Research Paper

Exploring Hedonic Methods for Constructing a House Price Index

New
Issue

Research Paper

Exploring Hedonic Methods for Constructing a House Price Index

Lujuan Chen, Shiji Zhao, Paul Romanis
and Poh Ping Lim

Analytical Services Branch

INQUIRIES

CONTENTS

ABBREVIATIONS

The following abbreviations have been used in this paper.

ABS   Australian Bureau of Statistics

CBD   Central Business District

DPIWE   Department of Primary Industries, Water and Environment

HPI   House Price Index

HPIm   Hedonic Price Imputation

HPIm–I   Hedonic Price Imputation indexes

MAC   Methodology Advisory Committee

SEIFA   Socio-economic Indexes for Areas

TD   Time Dummy

TD–I   Time Dummy indexes

EXPLORING HEDONIC METHODS FOR CONSTRUCTING A HOUSE PRICE INDEX

Lujuan Chen, Shiji Zhao, Paul Romanis and Poh Ping Lim
Analytical Services Branch

ABSTRACT

The construction of House Price Indexes (HPI) for Australian capital cities poses particular difficulties. Because the same houses are rarely sold in successive quarters, conventional price methods based on matched samples cannot be applied. Furthermore houses are heterogeneous goods in terms of location and other characteristics. Quarterly changes in price indexes could therefore be driven by compositional change rather than underlying price movements.

Hedonic methods involve application of regression techniques to express the price of a home in terms of its characteristics (such as age, size, construction material, location, neighbourhood etc.) and then constructing an index from the hedonic function. This research indicated that plausible hedonic models could be fitted for two cities.

Readers should note that since the completion of this research, ABS has revised the methods for compiling the house price index as described in ABS cat. no. 6417.0. While ABS has not adopted hedonic methods due to the lack of data available across the capital cities, they nevertheless remain of interest in ongoing research and have been used indirectly to inform the development of the new methods.

1. INTRODUCTION

The Australian Bureau of Statistics (ABS) compiles quarterly house price indexes for eight major Australian capital cities and an index at the national aggregate level. They are published in *House Price Indexes* (ABS cat. no. 6416.0). For many years, these indexes have been used by the Reserve Bank of Australia and the Department of the Treasury in their formulation of monetary and fiscal policies.

The surge of house prices starting from late 1990s and the recent slowing have engaged close attention of the policy agencies (including Treasury and the Reserve Bank) and media commentators on the House Price Index (HPI) and its ability to indicate the timing and magnitude of changes in house prices.

In late 2003, the ABS undertook a study to investigate the feasibility of applying hedonic methods to the construction of the HPI. The analysis was based on data from Hobart and Adelaide only, because the information required by hedonic methods were not available in other cities.

This paper was prepared for the ABS Methodology Advisory Committee (MAC) members to discuss methodological issues. It was presented to a MAC meeting in November 2004. In this context, readers should note that the discussion of the data and methodological issues reflects the situation and thinking in 2004. Also the views expressed in the paper are those of the authors and not an ABS position.

In 2005, the ABS revised the method of compiling HPI and the current HPI is compiled based on a method of geographic stratification. The new HPI did not use the hedonic method because the data required for this method (e.g. data about housing characteristics) are not available in most of the eight capital cities. Interested readers may refer to the information paper (Australian Bureau of Statistics, 2005b) published by the ABS about the methods used in the compilation of the new HPI.

This paper focuses on application of hedonic methods to the construction of house price index. However, this method has been applied by official statistical agencies to other areas. For example, the ABS used this method in the construction of the price index of personal computers and interested readers may refer to Australian Bureau of Statistics (2005a) for more information about the method used in the construction of that index.

In this paper, we seek comments and suggestions from MAC members to guide our future research of applying hedonic methods to house price indexes. Hedonic methods may be directly used in the construction of the HPI or indirectly to inform statistical choices regarding geographic stratification, design of weighting patterns and outlier treatment, etc.. We will use two case studies based on data from Hobart and Adelaide as a platform for asking questions. In the paper, we presented some early

findings for the purpose of exposing areas where we need advice and help and, at this stage, they are not conclusive.

Section 2 outlines the data and methodology used in the construction of the current HPI. Section 3 gives a brief overview of hedonic method and its application to price index. In Sections 4, we briefly describe the data from Hobart and Adelaide and present some early findings from applying hedonic methods to the data. Section 5 summarises questions for MAC members.

2. DATA AND METHODS FOR HPI COMPILATION IN 2004

The *House Price Indexes* (ABS cat. no. 6416.0) are published by the ABS on a quarterly basis. They cover each of the eight capital cities – Sydney, Melbourne, Brisbane, Perth, Adelaide, Hobart, Canberra and Darwin – other regions are not included. A weighted average of the eight capital cities price indexes is also published.

2.1 Data

In 2004, the HPI was based solely on data collected from the Offices of the Valuers-General or the State Real Estate Institutes. The data represent *transaction values* or the prices of houses being *sold* in the reference periods. Therefore, the HPI is about the values of the houses on the market and it does not necessarily reflect the 'values' of the housing stock. For some cities, the data are contract prices recorded at the date of exchange of contracts and, for other cities, they are recorded at the date of final settlement.

Included in the HPI are exclusively detached houses. Townhouses and apartments are not included in the current price index. This has two implications. First, the prices reflect not only the values of houses *per se* but also the values of the land. Second, the HPI may have a better representation of the market for owner occupied houses than the market for investment, particularly in the cities where there is a high concentration of investment properties.

An early investigation of the data revealed four prominent features of the sample of houses sold. First, the sample sizes varied enormously between cities and over time, depending on the sizes of the cities and the conditions of the housing market when the data were collected. For example, 433 prices were recorded in June 2002 and 1,128 in March 2003 for Hobart. For Sydney, 8,866 records were included in the sample of the June Quarter 2003.

Second, there is almost no overlap in the samples between adjacent quarters, obviously because very few houses are sold in two consecutive quarters. This feature made it impossible for the ABS to apply conventional price methods (i.e. the so-called 'matched sample' methods) to the HPI.

