

**Unclassified**

**STD/DOC(2011)1**

Organisation de Coopération et de Développement Économiques  
Organisation for Economic Co-operation and Development

**04-Feb-2011**

**English - Or. English**

**STATISTICS DIRECTORATE**

**HEDONIC PRICE INDEXES FOR HOUSING**

**STATISTICS DIRECTORATE**

**WORKING PAPER No 36**

*This paper has been prepared by Robert Hill (Department of Economics, University of Graz, Austria).*

*The complete document is only available in PDF format.*

For further information please contact:  
Robert Hill  
E-mail: [robert.hill@uni-graz.at](mailto:robert.hill@uni-graz.at)

**JT03295979**

Document complet disponible sur OLIS dans son format d'origine  
Complete document available on OLIS in its original format



STD/DOC(2011)1  
Unclassified

English - Or. English

## OECD STATISTICS WORKING PAPER SERIES

The OECD Statistics Working Paper Series - managed by the OECD Statistics Directorate - is designed to make available in a timely fashion and to a wider readership selected studies prepared by OECD staff or by outside consultants working on OECD projects. The papers included are of a technical, methodological or statistical policy nature and relate to statistical work relevant to the Organisation. The Working Papers are generally available only in their original language - English or French - with a summary in the other.

Comments on the papers are welcome and should be communicated to the authors or to the OECD Statistics Directorate, 2 rue André Pascal, 75775 Paris Cedex 16, France.

The opinions expressed in these papers are the sole responsibility of the authors and do not necessarily reflect those of the OECD or of the governments of its Member countries.

---

**<http://www.oecd.org/std/research>**

---

OECD/OCDE, 2010

Applications for permission to reproduce or translate all or part of this material should be made to: OECD Publications, 2 rue André-Pascal, 75775 Paris, Cedex 16, France; e-mail: [rights@oecd.org](mailto:rights@oecd.org)

## ABSTRACT

Every house is different. It is important that house price indexes take account of these quality differences. Hedonic methods which express house prices as a function of a vector of characteristics (such as number of bedrooms and bathrooms, land area and location) are particularly useful for this purpose. In this report I consider some of the developments in the hedonic methodology, as it is applied in a housing context, that have occurred in the last three decades. A number of hedonic house price indexes are now available. However, it is often difficult to see how these indexes relate to each other. For this reason the paper attempts to impose some structure on the literature by developing a taxonomy of hedonic methods, and then show how existing methods fit into this taxonomy. Also discussed are some promising areas for future research in the hedonic field, particularly the use of geospatial data and nonparametric methods for better capturing the impact of location on house prices. The main criticisms of the hedonic approach are evaluated and compared with the repeat sales and stratified median methods. The overall conclusion is that the advantages of the hedonic approach outweigh its disadvantages.

## RÉSUMÉ

Chaque maison est différente. Il est important que les indices de prix de l'immobilier tiennent compte de ces différences de qualité. Les méthodes hédoniques qui expriment des prix de l'immobilier en fonction d'un vecteur de caractéristiques (telles que le nombre de chambres et salles de bains, superficie et l'emplacement) sont particulièrement utiles à cette fin. Dans le présent rapport je considère certains développements dans la méthode hédonique, telle qu'elle est appliquée dans un contexte de logement, qui se sont produits au cours des trois dernières décennies. Un certain nombre d'indices des prix hédoniques pour l'immobilier est maintenant disponible. Cependant, il est souvent difficile de voir comment ces indices sont liés les uns aux autres. C'est pour cette raison que ce papier tente d'imposer une structure sur la littérature en développant une taxonomie des méthodes hédoniques et montre alors comment les méthodes existantes s'insèrent dans cette taxonomie. Sont aussi discutées certains domaines prometteurs pour les futures recherches dans le domaine hédonique, notamment l'utilisation des données géo-spatiales et de méthodes non paramétriques pour mieux capter l'impact de l'emplacement sur le prix de l'immobilier. Certains des principaux critiques de l'approche hédonique sont évalués et comparés avec les ventes répétées et des méthodes stratifiées médians sont aussi discutées. Dans l'ensemble, on conclut que les avantages de l'approche hédonique l'emportent sur ses inconvénients.

# TABLE OF CONTENTS

1. Introduction: The Importance of House Price Indexes
2. The Problem with Measuring Movements in House Prices using Median or Repeat-Sales Indexes
3. A Short History of Hedonic Indexes
  - 3.1 The origins of hedonic indexes
  - 3.2 Applications in a housing context
4. A Taxonomy of Hedonic Price Indexes for Housing
  - 4.1 Time-dummy methods
  - 4.2 Imputation methods
  - 4.3 Characteristics index methods
  - 4.4 Nonparametric methods
  - 4.5 Repeat sales and hedonic hybrid methods
5. Geospatial Data and Spatial Dependence
6. Estimation of Hedonic Models
7. Criticisms of Hedonic Price Indexes
  - 7.1 Omitted variables bias
  - 7.2 Functional form misspecification
  - 7.3 Data mining and lack of transparency and reproducibility
  - 7.4 Sample selection bias
8. Conclusion

