Land and Structures

Clapp (1980) [9] first proposed a model for the level of property values that allowed for the notion of dividing the property into additive land and structure values. Bostic et al., (2007) [5] proposed the concept of land leverage (the ratio of land value to overall property value) as an important indicator of residential property price dynamics, and followed the additive formulation of land and structure.

The conceptual model is (39)

$$V = L + S \tag{39}$$

where, V is the property value, L is the land value and S is the structure value.

The main issue faced by the modeller is that a standard hedonic regression cannot separate these two components. A log-linear specification cannot provide an additive decomposition. The regression must be linear. However, a standard linear regression with intercept (or time-dummies, or a time-varying intercept trend) and hedonic controls does not provide the required decomposition either. In this case two mixed - land and structure - components are obtained: 1) overall market condition and 2) a hedonic quality adjustment. Intercept time dummies, or a trend, capture the macroeconomic conditions of the property market under study (combining price trends in both the land and the structure). The reminder part of the regression provides a combined quality adjustment effect, where the individual estimates in the 'hedonic quality adjustment' component, are measures of the marginal effects of additional units of land size (inside margin), location, bedrooms, bathrooms, etc. The realisation that it is not possible to separate the value of the land from that of the structure using standard regression estimates, has led to a number of proposed alternatives that provide empirical identification strategies to separate the value of the land from that of the structure. Proposed approaches have included a non-linear systems approach, the use of exogenous information, and imposing of asymmetric behaviours on the dynamics of the land

and structure components.

Diewert(2007 [13], Section 5.1) proposed to combine an additive and a log-linear model to be estimated as a non-linear system, which would provide price indices per square metre of structure and land. Diewert et al., (2011) [14] explored a number of models and settled on a specification that used exogenous information to isolate the structure component, and thus providing identification of the land component. The model was then formalised in Diewert et al., (2015) [15] and labelled "the builder's model". The approach to separating the value of land from structure is based on replacing the set of parameters associated with the structure by an official price index of new building construction and a non-linear adjustment due to the depreciation of the asset with age. This framework has been applied in Diewert et al., (2015) [15] to data from the "Town of A" in the Netherlands, in Diewert and Shimizu (2015) [16] in an application to Tokyo residential property, and in Diewert et al., 2019 [17] in an application to British Columbia. By anchoring the model on an official price for new building construction, it is argued that the decomposition follows National Accounts principles and thus the estimates of land can be used in the computation of a country's productivity.

Rambaldi et al., (2010,2016) ([41],[42] proposed to approach the problem as the estimation of two unobserved components, where each component (land, structure) is uniquely mapped to a set of observable characteristics, and the behaviour of the components' prices is asymmetric. The underlying model is labelled "the valuer's model". The degree of asymmetry is determined by two bounded smoothing parameters which enter a modified Kalman filter algorithm. The land component is a function of land size and land location, and it is assumed to be the component that captures the largest proportion of price shocks in the market, an assumption that follows from earlier literature (Bostic et al., 2007, [5]). The structure component is a function of the structure's size (e.g. number of bedrooms, bathrooms, floor space, garages) and age[7], and its value is assumed to be more stable as its movements follow the trends in local markets' wages, construction's inputs prices and depreciation. The implementation is simple as the model depends on three parameters

[7]building quality, e.g. building materials types, can also be added to the controls for the structure. However, data on these are less likely to be available. Empirically, these do not seem to make a significant difference to the computed index. The key controls seem to be age and size of the structure.

that can be easily obtained. The first is the variance of the error term from a standard hedonic regression (obtained using least squares). The second is a pair of smoothing parameters which are bounded between zero and one and thus can be obtain by using a grid search. With estimates of these three parameters the algorithm to obtain the predictions of the value of the land and structure of each property, h, is just a set of formulae that does not require additional estimation. Rambaldi et al., (2016) [42] compared their estimated price indices for land and structure for the "Town of A" in the Netherlands to those obtained by Diewert et al., (2015) [15] to show they are not only comparable, but also smoother and can be computed at a monthly frequency even when the sample is small. Rambaldi and Tan (2019) [44] computed land price indices for three regions within the Greater Melbourne (Australia) metropolitan area, and compared the index's predicted growth in land prices to those computed by the state of Victoria's Valuer-General (VGV). To illustrate we draw from Rambaldi and Tan (2019)'s results.