Third, houses are very heterogeneous goods and their prices are determined by a wide range of factors: location, neighbourhood (i.e. natural and social) environment, convenience, size, age, quality of construction materials and many others. Without being able to collect and compare the prices of 'identical' houses, the price statisticians risk inconsistency of resulting indexes. In other words, the price indexes – rather than representing the underlying price movements – could be driven by changes in the composition of the 'quality' or 'characteristics' of houses on the

market. *The difficulty in controlling the 'quality' or 'characteristics' was the very reason ABS decided to explore hedonic methods.*

Fourth, the range of available information on house characteristics differed considerably among the eight cities. For example, the data from Sydney only included three variables: prices, addresses of the houses and the date when the houses were sold. However, the data from Hobart and Adelaide included prices, addresses, date and a good variety of characteristics of the houses. The limited information made it impossible for us to explore hedonic methods for cities beyond Hobart and Adelaide.

In addition, the ABS often had difficulty in receiving data on time. The HPI is scheduled to be published 9–10 weeks after the reference period. However, some data suppliers were unable to deliver all the data by the time when ABS finalised the preparation for publication. This is a significant problem that ABS is trying to solve.

2.2 The method

Mainly due to the data limitations, ABS has adopted a 'flexible' approach and tailored methods to suit the varying conditions of the data from the eight cities. To HPI construction, this means similar but not identical methods are applied across the cities and, to ensure the quality of resulting indexes, certain adjustments have been made (to the apparent outliers), where appropriate and necessary.

Indexation methods

Depending on the sizes of the city, the HPI samples are stratified into several geographical 'regions'. Price indexes are calculated for each of the regions. Then the regional indexes[1] are weighted up to the city level (to form a city price index). A national house price index is calculated by aggregating the prices indexes for the eight cities.

For Sydney, Melbourne, Brisbane and Adelaide, a 'tri-mean' method has been used in the calculation of the regional price indexes. Broadly speaking, this method involves following four steps:

1. In each region, 'extreme' (or 'unrepresentative') values are removed from the sample;

2. Prices are divided into three quantiles, representing low, medium and high valued houses;

3. An unweighted average price is calculated for each quantile;

---

1 In this paper, the term 'region' is used to describe a particular level of geographic stratification that was applied to the compilation of the former HPI. Hence 'regional' index and 'regional' weights are terms used to describe the index and weights at that level.

4.   A weighted arithmetic mean is used to aggregate the three 'quantile' prices to form a regional price index. The price for the medium quantile is weighted by 0.5 and the prices for the low and high quantiles are given 0.25 each;

A simpler method is used in the calculation of regional price indexes for Perth, Hobart, Darwin and Canberra. For these cities, after removing the 'extreme' (or 'unrepresentative') values, an unweighted average price is calculated for each region.

Stratification

As mentioned earlier, each city is stratified into several geographic regions. There are seven regions for Sydney, Brisbane and Perth, six regions for Melbourne, five regions for Adelaide, four regions for Hobart and three regions for Darwin and Canberra. For example, Sydney is stratified into Central, North, South, West, Burwood, Campbelltown, and Penrith and Canberra is divided into North, South, and Central.

The geographic stratification is very important to form a consistent index at city level.

> We are interested in any experience that MAC members may wish to share with us about design of geographic stratification for the housing market (e.g. investors/ home owners).

'Regional' and city weights

Weights are important components in calculating price indexes at city level and in the construction of the national HPI.

The regional weights used to derive capital city indexes were estimated based on the information extracted from the Population Census. The estimation was based on the distribution of the housing stock among the regions and it also took into account certain characteristics of the houses, in order to make sure that the weights are representative of the houses within the relevant region.

The indexes for each of the eight capital cities are arithmetically weighted together to form a national housing price index. The weights were estimated based on the value of *secured (individual) finance commitments* for the purchases of newly erected and established houses. The data were sourced from the ABS Survey of Housing Finance for Owner Occupation.

Treatment of outliers

To ensure the resulting price indexes are reasonably representative, it is important to properly deal with the prices which are suspiciously low or unreasonably high. The latter could represent houses which are too 'unique', such as the properties carrying certain non-replaceable features (e.g. neighbouring the Prime Minister's Lodge) or

unique historic values. Unfortunately the data available to ABS do not carry relevant information and, as a result, we have to rely on the price information to form a judgement and make adjustments accordingly.

A commonly used method to deal with such 'unrepresentative' prices is to exclude extreme values (or outliers) from the samples before they are used in the index calculation. The outliers may be identified through comparing the extreme values with the distribution of the observations in the whole sample or based on the experience and knowledge of local markets. An alternative method is to set upper and lower limits and to remove the observations outside the boundaries.

For Sydney, Melbourne and Brisbane, unusually high or low prices (i.e. generally below $100,000) are excluded, if they are considered not typical for the region. In determining whether the observations qualify as an 'outlier', comparisons are also made between the distributions of the prices in the current and the previous quarters. For Perth, the movements of the prices in the previous quarter are used as the reference in determining the upper and lower limits for the current quarter and prices that move out of the boundaries are removed.

For Darwin, $70,000 is set as the lower limit and no upper limit is applied. No fixed limits are set for Adelaide, Hobart and Canberra.

Although it may be argued that methods of identifying outliers and the subsequent treatments involve a certain degree of 'subjectivity', as we will show in the following sections, they appear to have worked fairly well for the current data. However, it is unrealistic to assume that the current methods will continue to work in the future, particularly after ABS uses different data from alternative sources.

Therefore we seek insights from MAC members about how to improve ABS methods of identifying and treating outliers that are practical, defensible and more 'scientific'.

3. AN OVERVIEW OF HEDONIC PRICE INDEXES

An important consideration in the constructing a credible HPI is being able to control for the differences in the 'quality', so that the houses sampled from different periods are comparable. It is reasonable to argue that the current methods would have effectively controlled certain aspects of quality through properly designed stratification, weighting structure and outlier treatments. However, it is not clear how well the method has controlled for other factors that are expected to have influenced housing prices. A prominent example is the varying attributes of individual houses (e.g. construction of materials, age and size, etc.). The main motivation of our study is to experiment and test the efficacy of hedonic methods to account for the impact of housing attributes on the HPI.