# 1 Introduction: The Importance of House Price Indexes

The real estate sector can be divided into residential and commercial categories. My focus in this report is on residential housing. According to Syz (2008) about one third of total wealth around the world (about \$21.6 trillion) is tied up in residential housing. Case, Quigley and Shiller (2005) find that changes in house prices have a larger impact than changes in stock-market prices on household consumption in the U.S. and other developed countries. In addition to the direct wealth effects, Case and Quigley (2008) predict that the main impact of a decline in US house prices will be felt through the income effect (i.e., due to falling employment in the construction and real estate industries) and the effect on financial markets. The importance of the housing market to the broader economy has been clearly demonstrated by the events that followed the collapse of the subprime mortgage market in the US. The subsequent fall in house prices, anticipated by Shiller (2007, 2008) and Case and Quigley (2008), triggered a global financial crisis.

This apparent propensity for boom and bust cycles in the housing market, and the way they impact on the rest of the economy, has acted to raise the profile of house price indexes and the level of scrutiny they receive. A housing boom can cause a significant redistribution of wealth from the younger to older generation, and increase inequality within each generation as renters (who are predominantly poorer) miss out on the capital gains received by home owners. Meanwhile, busts in the housing market can trigger a surge in defaults by households on their mortgages (particularly at the lower end of the income distribution), which can also destabilize the banking sector and cause credit markets to seize up.

More generally, house price indexes provide a barometer for the state of the economy that is useful for government fiscal policy, central bank monetary policy and financial markets. It is likely that central banks in particular may want to pay closer attention to asset prices (especially for housing) in future when setting monetary policy in the hope of avoiding the destabilizing effects of booms and busts. Potentially, this can be done in a number of ways ranging from adopting asset price as well as consumer price inflation targets to modifying the way the cost of owner-occupied housing services are calculated in the consumer price index (CPI). The treatment of housing in the CPI and other measures of inflation is a matter of some importance. Cecchetti (2007) makes the point as follows:

“With inflation targets winning the world of Central Banking, methods for measuring inflation have direct policy consequences. The big question for inflation measurement is how to handle housing.” (Cecchetti 2007)

Most countries currently either exclude owner-occupied housing completely from the CPI (e.g., the European Union harmonized index) or include it using the rental equivalence approach (e.g., the US). Rental equivalence entails imputing rents for owner-occupied housing. In the period leading up to the financial crisis, house prices in most countries rose rather faster than rents (the US case is discussed in Poole, Ptacek and Verbrugge 2005). As a result, the rise in house prices did not have much impact on the US CPI, thus perhaps explaining the failure of the Federal Reserve to raise interest rates more vigorously during the years preceding the financial crisis.

The main alternatives to rental equivalence for including owner-occupied housing in the CPI, namely the user cost and acquisitions approaches, also have their problems (see for example Diewert 2007a). User cost, which from a theoretical perspective probably has the most appeal, is normally defined as the sum of foregone interest, depreciation and capital gains:

$$UC_t = r_t P_t + \text{dep}_t - (P_{t+1} - P_t).$$

There is some debate as to whether the last term should be expected or ex post capital gains, and over the impact capital gains have on the volatility of the index (see Verbrugge 2008, Diewert and Nakamura 2009, Diewert, Nakamura and Nakamura 2009, Garner and Verbrugge 2009, and Schreyer 2009). What is clear, however, is that a house price index  $P_t$  is an important input into the user-cost formula.

The acquisitions approach attempts to treat housing in the same way as any other consumer durable. It measures the cost of all net dwellings acquired by households (which consist primarily of newly built dwellings). Practitioners of this approach include Australia and New Zealand. Eurostat is planning to use an acquisitions approach when owner-occupied housing is eventually included in the HICP. The acquisitions approach, as it is normally applied, simply measures the cost of building materials. This is because it is argued that the purchase of land is an investment and hence not consumption. A more sophisticated approach would try to separate structure and land prices econometrically (see Diewert, de Haan and Hendriks 2010).

Case, Shiller and Weiss (1993) and Shiller (2008) meanwhile stress another important

application of house price indexes. They argue that the development of derivative markets linked to house price indexes could improve the management of risk throughout the economy. Households may want to diversify their asset portfolios away from housing. According to Halifax Financial Services (2007), housing in 2007 accounted for 43 percent of household wealth in the UK, while according to Denk (2006) in 2004 it accounted for 39 percent of household wealth in the US. Owner-occupier households with most of their wealth invested in their house (and perhaps large mortgages) could reduce their exposure to housing by buying put options on a house price index. Pension funds and mutual funds, by contrast, hold very small amounts of their wealth portfolios in the form of residential housing, and hence could significantly diversify their portfolios by buying into the housing market through the purchase of call options on a house price index.

Until recently, however, the quality of house price indexes has been rather poor given their importance. This point was made forcefully for example by the Governor of the Reserve Bank of Australia in 2004.