Prior to 2019, revaluations from the VGV were run every two years and were part of the general valuation which also determines council rates. So, the valuation approach may have differed across Local Government Areas (LGAs) depending on the respective valuers' judgement. In early 2019 the VGV made available the revaluation outcomes for each LGA in the state of Victoria available on their website[8] since 2014[9]. These data contained the total site value (in $ amounts) for each LGA at a point in time. For example, the 2018 revaluation outcome determines the site value of properties as at 1 January 2018. These LGA site values were aggregated up to match the definition of inner, metro and outer regions of Greater Melbourne used by their model. From there, the biennial growth rate was calculated in line with that generated from the land value index (LVI[10]). Table 1 summarises the estimated revaluation outcome from the model's LVI, and compares to those from the VGV. They consider these results very encouraging given over this period the VGV only had oversight and there was a lack of standardisation. Table 2 compares the VGV's building cost index to that obtained from the model.

[8] https://www.propertyandlandtitles.vic.gov.au/valuation/council-valuations

[9] 2018, 2016 and 2014 revaluation rounds

[10] For example, an LGA in the inner area $(SiteValue\_inner\_2018/SiteValue\_inner\_2016 - 1) \times 100\%$ compared against $(LVI\_2018/LVI\_2016 - 1) \times 100\%$.

Table 1: Comparison of VGV land valuation versus model based Land Index

Region	Revaluation	Benchmark[1]	Model[2]	Difference
Inner				
	2016-2018	27.40%	29.30%	-1.90%
	2014-2016	30.54%	22.71%	7.83%
Metro				
	2016-2018	29.00%	35.60%	-6.60%
	2014-2016	33.81%	27.90%	5.91%
Outer				
	2016-2018	45.71%	36.48%	9.23%
	2014-2016	20.72%	17.86%	2.86%

[1] VGV valuations are at 1 January of corresponding year (2016, 2018).
[2] Increase over the periods: 2013Q4:2015Q4, 2015Q4:2017Q4

Source: Rambaldi and Tan (2019) [44]

Table 2: Comparison of VGV Building Cost Index and model based Structure Index

Period	VGV (residential construction)[1]		Model[2]		
July to June	Metropolitan	Regional	GM-Inner	GM-Metro	GM-Outer
2008-09	1.03	1.02	0.994	1.002	1.011
2009-10	1.03	1.03	1.028	1.042	1.031
2010-11	1.03	1.04	1.006	1.010	1.016
2011-12	1.03	1.05	0.993	0.994	0.998
2012-13	1	1.03	1.008	1.004	1.001
2013-14	1.02	1.04	1.018	1.021	1.011
2014-15	1	1	1.028	1.033	1.016
2015-16	1.03	1.01	1.022	1.025	1.022
2016-17	1.03	1.03	1.027	1.031	1.032
2017-18	1.035	1.035	1.003	1.014	1.026

[1] These are reported for Metropolitan and Non-metro/Regional Victoria. The Metropolitan does not overlap exactly with what is defined as Greater Melbourne.
Source: https://www.dtf.vic.gov.au/financial-reporting-policy/valuer-general-building-indices
[2] The model produces disaggregated figures for three areas within the Greater Melbourne area (GM).

Source: Rambaldi and Tan (2019) [44]

One of the motivations behind finding separate values for land and structure is to uncover the depreciation rate (impact of physical deterioration) of the stock of housing. The interested reader can consult Francke and Minne (2016) [22] and Diewert et al., (2017) [17] for a review of the literature and alternative approaches to the computation of the rate of depreciation of housing structures.