Hedonic method is a regression-based technique and, broadly speaking, it involves two steps. The first step is to estimate an equation (i.e. hedonic function) where market prices are used as a dependent variable and independent variables include a set of housing 'characteristics'. In order to ensure the consistency of the resulting index, the characteristic variables should cover the most important qualities that are valued by the market. The impact of changes in the composition of housing characteristics is minimised through the hedonic functions.

The second step is to use the coefficients of the hedonic function to construct an index. There are several methods that may be applied to the indexation.

3.1 Hedonic function

In the literature on hedonic price indexes for housing market (Goodman, 1989; Williams, 1991), housing characteristics are summarised into three broad types of variables:

- Structural attributes. These refer to 'qualities'[2] of individual houses such as size of houses or land, number of rooms, construction materials and age etc.. In this paper, we will use a capital letter "S" to represent structural attributes.

- Locational attributes are sometimes measured in terms of the distances to the Central Business District (CBD), local shopping centres, hospital and schools or closeness to popular places such as beaches etc.; In this paper, we will use a capital letter "L" to represent locational attributes.

- Neighbourhood attributes. They often refer to general social and natural conditions. In this paper we will use a capital letter "N" to represent neighbourhood attributes.

2　In this paper, 'attributes', 'characteristics' and 'qualities' are used interchangeably and they all represent factors that influence the prices of individual houses.

The following equation is a general hedonic function:

P= f(L, S, N)   (1)

where capital "P" represents prices of individual houses.

The partial derivative of the above hedonic function with respect to any attributes may be interpreted as the marginal price of the attribute, ceteris paribus (Rosen, 1974). If equation (1) perfectly describes the relationships between prices and the attributes and all the attributes (that have an influence on prices) are included in the estimation, then we will be able to predict accurately the prices of individual houses based only on the information on housing attributes.

In other words, if we have had a reliable hedonic function and obtained all the information about housing attributes, we will no longer need to rely on actual prices of identical or 'comparable' houses to construct a price index. The information about the housing attributes will enable us to construct a consistent index.

The first step of developing a hedonic price index is to define, estimate and test a hedonic function and its results may be applied *directly* or *indirectly* to price indexation.

3.2 Hedonic price indexes – A direct approach

Using the hedonic function directly in a price index is an approach favoured by academics and it has become increasing popular among official statistical agencies. Broadly speaking, price indexes may be constructed using two classes of methods – *Hedonic Price Imputation* method (HPIm) [3] and *Time Dummy* method (TD) which result in what we have called *Hedonic Price Imputation Indexes* (HPIm–I) and *Time Dummy Indexes* (TD–I).

Hedonic Price Imputation method (HPIm)

The following equation is an example of a specific form of a hedonic function.

$$Ln P_i = \beta_0 + \sum_{k=1}^{K} \beta_k Ln X_{ik} + \sum_{j=1}^{l} a_j C_{ij} + \varepsilon_i$$ (2)

In this function, housing characteristics are represented by a set of continuous variables (i.e. X_{ik} where $i = 1, 2, ..., I$ houses and $k=1, 2, ..., K$ variables) and a set of categorical variables (i.e. C_{ij} where $j = 1, 2, ..., J$ variables). β_k and a_j are coefficients. In equation (2), prices and continuous variables are expressed in log form where *Ln*

---

3   The term "Hedonic Price Imputation method" was used by Triplett (2001). For the purpose of constancy, we
    have used this term in our paper to describe this method.

denotes the natural logarithm and ε_i is assumed to be independently and identically distributed with expected value of 0 and variance of σ^2.

If data are available for more than one period, then several options are available for index construction. For example

- *Option 1(a)*: Equation (2) is estimated for each of the periods and the *predicted values* of P are used directly in price indexation. [4]

  This option relies on the data to determine whether specific variables should be included or dropped from the function and, as a result, the hedonic functions for different time periods may contain different variables in both X and C.

- *Option 1(b)*: This option differs from Option 1(a) only in that the variables in the hedonic function are predetermined and they are used in the estimation from the data in all the time periods.

- *Option 2*: Equation (2) is estimated from a set of data that covers current and (one or more) historical periods. [5] The coefficients are used to calculate the *predicted prices* for the base and current periods, which are used to construct an index.

Once we have obtained the predicted prices for both the base and current periods, the price index may be constructed based on various index formulae. Appendix 1 provides some of the most popular bilateral index formulae.

MAC members may suggest other options and here we will not go any further. However, it is noted that each of these options has its strengths and limitations. For example, although Option 1(a) is reasonably straightforward to implement, it has some complications. For example, coefficients available for index calculation could differ between base and current periods. If the variation is a reflection of reality (i.e. consumer preference), the index will be consistent. However, if the sample is not sufficiently large, the presence and absence of coefficients may be influenced by the variations of the statistical properties of the samples used in the estimation. In this circumstance, it is not clear to us whether the index will be consistent or not.

---

4   In the literature of hedonic methods, some authors argued that indexes should be constructed exclusively based on the predicted value of P and other authors prefer to maximise the use of actual prices. In the former case, predicted prices (of 'identical' houses) for both periods are used in the indexation and, in the latter case, both actual and predicted prices are used.

5   Such data are sometimes called 'pooled data'.

We are interested in any advice or insights from MAC members on

- the choice between the Options, particularly between 1(a) and 1(b);

- circumstances under which the 'statistically insignificant' variables should or should not be included in the index calculation;

- circumstances where 'predicted' prices should or should not be used in the index calculation.