Housing is the biggest asset in the country. Certainly for the household sector it is about 60 to 70 percent of their total wealth. It is an extremely important asset class for most people, yet the information we have on prices is hopeless compared with the information we have on share prices, bond prices, and foreign exchange rates, and even the information we have on commodity prices, export prices, import prices and consumer prices. It really is probably the weakest link in all the price data in the country so I think it is something that I would like to see resources put into. (Ian Macfarlane, Governor of the Reserve Bank of Australia, 4 June 2004).

The development of reliable house price indexes has been hampered by a combination of a lack of suitable data sets and the fact that every house is different both in terms of its physical characteristics and its location. The extreme heterogeneity of housing has required the development of new methods for quality-adjusting measured price changes.

The situation has improved in the last few years, although housing is still probably the weakest link in the list of macroeconomic price statistics. Hedonic methods which express house prices as a function of a vector of characteristics are starting to prove particularly useful for this purpose. This report considers some of the developments in the hedonic methodology, as it is applied in a housing context, that have occurred in the last three

decades.

One ongoing initiative, in particular, that is worth mentioning is the Eurostat Handbook on Residential Property Price Indices (RPPI). A preliminary draft is available online at the following website: [http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/methodology/owner\\_occupied\\_housing\\_hpi/rppi\\_handbook](http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/methodology/owner_occupied_housing_hpi/rppi_handbook). My report covers some similar ground, although it is more focused specifically on hedonic indexes.

More generally, quite a few quality-adjusted house price indexes are now available. The main indexes in the US use the repeat-sales approach. The two most notable are the Standard and Poor's/Case-Shiller (SPCS) Home Price and Office of Federal Housing Oversight (OFHEO) indexes. Elsewhere, the hedonic approach dominates. Notable examples are the Halifax Home Price Index in the UK, the permanent tsb index in Ireland, the Conseil Supérieur du Notariat (CSN) and INSEE (the national statistical office of France) index in France, the Zürcher Wohneigentumsindex (ZWEX) in Switzerland, the indexes published by the statistical offices of Finland, Norway and Sweden, and the RPData-Rismark indexes in Australia. Other less transparent hedonic indexes include the Verbund Deutscher Pfandbriefbanken (VDP) and Hypoport AG indexes in Germany and the Recruit Residential Price, Residential Market and Tokyo Area Condominium Market Indexes in Japan.<sup>1</sup>

Derivatives markets are also gradually starting to appear for some of these indexes. The first foray into house price index derivatives happened in 1991, when the now defunct London Futures and Options Exchange (FOX) launched derivatives on the Nationwide Anglia Building Society House Price (NAHP) index in the UK. The product was withdrawn after five months, due to low trading volume and accusations of attempts to artificially inflate it (see Kawaguchi 2007 and Shiller 2008). This botched start probably held back the development of housing derivatives by many years. In 2003 Goldman Sachs started trading derivatives on the Halifax House Price Index (HPI) in the UK. The Chicago Mercantile Exchange in 2006 started trading futures and options contracts on the SPCS Home Price Indexes, the Zurich Cantonal Bank (ZKB) also in 2006 started trading derivatives on the ZWEX index. The Australian Stock Exchange is due to start trading derivatives soon on the RPData-Rismark indexes. In spite of these significant developments, the impact of house price index derivatives markets has thus far been disappointing. Most of the existing derivatives markets suffer from low trading volume and liquidity. It remains to be seen

---

<sup>1</sup>Some details of the Japanese indexes can be found in Chapter 10 of the draft Eurostat RPPI Manual, although no actual formulas are provided.



whether this situation will improve with time.

## 2 The Problem with Measuring Movements in House Prices using Median or Repeat-Sales Indexes

House price indexes can be based on actual market data or expert surveys or even a combination of the two (for an example of the latter see de Vries et al. 2009).<sup>2</sup> Market prices for indexes using actual market data can take the form of asking prices, the price on which a mortgage backed offer is based, the price at which contracts are exchanged, or the actual price that is eventually officially recorded. Index providers trade off timeliness against accuracy depending on which market price they use. These tradeoffs are discussed further in Acadametrics (2009).

The simplest type of house price index is a median index that tracks the change in the price of the median dwelling from one period to the next. Examples include the National Association of Realtors (NAR) index in the US, and the Real Estate Institute of Australia (REIA) and LJ Hooker/BIS Shrapnel indexes in Australia. Derivatives on the NAR index have been traded on the Chicago Board Options Exchange since 2006.

The main attractions of median indexes are that they require less data, are easy to compute and easy to understand. Their main disadvantage is that they confound changes in prices with quality differences, and hence may provide very noisy estimates of the change in the cost of housing. This is because the quality of the median dwelling will tend to differ from one period to the next. For example, suppose there are two regions in a city denoted by  $A$  and  $B$ , and that region  $A$  is richer and hence has more expensive houses than region  $B$ . Suppose further that in 2006 and 2008 most of the houses sold (including the median) are from region  $A$ , while in 2007 most houses sold (including the median) are from region  $B$ . It is likely therefore that the median index will record a large rise from 2006 to 2007 and then a large fall from 2007 to 2008. Such an index could be a very poor indicator of what is actually happening in the housing market.