The price indices computed from the approach of Diewert and co-authors are of the time-dummy type as the index is based on a normalised set of time-period parameters that are estimated by the builder's model. The price indices computed from the approach proposed by Rambaldi and co-authors are of the hedonic imputation type. The model is used to compute the predictions of the price of land and structure for each sold property. Predictions of land prices are then used to compute formula (9), and similarly indices for the structure and the property (land + structure) can be obtained.

4 Extensions

4.1 Higher frequency indices

Traditionally property price indices have been computed at either the annual or quarterly frequency. Hedonic time-dummy based indices typically fitted annual dummies to the model that then determined the frequency of the resulting index. Hedonic imputed price indices are computed from regressions where all parameters (intercept and those attached to the hedonic characteristics) change at each time period (year, quarter, month, etc.) (see Section 2.1.3). The price index literature achieved this requirement by re-estimating the regression each year or quarter. Depending on the sample size, it is feasible to follow this approach to compute hedonic imputed price indices at a monthly frequency. However, samples are not random and thus the composition of sales within a given month can have a large impact on the estimated parameters and predictions, and thus unduly influence the resulting index. This issue was raised by Rambaldi and Fletcher (2014) [40], who proposed the use of time-varying parameter models to overcome the volatility induced by the composition of sales and varying sample sizes (due to issues such as seasonality of sales and periods of thin markets) when computing hedonic imputed price indices. Time-varying parameters build from the information from the previous and current periods producing a much smoother set of estimates and reducing the volatility of the imputations.

As more data are available, monthly price indices have become more common and until recently hailed as high frequency (see for example Bárcena et al., 2014 [2]; Bourassa and Hoesli, 2017 [6]). Bollerslev et al., (2016) [4] used an extended repeat-sales type model with data from ten major US cities to compute daily price indices. The model is estimated monthly, and then a moving-monthly window (i.e. it shifts the "month" by a day at a time) for the last month of the sample is used to produce a daily price index.

The first, to our knowledge, hedonic based high frequency index is that by Hill et al., (2021a) [28]. The model and index are computed at a weekly frequency using data for Sydney and a semi-parametric state-space model. Their model is a type of spatio-temporal specification, which have become popular in the real estate literature following the seminal work of Pace et al., (2000) [37] (see for example Liu, 2013 [33]; Hawkins and Habib, 2018 [24]; Chica-Olmo et al., (2019) [8];Otto and Schmid (2018) [36]; Teye and Ahelegbey (2017) [49]). Parametric spatio-temporal models have been used to compute monthly hedonic imputed price indices for property prices by Rambaldi and Fletcher (2014)[40] and for land prices by Rambaldi et al., (2016) [42] and Rambaldi and Tan (2019) [44].

An important finding of Hill et al., (2021a) [28] is that weekly indices are far more sensitive to the method of construction than those computed at a lower frequency such as quarterly. Hill and Scholz (2018) [29] and Diewert and Shimizu (*forthcoming*) [18] found hedonic imputed price indices obtained using postcode dummies to control for location in the hedonic model do not differ significantly from those obtained with models that use more sophisticated specifications, such as splines (see Section 2.2 for a presentation of alternative methods). Using the same metric as that proposed in Hill and Scholz (2018) [29] to compare indices - Index MSE(RS)- , Hill et al., (2021a) [28] find the indices obtained by a spatio-temporal model produces significantly and uniformly superior predictions of price relatives (i.e. the building blocks of a price index) to those obtained with Hill and Scholz (2018) [29] - GAM- and using postcode dummies in a time-varying parameter model (SS+PC) at monthly and weekly frequencies. This is shown in Tables 3 and 4, which have been reproduced from Hill et al., (2021a) [28] Their proposed spatio-temporal model is labelled SS+GAM.