Time Dummy method (TD)

The hedonic function estimation under the TD method always covers more than one period and it looks very similar to that used under the HPIm method. The following equation is an example.

$$Ln\, p_i^t = \beta_0 + \delta d_i^t + \sum_{k=1}^{n} \beta_k\, Ln\, x_{ik}^t + \sum_{j=1}^{l} a_j\, c_{ij}^t + \varepsilon_i \qquad (3)$$

This equation differs from equation (2) only in that it included an additional set of dummy variables (d_i^t). If only two (i.e. base and current) periods are involved (i.e. $t=0, 1$), then d_i^t takes a value of 1 if it is for the current period and 0 for the base period.[6]

The price index can be constructed in a straightforward manner. The predicted price of the ith house (in log form) for the base and current periods is calculated from the following equations, where a hat over a coefficients or a variable denotes its estimated or predicted value, respectively.

$$Ln\, \hat{p}_i^0 = \hat{\beta}_0 + \sum_{k=1}^{n} \hat{\beta}_k\, Ln\, X_{ik} + \sum_{j=1}^{l} \hat{a}_j\, C_{ij} \qquad (4)$$

$$Ln\, \hat{p}_i^1 = \hat{\beta}_0 + \sum_{k=1}^{n} \hat{\beta}_k\, Ln\, X_{ik} + \sum_{j=1}^{l} \hat{a}_j\, C_{ij} + \hat{\delta} \qquad (5)$$

So the price relative of ith house is simply $\hat{p}_i^1 / \hat{p}_i^0$ for all I (i=1, 2, ...n). Therefore a price index can be constructed directly from the coefficient(s) on the time dummy variable:

$$\hat{p}_{TD} = \Sigma\, \hat{p}_i^1 / \Sigma\, \hat{p}_i^0 = \exp(\hat{\delta}) \text{ for all } i.$$

---

6   In the literature, this method (involving only two consecutive periods) is known as the adjacent-period time dummy approach.

This method has an interesting statistical property when it is applied to a matched sample. Suppose we have a house sample S^0 in period 0 and S^1 in period 1. Assuming that the hedonic model (3) is estimated using Ordinary Least Square (OLS) on the pooled data $S^0 \cup S^1$. The regression residual are defined as

$$u_i^t = \ln p_i^t - \ln \hat{p}_i^t = \ln (p_i^t / \hat{p}_i^t), \text{ t= 0,1.}$$

Due to the inclusion of a constant term and a dummy variable, the residuals sum to zero in both periods, or:

$$\sum_{i \in S^0} \ln \left(\frac{p_i^0}{\hat{p}_i^0} \right) = \sum_{i \in S^1} \ln \left(\frac{p_i^1}{\hat{p}_i^1} \right) = 0. \tag{6}$$

Taking antilogarithms yields

$$\prod_{i \in S^0} \left(\frac{p_i^0}{\hat{p}_i^0} \right) = \prod_{i \in S^1} \left(\frac{p_i^1}{\hat{p}_i^1} \right). \tag{7}$$

If the sample does not change (i.e. prices of identical houses are observed in both periods), then $S^1 = S^0 = S$, and it follows from (7) that

$$\prod_{i \in S} \left(\frac{\hat{p}_i^1}{\hat{p}_i^0} \right)^{\frac{1}{n}} = \prod_{i \in S} \left(\frac{p_i^1}{p_i^0} \right)^{\frac{1}{n}}.$$

This is unweighted geometric mean of price relatives between periods 0 and 1, which is referred to in the index number literature as the Jevons index formula (Triplett, 2001).

Choice between HPIm and TD

The choice between the HPIm and TD has been subject to investigation by academics. According to Silver and Heravi (2004), the decision should be made based on the stability / instability of the hedonic function and the changeability of the (value of the) characteristics over time. The authors provided the following rules as a practical guidance:

• HPIm should be avoided and TD should be used when there is evidence of significant changes in the characteristics;

• TD should be avoided and HPIm should be used when there is evidence of significant parameter instability;

- If neither the change in parameters nor in characteristics is particularly large relative to the other, then a symmetric average, say geometric mean, of the two indexes is preferred;

- If both changes are significant and the values differ exponentially, a more appropriate estimate might be a weighted mean of the two indexes;

- Either the HPIm or TD will be acceptable if either the parameters are relatively stable or the characteristics do not change over time.

These rules appear to be quite useful when we choose between HPIm and TD.

3.3 Hedonic price indexes – An indirect approach

Although hedonic methods have many desired properties, they have certain limitations that have made it difficult to use them in the production of official statistics. For example, estimation and maintenance of a hedonic function can be costly and they require certain (economic and econometric) skills that are not commonly available among statistics practitioners. As a result, they could put excessive pressure on statistical agencies which are required to produce publications regularly and with nonnegotiable time and resource constraints. Therefore hedonic methods have not been extensively used *directly* in the production of official statistics, although they are becoming increasingly popular (Conniffe and Duffy, 1999).

However, while there may be significant resource costs involved in statistical agencies using hedonic methods broadly, the methods can be applied to a limited degree in less costly indirect way. For example,

- During normal production cycles, statisticians may still apply conventional and inexpensive methods (such as the one currently used by the ABS) to the production of regular publications and, at the same time, use a hedonic index (as a benchmark) to monitor the quality of indexes; or

- When statistical agencies design a new production system or the existing method undergoes a review, a hedonic function may be used to inform statistical choices (particularly in the areas of stratification, design of weighting patterns and treatment of outliers).

The ABS is unlikely to use hedonic methods *directly* in the production of the HPI in the short term. However there are possibilities for *indirect* use of hedonics. Therefore, it is useful to continue exploring hedonic methods and their application to the HPI.

4. TWO CASE STUDIES: HOBART AND ADELAIDE

The case study on Hobart and Adelaide was the first attempt by the ABS to apply hedonic methods to the construction of the house price index. In the study we applied various indexation methods mentioned in the previous section. In this paper, we will outline the main characteristics of the data and briefly discuss issues that we encountered in the process of data cleaning and econometric estimation.

4.1 Data for Hobart and Adelaide

The data for Hobart were provided by the Tasmanian Department of Primary Industries, Water and Environment (DPIWE). They cover the whole Hobart metropolitan area for eight consecutive quarters from June quarter 2002 to March quarter 2004.

The data for Adelaide were provided by the South Australian State Valuer-General's Office. They cover 11 consecutive quarters from September 2001 to March 2004.

The data from the two States contain a good deal of information about structural attributes (S) of individual houses, but they do not have any information about neighbourhood attributes (N) and locational attributes (L), other than address details of the houses.

Structural attributes (S)

Table 4.1 summarises the data obtained from Tasmania and South Australia.

From table 4.1, it is clear that only four variables are numerical and all others are categorical. Among the categorical variables, the 'house condition' is ranked from 1–10 by the data providers and we simply used the existing ranking in our estimation.