Some median index providers try to address this problem by computing stratified medians. Stratification (often alternatively referred to as mix-adjustment) in its simplest form

---

<sup>2</sup>This section draws extensively on Hill, Melser and Syed (2009).

divides a city into geographical regions and then computes a separate median for each region. The changes in the median indexes for each region are then averaged, usually by taking an arithmetic or geometric mean to obtain the overall price index for that period. While stratification should reduce the amount of noise in the index, it will not eliminate it. Within each region, it will still be the case that the median dwelling sold in one period will tend to be of either superior or inferior quality to the median sold in the previous period. These differences will not necessarily offset each other from one region to the next. More sophisticated median indexes (see for example Prasad and Richards 2006) stratify by structural attributes of dwellings within regions, the physical location of the dwelling, and neighborhood characteristics of regions. The Established Homes Price Index published by the Australian Bureau of Statistics (ABS) is an example of such an index (see Australian Bureau of Statistics 2006). Stratified median indexes, such as those of Prasad and Richards and of ABS, can be viewed as an intermediate step between a simple median and a truly hedonic index.

As well as providing a noisy estimate of price changes, a median index may also be subject to systematic bias. Suppose for example that the average quality of houses sold improves over time. A median index will ignore this fact, and hence will be an upward-biased measure of the quality-adjusted price of housing. Stratification is of little use for dealing with this problem.

A repeat-sales index, as the name suggests, is computed only from repeat-sales data. Restricting the comparison to repeat sales ensures that each price relative compares like with like. One problem with this reasoning, however, is that the same dwelling at two different points in time is not necessarily the same. The best known repeat-sales indexes are the Standard and Poor's/Case-Shiller (SPCS) Home Price Indexes in the US. These are computed for 20 cities (see Standard and Poor's 2008). The Office of Federal Housing Oversight (OFHEO) also computes repeat-sales indexes for the US (see Calhoun 1996). Residex and RPData-Rismark compute repeat-sales indexes for Australian cities, while the UK and Dutch Land Registries compute repeat-sales indexes for the UK and the Netherlands, respectively. Another repeat-sales index in the Netherlands is the Woningwaarde Index Kadaster (House Price Index Kadaster), see Jansen et al. (2008).

The repeat-sales method is usually attributed to Bailey, Muth and Nourse (1963), although Shiller (2008) traces back its origins to Wyngarden (1927) and Wenzlick (1952). The method was extended by Case and Shiller (1987, 1989) to better account for het-

eroscedasticity.

The basic repeat-sales method estimates the following regression model by ordinary least squares (OLS):

$$\ln p_{th} - \ln p_{sh} = \sum_{\tau=0}^T \beta_{\tau} D_{\tau h} + \varepsilon_h, \quad (1)$$

where  $h$  indexes a particular dwelling,  $p_{sh}$  and  $p_{th}$  denote the prices at which dwelling  $h$  was sold in periods  $s$  and  $t$ ,  $\varepsilon_h$  is an error term, and  $D_{\tau h}$  is a dummy variable that takes the value 1 if the price of dwelling  $h$  is observed for the second time in period  $\tau$ , -1 if the price of dwelling  $h$  is observed for the first time in period  $\tau$ , and zero otherwise. The price indexes  $P_t$  are obtained by exponentiating the estimated parameters, denoted here by  $\hat{\beta}_t$ :

$$P_t = \exp(\hat{\beta}_t).$$

Case and Shiller (1987, 1989) argue that the change in house prices includes components whose variances increase with the interval between sales. They estimate this heteroscedastic variance by regressing the square of the ordinary least squares (OLS) error on a constant and the time interval between sales. Calhoun (1996), however, argues that the heteroscedastic variance can be expected to be non-linear in time intervals. Hence he proposes estimating the square of the error as a function of a constant, the time interval and the square of the time interval. The main extension of Calhoun (1996) therefore is the inclusion of a quadratic term in the time intervals when estimating the variance of the error.

The Calhoun (1996) method, which is used by the Office of Federal Housing Oversight (OFHEO) in the US, uses the OLS estimates  $\hat{\beta}_t$  from (1) to predict the price in period  $t$  of a dwelling with a transaction price  $p_{sh}$  in period  $s$  as follows:

$$\ln \hat{p}_{th} = \ln p_{sh} + \hat{\beta}_t - \hat{\beta}_s.$$

These predicted values can in turn be used to calculate squared deviations of observed house prices around the estimated market index:

$$d_h^2 = [\ln p_{th} - \ln \hat{p}_{th}]^2 = [\ln p_{th} - \ln p_{sh} - \hat{\beta}_t + \hat{\beta}_s]^2.$$

These squared deviations are then used as the dependent variable in the following regression:

$$d_h^2 = a + b(t - s) + c(t - s)^2. \quad (2)$$

The predicted squared deviations  $\hat{d}_h^2$  obtained from (2), modeled as a function of time between sales and time between sales squared, provide weights that correct for heteroscedasticity in the generalized least squares (GLS) estimation of  $\beta_t$ .

$$\frac{\ln p_{th} - \ln p_{sh}}{\sqrt{\hat{d}_h^2}} = \sum_{\tau=0}^T \beta_\tau \frac{D_{\tau h}}{\sqrt{\hat{d}_h^2}} + \frac{\varepsilon_h}{\sqrt{\hat{d}_h^2}}. \quad (3)$$