Table 3: Model Prediction and Index Quality Comparison

	Model RMSPE				Index MSE(RS)	
	Sydney	harbour	Bondi beach	Blue Mountains	Weekly	Monthly
Radius		5 Km	2.5 Km	30Km		
GAM	0.1857	0.3136	0.3008	0.1260	0.0233	0.0245
SS+GAM	0.1775	0.3067	0.2954	0.1315	0.0102	0.0112
SS+PC	0.2088	0.3518	0.3239	0.1540	0.0246	0.0264
Sample	433202	13222	6950	19089		

Note: The mean square prediction error of prices (RMSPE) are uniformly higher for the model with postcodes across all geographical alternatives. Similarly, the mean square error of the prediction of price relatives (MSE(RS)) are higher at both the SS+PC at both weekly and monthly frequency. The RMSPE is lowest for the SS+GAM model except in one case (the Blue Mountains) when GAM is the lowest. The SS+GAM is uniformly the lowest in MSE(RS) for both weekly and monthly frequencies.
- Reproduced from Hill et al., (2021a) [28]- Table 3

Table 4: p-values for H_0: $MSE(RS)_{M1} - MSE(RS)_{M2} = 0$

	Weekly	Monthly
SS+PC vs. SS+GAM	0.0000	0.0000
SS+PC vs. GAM	0.0483	0.0014
GAM vs. SS+GAM	0.0000	0.0000

Note: These p-values imply that SS+GAM is highly significantly different from both SS+PC and GAM at both the weekly and monthly frequencies.
- Reproduced from Hill et al., (2021a) [28] - Table 4

4.2 Measuring price changes for the stock of housing

The combination of data availability, mass-imputation techniques and high computing power provides an ideal environment to consider constructing price changes for the stock of housing. This issue was mentioned in the HRPPI ([19]), chapters four and eight. The HRPPI ([19]) indicated that the use of stratification can approximate a stock based residential property price index. Diewert et al., (2017) [17] propose to use sales data over a reasonably long period of time to approximate the quantity (stock) of residential property. The construction of the relevant "stock" of housing is the key issue. The availability of administrative data would seem to be a promising path. Administrative land titles' data can provide the population of property by use (e.g. residential detach, attached, etc.). However, these would need to be linked to other datasets that capture renovations and improvements to provide a reasonable approximation of the stock at each point in time. Once a stock dataset of characteristics at the level of individual properties has been constructed, a hedonic model can be used to estimate prices for these properties each period. A hedonic imputation price index for the housing stock can then be computed. This is an area likely to see more research in the near future.

5 Conclusion

In recent years the rolling-time dummy (RTD) and repricing methods have become popular with National Statistical Institutes (NSIs) for constructing official house price indices. Indeed most NSIs in Europe use one of these two methods. One reason for this is that both methods are better suited for use with smaller datasets than the average characteristics or hedonic imputation methods.

Location is typically controlled for in hedonic models using postcode/zipcode dummy variables. However, a number of more sophisticated methods are now available, particularly given the increasing availability of geo-coded longitudes and latitudes at the level of individual properties. While such methods are generally not currently being used by NSIs in their official indices, this could change in the future.

Another active area of research is the use of hedonic methods to construct separate price indices for land and structures. A key concern here is that house price indices may be upwardly biased if they fail to account for depreciation of the structures. Separating land from structure ensures that the resulting land price index is not distorted by depreciation.

There is growing demand for higher frequency (e.g., weekly) indices. In some cases the housing datasets may not be large enough to easily accommodate say weekly indices. Recently a number of approaches have been developed to allow more robust house price indices to be computed at higher frequencies and/or on smaller datasets.

Finally progress is also being made on the construction of price indices for the stock of housing. The hedonic imputation method is ideal for this purpose as long as characteristic information is available on a sufficiently large portion of the housing stock.