4.1 Data from Hobart and Adelaide

Characteristics	Hobart		Adelaide		Definition
	used	available	used	available	
Floor area (Floor) [1]	✓	✓	✓	✓	Numerical variable; size of floor area (m^2).
Land area (Land)	✓	✓	✓	✓	Numerical variable; size of land area (m^2).
Bedrooms	No	No	✓	✓	Numerical variable; number of bedrooms.
Number of main rooms (Rooms)	✓	✓	✓	✓	Numerical variable; number of main rooms.
Year of construction (Age)	✓	✓	✓	✓	Numerical variable; number of years after construction.
Wall material (Wall)	✓	✓	✓	✓	Categorical variable, 1 for brick and 0 for other wall materials (iron, rendered, stone etc.)
Roof material (Roof)	✓	✓	✓	✓	Categorical variable, 1 for galvanised iron or tile roof and 0 otherwise.
Municipality (LGA) [2]	✓	✓	No	No	
More than one bathroom (Bathroom)	No	No	✓	✓	Categorical variable, 1 for more than one bathroom and 0 for others.
Multi-storey house (Multi-storey)	No	No	✓	✓	Categorical variable, 1 for multi-storey and 0 for other houses.
Conventional house style (Conventional)	No	No	✓	✓	Categorical variable, 1 for conventional style and 0 for other styles.
House condition (Condition)	No	No	✓	✓	Numerical variable, Ranked from 1 (uninhabitable) to 10 (top quality and excellent).
Common plan type (Plan Type)	No	No	✓	✓	Categorical variable, 1 for common plan type and 0 for other plan types.
Property frontage (Frontage)	No	No	✓	✓	Categorical variable, 1 for irregular frontage and 0 for others.

1) In the first column, the terms in the bracket will be used in table 4.3 which contains a summary of the estimates of a hedonic function. The last column also defines how the categorical variables are defined when they are used as dummy variables in the estimation of hedonic functions.

2) LGA stands for 'local government area'.

Locational attributes (L) and Neighbourhood attributes (N)

To fill up the information gap required by hedonic methods, we have obtained two additional sets of data from sources within the ABS. The geographic data covers information such as the distances from specific streets to places such as the CBD, nearest school, shopping centre and hospital etc.. Once we know the address of a house, we are able to calculate the distance between the *centre point of the street* (in which the house is located) and the facilities (i.e. shops, schools and hospitals etc.).

However, this dataset has several limitations. For example, one would imagine the actual distance that residents needs to travel (by car, bus or train) between their homes and the CBD will be a sensible measure for this purpose. However, a 'straight line' distance is measured in the data. Despite this weakness, the data still contain information that is very useful for this project.

The second panel of table 4.2 summarises the main characteristics of the data.

4.2 Other variables included in modelling for Hobart and Adelaide

....

Socio-economic Indexes for Areas (SEIFA, composed from Population Census)

....

Variables	*Definition*

....

SEIFA(HD) = Highly disadvantaged,	SEIFA(HD) = 1 if SEIFA in the bottom quintile, else SEIFA(HD) =0
SEIFA(D) = Disadvantaged	SEIFA(D) = 1 if SEIFA in the second quintile, else SEIFA(D) =0
SEIFA(M) =Median	SEIFA(M)= 1 if SEIFA in the median quintile, else SEIFA(M) =0;
SEIFA(A) = Advantaged	SEIFA(A)= 1 if SEIFA in the forth quintile, else SEIFA(A) =0;
SEIFA(HA) = Highly Advantaged	SEIFA(HA)= 1 if SEIFA in the top quintile, else SEIFA(HA) =0;

....

Data from Geography Section

....

Variables	*Definition*

....

CBD	Numerical variable, distance to the CBD
Primary School	Categorical variable, 1 if Primary School presence within 1 km, 0 otherwise.
Secondary School	Categorical variable, 1 if Secondary School presence within 1 km, 0 otherwise.
College	Categorical variable, 1 if College School presence within 1 km, 0 otherwise.
University	Categorical variable, 1 if University presence within 2 km, 0 otherwise.
Shops	Categorical variable, 1 if Shopping Centre presence within 2 km, 0 otherwise.
Hospital	Categorical variable, 1 if Hospital presence within 5 km, 0 otherwise.
Emergency	Categorical variable, 1 if Emergency Services presence within 5 km, 0 otherwise.

....

A second set of data were extracted from the 2001 Population Census. The Population Census contains a wide range of socio-economic variables that have been summarised at various aggregation levels (i.e. defined in terms of geographic strata).[7] We used them for the purpose of quantifying the neighbourhood attributes (N). Quite obviously, it is not a straightforward task to build a credible and defensible profile of neighbourhood attributes (N) and the task itself could constitute a major research project. We had neither the time nor resources to conduct a full-scale research on this topic.

To overcome this problem, we used the *Index of Relative Socio-economic Advantage and Disadvantage* derived from the 2001 population census (See *Census of Population and Housing – Socio-economic Indexes for Areas*, ABS cat. no. 2039.0).[8]

Although we are not certain about how well the Index of Relative Socio-economic Advantage and Disadvantage fits the purpose of this study, the preliminary findings suggested that it appeared to have worked reasonably well. However, once there is a

---

7   Details of the variables contained in the Population Census can be found in ABS (2001).

8   More details of the methodology can be found in ABS (2003).

strong prospect of applying hedonic methods to all the eight capital cities, we will further investigate this issue.

MAC members may wish to share with us their experience and insights or direct us to useful literature on this particular issue.

4.2 Hedonic function

In this project, we explored both the Hedonic Price Imputation (HPIm) and time dummy (TD) methods in the estimation of hedonic functions. In this section, we have chosen one set of estimates from each approach as examples to explain issues that we have confronted in the study.

Data treatments

The data provided by the Tasmanian DPIWE and South Australian Valuer-General contain incomplete information. A typical problem is that, for a few houses, data on certain characteristics are missing. In this study, we did not attempt to impute the missing values. Instead we removed all the observations that were found to have contained incomplete information, before the functions were estimated. We expected that exclusion of observations (with missing values) would cause a negligible impact on the results because, from the datasets for both Hobart and Adelaide, a very small number of observations were excluded.