The WRS price indexes  $P_t$  are obtained by exponentiating the estimated GLS parameters, denoted here by  $\tilde{\beta}_t$ :

$$P_t = \exp(\tilde{\beta}_t).$$

It can be shown that this index is a biased estimate of the desired population parameter since it entails taking a nonlinear transformation of a random variable. Goetzmann (1992) suggests the following correction:

$$P_t = \exp(\tilde{\beta}_t + \hat{\sigma}_t^2/2),$$

where  $\hat{\sigma}_t^2$  is an estimate of the variance of the house price index (see Calhoun 1996 for further details on how this variance is estimated). Similar types of corrections are proposed for semi-log hedonic models (see below).

The main advantages of the repeat-sales method are that it generates quality-adjusted indexes that are relatively easy to compute, that only require transaction prices and unique dwelling identifiers, and that it allows the provider only limited discretion. Its main disadvantages are that it throws away a lot of data (i.e., the prices of all dwellings that sell only once in the data set), and that one cannot be sure that like is compared with like when comparing the price of the same dwelling at two different points in time. This is because the dwelling may have been renovated, extended, neglected, etc., between the two transaction dates.

These problems are well-known in the literature, and attempts are made to correct for the latter. The S&P/Case-Shiller Home Price Indices for example exclude repeat sales that occur within six months on the grounds that such high frequency repeat sales suggest that the dwelling may have been renovated. Also, repeat sales with a relatively long time interval between sales are given less weight in the index because of the increased likelihood that they have experienced physical changes (see Standard and Poor's 2008).

Two further problems are first that all the results change when a new period is added to the data set (see Clapp and Giacotto 1999, and Clapham, Englund, Quigley and Redfearn

2006). This could cause confusion for users, especially when the index forms the basis of derivatives contracts.

Second and more importantly, the data set may suffer from sample selection bias which may in turn cause bias in the index. Clapp and Giaccotto (1992) argue that a repeat-sales sample has a “lemons” bias, since starter homes sell more frequently as a result of people upgrading as their wealth rises. This lemons bias has also been documented by Meese and Wallace (1997), and Steele and Goy (1997). Suppose further that better quality dwellings rise in price on average at a differing rate than worse quality dwellings. In this case, a repeat sales index may be biased. This seems to be the situation observed by Hill, Melser and Syed (2009) in their data set for Sydney, Australia over the period 2001-2006. Furthermore, Shimizu, Nishimura and Watanabe (2010) find that due to sample selection bias repeat-sales indexes are up to two years late in identifying the low point in the Tokyo housing market in 2002.

Hedonic methods provide an alternative way of constructing quality-adjusted price indexes. It is to such methods that I now turn.

## **3 A Short History of Hedonic Indexes**

### **3.1 The origins of hedonic indexes**

The hedonic method dates back at least to Waugh (1928). Other early contributors include Court (1939) and Stone (1954). It was, however, only after Griliches (1961, 1971) that hedonic methods started to receive serious attention (see Schultze and Mackie 2002 and Triplett 2004). The conceptual basis of the approach was laid down by Lancaster (1966) and Rosen (1974).

A hedonic model regresses the price of a product on a vector of characteristics (whose prices are not independently observed). The hedonic equation is a reduced form equation that is determined by the interaction of supply and demand. The hedonic method is used for two main purposes. The first purpose is to quality adjust the observed prices on the lefthand side of the hedonic equation so as to allow the construction of a quality-adjusted price index. The second purpose is to obtain estimates of the willing to pay for, or marginal cost of producing, the characteristics on the righthand side of the hedonic equation, or even

of the underlying demand or supply functions of these characteristics and corresponding consumer or producer surpluses. For example, hedonic models are widely used in the environmental field to value local amenities and the damage caused by pollution (see for example Bayer, Keohane and Timmins 2009). However, the derivation of demand or supply functions is typically problematic since the observed prices and characteristic quantities are determined jointly by supply and demand (see Rosen 1974). Pakes (2003) goes further and argues that interpreting the characteristic shadow prices obtained from a hedonic regression as marginal costs is also problematic since it ignores markups charged by producers.

Fortunately, none of these problems are relevant if our sole purpose in using a hedonic model is to construct a quality-adjusted price index. The majority of research in this field has focused on products subject to rapid technological change, such as computers (see for example Dulberger 1989, and Berndt, Griliches and Rappaport 1995) and cars (see for example Triplett 1969, and Arguea and Hsiao 1993). Some early applications were agricultural in focus, including Waugh's 1928 study on asparagus. More recently, the research focus has broadened to include the healthcare industry (see Cockburn and Anis 2001, and Berndt et al. 2002), the housing market (see below), and artwork (see Chanel, Grard-Varet and Ginsburgh 1996, and Collins, Scorcu and Zanola 2009).