In conclusion, the application of hedonic methods to the construction of house price indices is an active research area in which significant progress has been made in the last few years. This is helping to improve the accuracy of house price indices and broadening the range of indices that can be computed.

References

[1] Australian Bureau of Statistics (2006), A Guide to the House Price Indexes, Cat 6464.0

[2] Bárcena, M.J., P. Menéndez, M.B. Palacios, F. Tusell (2014), A Real-Time Property Value Index Based on Web Data, Data Mining Applications with R, Chapter 10. Elsevier.

[3] Bidanset, Paul E. and Lombard, John R. (2014), Evaluating Spatial Model Accuracy in Mass Real Estate Appraisal: A Comparison of Geographically Weighted Regression and the Spatial Lag Model. School of Public Service Faculty Publications. 26, https://digitalcommons.odu.edu/publicservice_pubs/26.

[4] Bollerslev, T. A. J. Patton and W. Wang (2016), Daily House Price Indices: Construction, Modeling, and Longer-run Predictions, Journal of Applied Econometrics, 31, 1005–1025, doi: 10.1002/jae.2471.

[5] Bostic, R.W., S.D. Longhofer and C.L. Readfearn (2007), Land Leverage: Decomposing Home Price Dynamics, Real Estate Economics 35,2, 183-2008.

[6] Bourassa, S. C. and Martin Hoesli (2017) High-Frequency House Price Indexes with Scarce Data, Journal of Real Estate Literature, Vol. 25, No. 1 (2017), pp. 207-220

[7] Case, K.E. and J.M. Quigley (1991), The Dynamics of Real Estate Prices, The Review of Economics and Statistics, 1991, 73(1), 50–8.

[8] Chica-Olmo, J., R. Cano-Guervos and M. Chica-Rivas (2019), Estimation of Housing Price Variations Using Spatio-Temporal Data, Sustainability, 11, 1551; doi:10.3390/su11061551.

[9] Clapp, J.M. (1980), The Elasticity of Substitution for Land: The Effects of Measurement Errors, Journal of Urban Economics, 8, 255-263.

[10] Colwell, P. F. (1998), A Primer on Piecewise Parabolic Multiple Regression Analysis via Estimations of Chicago CBD Land Prices, Journal of Real Estate Finance and Economics, 17,1, 87-97.

[11] D'Acci, L. (2013), Monetary, Subjective and Quantitative Approaches to Assess Urban Quality of Life and Pleasantness in Cities (Hedonic Price, Willingness-to-Pay, Positional Value, Life Satisfaction, Isobenefit Lines), Social Indicators Research, 115, 531-559, doi:10.1007/s11205-012-0221-7

[12] de Haan J. (2004), Direct and Indirect Time Dummy Approaches to Hedonic Price Measurement, Journal of Economic and Social Measurement, 29(4), 427-443.

[13] Diewert, W.E. (2007), The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions, Discussion Paper 07-01, Department of Economics, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z1.

[14] Diewert, W.E., J. de Haan and R. Hendriks (2011), The Decomposition of a House Price Index into Land and Structures Components: A Hedonic Regression Approach, The Valuation Journal 6, 58-106.

[15] Diewert, W.E., J. de Haan and R. Hendriks (2015), Hedonic Regressions and the Decomposition of a House Price index into Land and Structure Components, Econometric Reviews 34, 106-126.

[16] Diewert and C. Shimizu, 2015, Residential Property Price Indexes for Tokyo Macroeconomic Dynamics 1, 8, 1-56, doi: 10.1017/S1365100514000042

[17] Diewert, W.E., N. Huang, and K. Burnett-Isaacs (2017), Alternative Approaches for Resale Housing Price Indexes, Discussion Paper 17-05, Vancouver School of Economics, The University of British Columbia, Vancouver, Canada, V6T 1L4.

[18] Diewert, W.E., and C. Shimizu, (*forthcoming*), Residential Property Price Indexes, Spatial Coordinates versus Neighbourhood Dummy Variables, Review of Income and Wealth.