We also need to deal with prices that appeared to be 'unrepresentative'. In this study, we used a simple method which is different from the one the ABS has used in the construction of the HPI. Before we ran the regressions, we removed the top and bottom 5 percent of the observations from each quarter's data. Although this treatment is somewhat arbitrary, it appears to have worked well.

It is interesting to note that removing certain percentages of top and bottom observations does not guarantee an absence of 'unusual' values. A further examination of data revealed some other problems. For example, according to the original data files from Hobart, one house had 72 bedrooms which was not impossible but it was unlikely and certainly unrepresentative. We also excluded this and similar observations.

We are interested in any experience that MAC members wish to share with us, that may enable us to deal with missing observations and outliers in a more sophisticated and 'scientific' way.

Selection of variables and functional form

In estimating a hedonic function, it is very important to determine its functional form and choose an appropriate procedure to select variables to be included in the function. It was not a straightforward exercise to find proper solutions (to both issues) and they were complicated by the fact that the two decisions were not independent.

We planned but did not have time to conduct an analysis that would give us confidence in these choices. So the methods we have adopted may appear quite inadequate.

This is an area we particularly seek advice and guidance from the MAC members.

In this project, we decided to transform all the numerical variables into double-log form before estimating hedonic functions. In the literature, log-linear and linear forms are also used in the literature and Box–Cox tests may be used to determine an appropriate form.

Our decision was based on three reasons. First, in the literature, it was suggested double-log was more appropriate for the housing market (Gatzlaff and Ling, 1994). We were also more comfortable with the interpretation (of the relationships between the price and housing characteristics) under the double-log form. For example, the relationship between the prices (that consumers are willing to pay) and the size of a house (or its land) is unlikely to be linear – the 'marginal value' of additional land is likely to vary depending on the price level.

Second, the double-log form has certain superior statistical properties. For example, according to Diewert (2001, 2003), the residual from a double-log function is less likely to suffer heteroskedasticity and it is more consistent with the economic theory.

Third, if we choose the time dummy method (TD) with the variables in double log form, it becomes a very straightforward exercise to transform the coefficients of the time dummy variable into a price index (see details in Section 3.2). This feature makes the hedonic index easier to implement.

We used a backward elimination techniques incorporated in SAS (i.e. SAS Proc Reg / selection = backward) in the variable selection. This is an automatic procedure and it begins by calculating F statistics for a model, including all of the independent variables. Then the variables are deleted from the model one by one until all the variables remaining in the model produce F statistics at a predetermined significance level.

This variable selection procedure has an advantage of being easy to use, particularly in the regular production of statistical publications. But using this method runs certain

risks. For example, if the statistic properties of the samples (e.g. the strengths of relationships between important variables) varied significantly from one period to another, certain important variables could be mistakenly excluded.

> We hope that MAC members could provide some advice or guidance on developing sound procedures (or rules) of determining functional form and choosing variables for the purpose of improving the HPI.

Estimates of hedonic function

Table 4.3 is an example of the estimates of a hedonic function using hedonic price imputation method (HPIm). This function was estimated based on equation (2) and the estimation was run on the March quarter 2004 data for Adelaide. Other estimates (based on Adelaide and Hobart data) are broadly similar. A full set of estimates of hedonic functions are available to MAC members on request.

4.3 Coefficients of a hedonic function for Adelaide (March 2004)

Variables [a]	Coefficients	Standard errors	t–Statistics	Pr > t
Intercept	3.4182	0.1039	32.90	<0.0001
Log (Land))	0.1362	0.0120	11.35	<0.0001
Log (Age)	−0.0500	0.0062	−8.06	<0.0001
Log (Floor)	0.3726	0.0135	27.60	<0.0001
Log (Condition)	0.2081	0.0307	6.78	<0.0001
Log (CBD)	−0.2823	0.0091	−31.02	<0.0001
(Brick) Wall	−0.0459	0.0088	−5.22	<0.0001
Multistorey	0.0687	0.0113	6.08	<0.0001
Bathroom	0.1063	0.0135	7.87	<0.0001
Conventional	−0.0641	0.0085	−7.54	<0.0001
Plan Type	−0.0795	0.0112	−7.10	<0.0001
Emergency	0.0506	0.0127	3.98	<0.0001
Shops	−0.0176	0.0082	−2.15	0.0308
Primary School	−0.0473	0.0115	−4.11	<0.0001
University	−0.0288	0.0089	−3.24	0.0012
SEIFA (HD)	−0.1414	0.0131	−10.79	<0.0001
SEIFA (D)	−0.0885	0.1171	−7.56	<0.0001
SEIFA (A)	0.0895	0.0125	7.16	<0.0001
SEIFA (HA)	0.1537	0.0157	9.79	<0.0001
F Value	416.0300			<0.0001
R–Square	0.6873			

a) We initially included 'number of main bedrooms', 'major roof material', 'frontage', 'presence of secondary school' and 'presence of colleges' in the regression (see the definitions of these variables in table 4.1). They were excluded in the variable selection process. The variable names are explained in tables 4.1 and 4.2.

Table 4.3 only presents the estimates for the variables that were found to be statistically significant.

Several observations about the estimates are worth noting.

The signs of the (statistically significant) variables are plausible and consistent with our expectations. For example, the estimates suggest that the price of a house will be higher if

- land or/and floor areas are larger;

- the house is newer, the condition is better, it has brick wall or more than 1 bathroom;

- the house is close to the CBD and emergency services; and

- the house is located in the better socio-economic environment (indicated by "HA" or "A" in the SEIFA variables).

The estimates also revealed certain aspects of consumer preference, where it is difficult to form some kind of *a priori* belief. For example, consumers appeared to be willing to pay a higher price if

- the house is a multi-story building of an unconventional style and it is built according to a 'common plan'; and

- the house is not too close to its local primary (but not necessarily secondary) school, local shopping centre and universities.

The F statistic (>416.03) and R–square value ($=0.687$) indicate that the hedonic function gives a reasonable approximation of the relationships between house prices and their characteristics.