The US has taken the lead in incorporating the research findings on quality adjustment emerging from the hedonic literature into its official statistics. The Bureau of Economic Analysis (BEA) in 1985 started producing quality-adjusted price indexes for computers and peripheral equipment for use in its national accounts (see Cole et al. 1986, and Cartwright 1986). In the next few years BEA also developed hedonic indexes for multifamily residential structures (de Leeuw 1993), semiconductor chips, digital telephone switching equipment, spreadsheets and word processing programs, and photocopying equipment (see Wasshausen and Moulton 2006).

In parallel, the Bureau of Labor Statistics (BLS) in 1988 started including hedonic price indexes in the consumer price index. This process received a large stimulus as a result of the report of the Boskin Commission (see Boskin et al. 1996), which claimed that the US consumer price index (CPI) has an upward bias of about 1.1 percentage points per year. A significant component of this is quality change bias. Given the large potential implications of such a bias for the government budget (due to the significant number of payments that are indexed to the CPI) the BLS came under immediate pressure to fix the bias as quickly as possible. In most cases, the hedonic approach offers the best chance of eliminating

quality change bias. The BLS now uses hedonic methods to quality adjust a number of commodity classes in the CPI including clothing, rental housing, owners' rental equivalence, computers, televisions, audio equipment, VCRs, DVD players, camcorders, refrigerators, washers, dryers, microwave ovens, and college textbooks (see Johnson, Reed and Stewart 2006). Other countries have lagged behind the US in this process, although some are showing signs of gradually moving in the same direction (see Ahnert and Kenny 2004). One unfortunate consequence of this mismatch is that it has the potential to undermine the international comparability of US GDP growth rates and inflation.

A further concern among some experts in the field is the inherent complexity of some of the hedonic adjustment processes and the perceived risk that a too rapid expansion of hedonic methods could lead to mistakes that erode confidence among users (see Schultze and Mackie 2002 and von Auer and Brennan 2007).

### **3.2 Applications in a housing context**

Housing is an extreme case of a differentiated product in the sense that every house is different. Also, unlike computers or cars, one of the most important characteristics of a house is its location. More generally, one can distinguish between physical and locational attributes. Examples of the former include the number of bedrooms and land area, while examples of the latter include the exact longitude and latitude of a house, and the distance to local amenities such as a shopping center, park or school. Hedonic regressions for housing also typically suffer from a severe omitted variables problem, both in terms of the physical and locational characteristics.

Hedonic methods are used for three main purposes in a housing context. First and foremost in our context, hedonic methods are used to construct quality-adjusted house price indexes.<sup>3</sup> Second, they are used to provide automated valuations (or general appraisals) of properties. Third, they are used to explain variations in house prices or to determine the impact on house prices of certain characteristics, such as environmental bads such as pollution (see Kiel and Zabel 2000, and McMillen 2004) or goods such as public parks (see Song and Knaap 2004, and Rouwendal and van der Straaten 2008), local taxes and public school provision (see Oates 1969, and Gibbons and Machin 2003) and crime (see Gibbons

---

<sup>3</sup>The main focus here is on house prices. However, the same methods can be applied to the rental market.

2004, and Naroff, Hellman and Skinner 2006). Coulson (2008; Chapter 3) provides a useful survey of this literature.

The first applications of hedonic methods to the housing market tended to address the last of these objectives. Perhaps the first such study was Ridker and Henning (1967), which focused on air pollution. Research in the field began in earnest in the 1970s and 1980s. Notable early contributions include Oates 1969, Kain and Quigley (1970), Berry and Bednarz (1975), Gillingham (1975), Chinloy (1977), Ferri (1977), Maclennan (1977), Goodman (1978), Follain and Malpezzi (1980), Linneman (1980), Palmquist (1980), and Halvorsen and Pollakowski (1981). These early contributions typically only had access to rather limited data sets and computing power. As a result of the combination of the development of new data sets, increased computing power, and the growing recognition of the economic importance of the housing sector, this field has since become a very active area of research. For example, computer intensive nonparametric methods are becoming increasingly popular (see Pace 1993, and Wallace 1996). Also, it is now standard to make use of the longitude and latitude data of individual properties in hedonic regressions. This is usually achieved by constructing a matrix of distances between all properties in the data set and then using methods developed in the spatial econometrics literature (see for example Anselin 1988) to allow for spatial dependence in the estimated model. Explicitly accounting for spatial dependence should ameliorate the omitted (locational) variables problem.

Probably the first maintained hedonic house price index was produced by the US Census Bureau in 1968 (see Triplett 2004). In the UK, the Halifax house price index and the Nationwide index both date back to the 1980s. More recently, a third UK hedonic index – the Communities and Local Government (CLG) index – was developed by the Office of National Statistics (ONS) (see Acadametrics 2009 for a discussion of the various UK indexes). In Ireland, the permanent tsb index calculated using the same methodology as the UK’s Halifax index dates back to 1996 (see Duffy 2009). Conseil Supérieur du Notariat (CSN) and INSEE (the national statistical office of France) compute hedonic indexes for regions in France since 1998 (see Gouriéroux C. and A. Laferrère 2009). Also, hedonic indexes are computed by Statistics Finland (see Saarnio 2006), Statistics Norway (see Thomassen 2007), Statistics Sweden (see Ribe 2009), the Statistical Office of the Republic of Slovenia on an experimental basis (see Pavlin 2006), RPData-Rismark for Australia, Informations und Ausbildungszentrum für Immobilien für Zurich, Switzerland (the Zürcher Wohneigentumsindex ZWEX index – see Syz, Vanini and Salvi 2008), Verband Deutscher Pfandbriefbanken (VDP) and Hypoport AG for Germany (although these indexes lack



transparency), and Recruit IPD Japan, the Japan Real Estate Institute and the Japan Research Institute (these indexes likewise lack transparency). According to Hoffman and Lorenz (2006), the German Federal Statistical Office is also in the process of developing hedonic indices.