[19] European Commission, Eurostat, OECD, and World Bank (2013) Handbook on Residential Property Price indices (RPPIs). Luxembourg: Publications Office of the European Union, doi:10.2785/34007

[20] Eurostat (2015), Detailed Technical Manual on Owner-Occupied Housing for Harmonised Index of Consumer Prices (DF V4). Eurostat: Luxembourg.

[21] Francke, M.K. (2010),Repeat Sales Index for Thin Markets: A Structural Time Series Approach, Journal of Real Estate Finance and Economics, 41, 1, 24–52.

[22] Francke, M.K. and A. M. van de Minne (2016), Land, Structure and Depreciation, Real Estate Economics, 0,0, 1-37, doi:10.1111/1540-6229.12146.

[23] Guntermann, K. L., C. Liu and A. D. Nowak (2016), Price Indexes for Short Horizons, Thin Markets or Smaller Cities, The Journal of Real Estate Research, Vol. 38(1), January–March, 93- 128.

[24] Hawkings, J. and K.N. Habib (2018), Spatio-Temporal Hedonic Price Model to Investigate the Dynamics of Housing Prices in Contexts of Urban Form and Transportation Services in Toronto, Transportation Research Record, 2672, 6, 21-30, doi:10.1177/0361198118774153.

[25] Hill, R.C., J.R. Knight, and C.F. Sirmans (1991), Estimating Capital Asset Price Indexes, Review of Economics and Statistics, 1997, 79(2), 226–33.

[26] Hill, R.J. (2013), Hedonic Price Indexes for Residential Housing: A Survey, Evaluation and Taxonomy, Journal of Economic Surveys, 27, 5, 879–914. doi: 10.1111/j.1467-6419.2012.00731.x

[27] Hill, R. J. and D. Melser (2008), Hedonic Imputation and the Price Index Problem: An Application to Housing, Economic Inquiry, 46,4, October, 593-609. doi:10.1111/j.1465-7295.2007.00110.x.

[28] Hill, R. J., A. Rambaldi and M. Scholz (2021), Higher Frequency Hedonic House Price Indices: A State-Space Approach, Empirical Economics, 61, 417–441, doi:10.1007/s00181-020-01862-y.

[29] Hill, R. J. and M. Scholz (2018), Can Geospatial Data Improve House Price Indexes? A Hedonic Imputation Approach With Splines, Review of Income and Wealth, 64, 4, 737-756 doi:10.1111/roiw.12303.

[30] Hill, R. J., M. Scholz, , C. Shimizu and M. Steurer (2018). An evaluation of the methods used by European countries to compute their official house price indices. Economie et Statistique / Economics and Statistics, 500-501-502, 221–238, doi:10.24187/ecostat.2018.500t.1953.

[31] Hill, R. J., M. Scholz, C. Shimizu and M. Steurer (2021), Rolling-Time-Dummy House Price Indexes: Window Length, Linking and Options for Dealing with Low Transaction. Journal of Official Statistics, forthcoming.

[32] Kelejian, H. and G. Piras (2017), Spatial Econometrics, Academic Press, Elsevier, ISBN: 978-0-12-813387-3.

[33] Liu, X. (2013), Spatial and Temporal Dependence in House Price Prediction, Journal of Real Estate Finance and Economics, 47, 341-369, doi:10.1007/s11146-011-9359-3.

[34] O'Hanlon, N. (2011), Constructing a National House Price Index for Ireland, Journal of the Statistical and Social Inquiry Society of Ireland, 40, 167-196.

[35] Otto, P. and W. Schmid (2018), Spatiotemporal analysis of German real-estate prices, Annals of Regional Science, 60, 41–72, doi:10.1007/s00168-016-0789-y

[36] Oust, A., S.N. Hansen and T. R. Pettrem (2020), Combining Property Price Predictions from Repeat Sales and Spatially Enhanced Hedonic Regressions, Journal of Real Estate Finance and Economics, 61, 183-207, doi:10.1007/s11146-019-09723-x.