4.3 Hedonic price indexes

This section presents three hedonic price indexes (for Adelaide) and compares them with an index constructed based on the current HPI method. More results can be found in Appendix B.

Four indexes are presented in table 4.4. The first column (HPI–S) shows the index using the current HPI method. It should be noted that the index is not perfectly consistent with the published figures, because the treatment of large outliers is different.

In the construction of hedonic price imputation index (HPIm–I), Option 1(a), was adopted to estimate the hedonic functions. Using this option, a hedonic function was estimated on each cross-section data for each of all the quarters and the characteristic variables were selected using the automatic backward elimination technique.

4.4 Comparison of indexes – Adelaide

	HPI–S	DI–I	TD–Ia	TD–Ib
September 2001	100.0	100.0	100.0	100.0
December 2001	104.1	110.3	104.8	104.6
March 2002	109.0	104.7	110.1	110.0
June 2002	113.5	114.3	114.2	114.1
September 2002	119.9	119.8	119.8	119.7
December 2002	126.1	120.6	120.5	120.3
March 2003	131.5	133.2	132.6	132.4
June 2003	141.5	141.4	141.2	141.0
September 2003	147.3	150.4	149.7	149.6
December 2003	156.7	161.9	161.7	161.9
March 2004	161.4	163.9	163.7	164.1

In table 4.4, we presented two time dummy indexes. For TD–Ia, hedonic functions were estimated from a pooled data that covered only two consecutive quarters and this is sometimes called 'chained time dummy method'. For TD–Ib, one hedonic function was estimated from a pooled data that covered all the periods (from September 2001 to March 2004) and the coefficients of the time dummy variables were transformed into a price index (see details in Section 3.2).

Three indexes (i.e. HPI–S, HPIm–I and TD–Ia) are presented in figure 4.5.

4.5 House price indexes – Adelaide

Several observations can be made about the results. First, apart from a couple of periods, the two hedonic indexes trace each other closely. This implies that choosing different hedonic methods (i.e. HPIm and TD) may not make a big difference for this dataset.

Second, the hedonic indexes do not differ significantly from the HPI–S, in terms of their 'long term' trends. This suggests that the structural attributes (S) did not make a significant impact on the indexes. There are at least three possible interpretations.

1. Relative to the impact of locational attributes (L) and neighbourhood attributes (N), the changes in the composition of the qualities of individual houses (defined in the structural attributes) were not significant enough to influence the behaviour (of the long term trend) of the HPI.

2. This is a unique time period where prices of all houses (of different qualities) increased at a similar pace, meaning that there were no large compositional changes reflected in the sample.

3. The effect of increasing land prices in the period swamped all other effects.

If the second interpretation is true, then we will observe quite different behaviours between the HPI–S and the hedonic indexes if we apply hedonic methods to other cities or different time periods. However, our results from the Hobart data suggested that the similarity between HPI–S and hedonic indexes was not unique to Adelaide. Whether we believe this interpretation or not, it will be necessary to take housing qualities (i.e. structural attributes) into account, in order to produce a consistent HPI.

If the first interpretation is true, then this is an encouraging result. This is because, the HPI will be a reliable indicator of house prices, so long as we properly control locational attributes (L) and neighbourhood attributes (N). This can be achieved through properly stratifying the geography of the eight cities.

We plan to further investigate this issue in the future and would appreciate MAC members views on this matter.

Finally, the HPI–S appears to be much smoother than the two hedonic indexes. This implies that the current methods of identifying outliers and the weights used in the index compilation may have worked quite well. Of course, this is only indicative and we plan to further investigate this issue before drawing a firm conclusion.

4.4 Limitations

In this section, we reported some of the earlier findings from our exploration of hedonic methods based on the data from Adelaide. The results are encouraging and they signalled a promise of using hedonic methods, directly or indirectly, for the future HPI.

However, we are hesitant to draw firm conclusions at this stage due to limitations in terms of both data and the statistical tests that we were able to run. In particular, we

are conscious of the following major limitations that we wished to, but we were unable to, overcome.

- We only had data from Hobart and Adelaide. We may obtain very different results when we apply hedonic methods to large cities (such as Sydney and Melbourne) where the situations (e.g. housing markets, socio-economic conditions and consumer preferences) are more complex. Therefore, it is too early to make any generalisations;

- The data used in this study covered only a short time period and, more importantly, this was a period where housing markets were booming everywhere. The results could be unique to this stage of housing market cycle;

- The data that are available to us only covered detached houses. Town houses and apartments were not included in the HPI and our analysis. This means that our sample might have a narrow representation in both the market (i.e. more about market of owner-occupied than investment properties) and the goods involved in the transactions (i.e. properties with well defined and separate titles). When investment properties, town houses and apartments are included in the analysis, the situation could become more complex because certain attributes (e.g. land size etc.) will be less well defined and the relationships between price and structural, locational and neighbourhood attributes could be more complex.

- A proper hedonic function depends on the availability of well-defined attributes variables (i.e. S, L and N). According to our estimation results, the attributes variables used in estimating our hedonic functions served our purpose pretty well. However, they could certainly be further improved in terms of coverage and definition. For example,

    - our structural attributes appeared to have missed some important variables (e.g. number of garages);

    - certain locational attributes should be better defined (i.e. using actual travel distances rather than "straight-line" distances), if possible; and

    - SEIFA variables could be decomposed to represent various aspects of neighbourhood attributes.

We hope that these variables could be further improved in the future analysis.

5. FUTURE ANALYSIS OF HPI AND QUESTIONS FOR MAC MEMBERS

The two major focuses of our future analyses are

- more refined research on hedonic methods; and

- the use of the research outcomes to inform major statistical choices required in the production of the HPI.

However, an important background of the research is the review of the HPI currently being undertaken by the ABS.[9] The review is likely to lead to demand for analysis of new data (of different contents, specifications and format) from alternative sources.

At this stage, however, we are uncertain about what future data will look like and it is very hard to predict analytical questions that may arise in the future. In this circumstance, we would like to open the whole study for suggestions. MAC members may wish to share experience or insights on general issues such as

- Methods that may be used to deal with complex issues (e.g. housing markets, socio-economic conditions and consumer preferences), when we analyse the housing market in large cities; or

- Methods that may be used to apply hedonic methods to investment properties (i.e. town houses and apartments in particular); or

- Methods to better define locational and neighbourhood attributes.