## 4 A Taxonomy of Hedonic Price Indexes for Housing

### 4.1 Time-dummy methods

#### 4.1.1 Description of the method

The hedonic approach can be implemented in a number of different ways. To see the underlying structure of these methods and how they relate to each other it is useful to divide them into taxonomic groups. The taxonomy that follows where possible uses the terminology laid down by Triplett (2004).

The time-dummy method is the original hedonic method. It typically uses the semi-log functional form. A standard semi-log formulation is as follows:<sup>4</sup>

$$y = Z\beta + D\delta + \varepsilon, \quad (4)$$

where  $y$  is an  $H \times 1$  vector with elements  $y_h = \ln p_h$ ,  $Z$  is an  $H \times C$  matrix of characteristics (some of which may be dummy variables),  $\beta$  is a  $C \times 1$  vector of characteristic shadow prices,  $D$  is an  $H \times T - 1$  matrix of period dummy variables,  $\delta$  is a  $T - 1 \times 1$  vector of period prices, and  $\varepsilon$  is an  $H \times 1$  vector of random errors.<sup>5</sup> Finally,  $H$ ,  $C$  and  $T$  denote respectively the number of dwelling, characteristics and time periods in the data set. The first column in  $Z$  consists of ones, and hence the first element of  $\beta$  is an intercept. Examples of characteristics include the number of bedrooms, number of bathrooms, land area, and dwelling type of a property (house or unit). It is possible also to include functions of characteristics (such as land size squared), and interaction terms between characteristics. For example, one might want to interact bedrooms and land area, bathrooms and land area, and bedrooms and bathrooms. Focusing specifically on the last of these, the inclusion of bedroom-bathroom

---

<sup>4</sup>See Diewert (2003) and Malpezzi (2003) for a discussion of some of the advantages of the semi-log model in this context.

<sup>5</sup>The base period price index is normalized to 1.

interaction terms could be justified by the fact that the value of an extra bathroom may depend on how many bedrooms there are.

When the objective of the exercise is to construct a quality-adjusted price index, the primary interest lies in the  $\delta$  parameters which measure the period specific fixed effects on the logarithms of the price level after controlling for the effects of the differences in the attributes of the dwellings. One attraction of the semi-log time-dummy model is that the price index  $P_t$  for period  $t$  is derived by simply exponentiating the estimated  $\delta_t$  coefficient obtained from the hedonic model:

$$\hat{P}_t = \exp(\hat{\delta}_t). \quad (5)$$

Consider for example the semi-log time-dummy regression equation in Table 1.

**TABLE 1: A Semi-Log Time-Dummy Regression**

Variable	Estimate	Standard Error	t statistic
Intercept	12.908	0.00619	2086.34
Bed1	-0.00434	0.02045	-0.21
Bed3	0.00688	0.00802	0.86
Bed4	0.09000	0.00604	14.91
Bed5	0.11762	0.00505	23.27
Bath2	0.36727	0.00382	96.20
Bath3	0.63106	0.00583	108.30
Bath4	0.95663	0.01181	81.01
LandArea	0.03217	0.00523	6.15
$\delta_{02}$	0.1571	0.00570	27.57
$\delta_{03}$	0.2752	0.00563	48.88
$\delta_{04}$	0.3421	0.00592	57.77
$\delta_{05}$	0.2697	0.00536	50.34
$\delta_{06}$	0.2518	0.00518	48.56

This regression, which has been slightly modified for expositional purposes, was calculated using an Australian Property Monitors data set consisting of 100,887 houses, covering the period 2001 to 2006 for Sydney, Australia.

Bed1 is a dummy variable that takes the value 1 when a dwelling has one bedroom and zero otherwise, Bath2 takes the value 1 when a dwelling has two bathrooms and zero

otherwise, etc. Landarea is measured in square meters divided by 1000. The bedroom coefficients are defined relative to a two bedroom dwelling. This explains why the Bed1 coefficient is negative and the Bed2 coefficient is omitted. The estimated coefficient on Bed4 for example can be interpreted as follows:  $\exp(0.09) = 1.094$ , which implies that increasing the number of bedrooms from 2 to 4 acts to increase price by on average 9.4 percent, other things equal.