[37] Pace, R. K., Ro. Barry, O. W. Gilley, C.F. Sirmans (2000) International Journal of Forecasting, 16 (2000) 229–246, doi:10.1016/S0169-2070(99)00047-3.

[38] Pakes A. (2003), A Reconsideration of Hedonic Price Indices with an Application to PC's, American Economic Review, 93(5), 1578-1596.

[39] Rachman, A. N. (2019), "An Alternative Hedonic Residential Property Price Index for Indonesia Using Big Data: The Case of Jakarta," Paper presented at the International Conference on Real Estate Statistics, February 20-22, 2019 Eurostat, Luxembourg. https://www.crrem.eu/crrem-at-the-eurostat-international-conference-on-real-estate-statistics-2019/

[40] Rambaldi, A.N. and C. S. Fletcher (2014), Hedonic Imputed Property Price Indexes: The Effects of Econometric Modeling Choices, Review of Income and Wealth, 60, S423-S448, doi:10.1111/roiw.12143

[41] Rambaldi, A.N., R.R.J McAllister, K. Collins and C.S. Fletcher (2010), Separating Land from Structure in Property Prices: A Case Study from Brisbane Australia, School of Economics, The University of Queensland, St. Lucia, Queensland 4072, Australia.

[42] Rambaldi, A.N., R.R.J McAllister and C.S. Fletcher (2016), Decoupling land values in residential property prices: smoothing methods for hedonic imputed price indices, 34th IARIW General Conference. Dresden, Germany, August 2016. http://old.iariw.org/dresden/rambaldi.pdf.

[43] Rambaldi, A. N. and D. S. P. Rao (2013), Econometric Modeling and Estimation of Theoretically Consistent Housing Price Indexes, CEPA Working Papers Series WP042013, School of Economics, The University of Queensland, Australia.

[44] Rambaldi, A. N., Tan, M. S. (2019) Land Value Indices and The Land Leverage Hypothesis in Residential Housing. International Conference on Real Estate Statistics. EuroStat. Luxembourg. https://ec.europa.eu/eurostat/about/opportunities/conferences

[45] Real Estate and Construction Economy Bureau (2020), Methodology of JRPPI: Japan Residential Property Price Index. Ministry of Land, Infrastructure, Transport and Tourism, Japan. https://www.mlit.go.jp/common/001360414.pdf

[46] Shimizu, C. and W. E. Diewert (2019), Residential Property Price Index in Japan: Discussion in Methodology and Data Sources, Paper presented at the International Conference on Real Estate Statistics, Eurostat, Luxembourg, 2019. https://www.crrem.eu/crrem-at-the-eurostat-international-conference-on-real-estate-statistics-2019/

[47] Shimizu, C, H.Takatsuji, H.Ono and K. G. Nishimura (2010), Structural and Temporal Changes in the Housing Market and Hedonic Housing Price Indices, International Journal of Housing Markets and Analysis, 3(4), 351-368.

[48] Silver M. and S. Heravi (2001), Quality Adjustment, Sample Rotation and CPI Practice: An Experiment, Presented at the Sixth Meeting of the International Working Group on Price Indices, Canberra, Australia, April 2-6.

[49] Teye, A. L. and Ahelegbey, D. F. (2017) Detecting spatial and temporal house price diffusion in the Netherlands: A Bayesian network approach. Regional Science and Urban Economics, doi:10.1016/j.regsciurbeco.2017.04.005

[50] Triplett J. E. (2004) Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products. STI Working Paper 2004/9, Directorate for Science, Technology and Industry, Organisation for Economic Co-operation and Development, Paris. https://www.oecd-ilibrary.org/science-and-technology/handbook-on-hedonic-indexes-and-quality-adjustments-in-price-indexes_643587187107