We will also be appreciative if MAC members choose to advise on one or more specific issues mentioned in the text, which include

- Design of geographic stratification for the housing market (see details in Section 2.2)

- Methods of identifying and removing 'outliers' (Sections 2.2 and 4.2)

- Method to choose between two hedonic price imputation methods (Section 3.2)

- Procedure (or rules) that may be used in determining hedonic functional form and variable selection (Section 4.2)

- The reasons why hedonic indexes are similar to the indexes based on the current HPI method (Section 4.2)

We will be very happy to hear comments and suggestions on any other issues, specific or general.

---

9   This review was completed during 2005 and led to a revised method of compiling the HPI (See Australian Bureau of Statistics, 2005b).

REFERENCES

Australian Bureau of Statistics (2001) *Census Dictionary*, cat. no. 2901.0, ABS, Canberra.

—— (2003) *Information Paper: Census of Population and Housing – Socio-economic Indexes for Areas*, cat. no. 2039.0, ABS, Canberra.

—— (2004) *House Price Indexes: Eight Capital Cities*, cat. no. 6416.0, ABS, Canberra.

—— (2005a) *Introduction of Hedonic Price Indexes for Personal Computers*, ABS Information Paper, cat. no. 6458.0, ABS, Canberra.

—— (2005b) *Renovating the Established House Price Index, Australia*, ABS Information Paper, cat. no. 6417.0, ABS, Canberra.

Conniffe, D. and Duffy, D. (1999) "Irish House Price Indices – Methodological Issues", *Economic and Social Review*, 30(4), pp. 403–423.

Crone, T.M.; Nakamura, L.I. and Voith, R.P. (2004) "Hedonic Estimates of the Cost of Housing Services: Rental and Owner-Occupied Units", Paper presented at the International Conference on Index Number Theory and the Measurement of Prices and Productivity, Vancouver, British Columbia, 30 June–3 July, 2004.

de Haan, J. (2003) "Time Dummy Approaches to Hedonic Price Measurement", Paper presented at the Seventh Meeting of the International Working Group on Price Indices, 27–29 May 2003, INSEE, Paris.

Diewert, W.E. (2001) *Hedonic Regressions: A Consumer Theory Approach*, Discussion Paper, Department of Economics, University of British Columbia.

Fox, K.J., Hill, R. and Diewert, W.E. (2004) "Identifying Outliers in multi-output Models", *Journal of Productivity Analysis*, 22, pp. 73–94.

Gatzlaff, D.H. and Ling, D.C. (1994) "Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies", *Journal of Urban Economics*, 40, pp. 221–244.

Goodman, A.C. (1989) Topics in Empirical Urban Housing Research, "The Economics of Housing Markets", in R. Muth & A. Goodman, Harwood Academic, Chur, Switzerland, pp. 49–146.

Melser, D. (2004) "The Hedonic Regression Time-Dummy Method and the Monotonicity Axioms", (draft working paper), School of Economics, University of New South Wales, Sydney, NSW 2052, Australia.

Silver, M. and Heravi, S. (2004) "The Difference Between Hedonic Imputation Indexes And Time Dummy Hedonic Indexes for Desktop PCs", Working Paper (2004/02), Centre for Applied Economic Research, University of New South Wales.

Triplett, J.E. (2001) *Handbook on Quality Adjustment of Price Indexes for Information and Communication Technology Products*, OECD, Paris.

Williams, A. (1991) "A Guide to Valuing Transport Externalities by Hedonic Means", *Transport Review*, 11(4), pp. 311–324.

APPENDIXES

A. PRICE INDEX FORMULAE

Let p_{it} be the price and q_{it} be the quantity of the ith (i = 1, 2, ..., n) commodity transacted in the time period t. The current value v_{it} and the expenditure share w_{it} of commodity i in period t are given by

$$v_{it} = p_{it} q_{it}, \quad w_{it} = \left(\frac{v_{it}}{\sum_{i=1}^{n} v_{it}} \right) \text{respectively.}$$

Laspeyres Price Index is defined by

$$L_t = \left(\frac{\sum_{i=1}^{n} p_{it} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \right) \times 100 = \sum_{i=1}^{n} w_{i0} \left(\frac{p_{it}}{p_{i0}} \right) \times 100$$

Paasche Price Index is defined by

$$P_t = \left(\frac{\sum_{i=1}^{n} p_{it} q_{it}}{\sum_{i=1}^{n} p_{i0} q_{it}} \right) \times 100 = \left\{ \sum_{i=1}^{n} w_{it} \left(\frac{p_{i0}}{p_{it}} \right) \right\}^{-1} \times 100$$

Fisher Price Index is defined by

$$F_t = (L_t P_t)^{\frac{1}{2}}$$

Tornqvist Price Index is defined by

$$T_t = \prod_{i=1}^{n} (P_{it}/P_{i0})^{\frac{1}{2}(w_{i0} + w_{it})} \times 100$$

B. HEDONIC INDEXES FOR HOBART

B.1 House price indexes – Hobart

B.2 Comparison of indexes – Hobart

	Weighted Mean	Hedonic Imputation	Flex-Hedonic Imputation	Time Dummy Hedonic	Chained TD Hedonic
June 2002	100.0	100.0	100.0	100.0	100.0
September 2002	104.1	108.0	108.6	107.3	107.5
December 2002	114.1	113.1	114.1	112.8	113.0
March 2003	133.7	126.8	127.0	125.7	125.7
June 2003	140.9	140.3	141.8	141.1	141.7
September 2003	165.0	163.8	165.6	163.9	164.2
December 2003	176.1	181.3	185.3	182.6	183.4
March 2004		183.2	189.8	186.6	187.1

FOR MORE INFORMATION . . .

INTERNET **www.abs.gov.au**   the ABS web site is the best place for data from our publications and information about the ABS.

LIBRARY A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our web site for a list of libraries.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our web site, or purchase a hard copy publication. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All ABS statistics can be downloaded free of charge from the ABS web site.

WEB ADDRESS www.abs.gov.au