The price indexes for the six years are obtained by exponentiating the estimated  $\delta_t$  coefficients from Table 1 as follows:

$$\begin{aligned} P_1 &= \exp(\delta_{01}) = \exp(0.0000) = 1.0000 \\ P_2 &= \exp(\delta_{02}) = \exp(0.1571) = 1.1701 \\ P_3 &= \exp(\delta_{03}) = \exp(0.2752) = 1.3168 \\ P_4 &= \exp(\delta_{04}) = \exp(0.3421) = 1.4078 \\ P_5 &= \exp(\delta_{05}) = \exp(0.2697) = 1.3096 \\ P_6 &= \exp(\delta_{06}) = \exp(0.2518) = 1.2863 \end{aligned}$$

In other words, prices in 2002 are 17 percent higher than in 2001, etc.

Box-Cox tests of the semi-log model almost invariably reject the semi-log model (see for example Halvorsen and Pollakowski 1981). One problem with using the Box-Cox method to choose the functional form is that the derivation of price indexes directly from the hedonic equation may become far from straightforward.

Brushing concerns over the semi-log model aside, it can be shown that the price index  $P_t$  derived from (5) is a biased estimate of the desired population parameter since it entails taking a nonlinear transformation of a random variable (see Goldberger 1968, Kennedy 1981, Giles 1982 and Triplett 2004). More precisely, we have that  $E[\exp(\hat{\delta})] \neq \exp(\delta)$ . The standard correction for this bias is to add half the variance of the estimated coefficient when calculating the price index. The corrected price index for period  $t$ , denoted here by  $P_t^*$ , is therefore calculated as follows:

$$P_t^* = \exp \left[ \hat{\delta}_t + \frac{1}{2} \hat{\sigma}_t^2 \right],$$

where  $\hat{\sigma}_t^2$  is an estimate of the variance of  $\hat{\delta}_t$ .<sup>6</sup>

---

<sup>6</sup>See van Garderen and Shah (2002) for a discussion of ways of estimating  $\hat{\sigma}_t^2$ .

There is, however, a problem with this correction that has thus far been ignored in the literature. By construction, no correction is made to the price index of the base period which is still normalized to 1. The correction therefore acts to systematically increase the price index of every other period relative to that of the base period. Given that in the absence of this correction the time-dummy method is base-period invariant (subject to rescaling), the imposition of the correction therefore causes a violation of base-period invariance. The fact that the original time-dummy method is base-period invariant implies that the uncorrected indexes cannot be systematically biased relative to the base period. It follows that the corrected indexes, while correcting another source of bias, generate indexes that are systematically biased for each period relative to the base period. One resolution to this dilemma is to generate  $T$  sets of time-dummy results using each of the  $T$  periods in turn as the base period, then make the bias correction discussed above, and finally take a geometric average of the  $T$  sets of results. Such an approach, which as far as I know has never been tried, while more laborious will correct for both types of bias.

Syed, Hill and Melser (2008) investigate the magnitude of the difference between  $P_t$  and  $P_t^*$  for a single base period using data for Sydney, Australia over the period 2001-2006. They find that the difference manifests itself typically only in the third or fourth decimal place of the price indexes and hence the bias can in most cases be ignored. It remains to be seen whether the same is true for other housing data sets.

One of the key determinants of house prices is location. The explanatory power of the hedonic model can therefore be significantly improved by exploiting information on the location of each property. Probably the simplest way to do this is to include postcode identifiers for each dwelling in the hedonic model.<sup>7</sup> These postcode identifiers can take the form of dummy variables. In the case of the Sydney data set, the inclusion of postcode dummies acts to increase the R-squared coefficient from about 0.56 to 0.76 (see Hill, Melser and Syed 2009). The hedonic model now takes the following form:

$$y = Z\beta + B\gamma + D\delta + \varepsilon, \tag{6}$$

where the additional term  $B$  is an  $H \times (M-1)$  matrix of dummy variables,  $\gamma$  is an  $(M-1) \times 1$  vector of parameters, and  $M$  is the number of postcode identifiers.

---

<sup>7</sup>When longitude and latitude data are available for each dwelling, spatial dependence can be modeled in a more rigorous way (see below).

### 4.1.2 Strengths and weaknesses

Perhaps the main strength of the semi-log time-dummy method is its simplicity. The price indexes are derived straight from the estimated hedonic equation. It also generates standard errors on the price indexes.

It could be argued, however, that the time-dummy method lacks flexibility. The period dummies and dwelling characteristics enter the hedonic function additively. In other words, the function exerts quite severe restrictions on the potential interactions between periods and characteristics. One way to increase the flexibility of the model is by allowing the periods to interact with the characteristics. This approach however requires the estimation of more parameters reducing the number of degrees of freedom. It also makes the derivation of the price indexes from the estimated hedonic model more complicated (see for example Syed, Hill and Melser 2008).

To see more clearly some of the implications of fixing the reference characteristic shadow prices it is useful to write out the price index formula implied by the time-dummy model in (4).<sup>8</sup> The estimated version of the time-dummy hedonic model can be written as follows:

$$\hat{y} = Z\hat{\beta} + D\hat{\delta},$$

The estimated shadow price vector and price index vector are derived as follows:

$$\hat{\beta} = (Z^T Z)^{-1} Z^T (y - D\hat{\delta}), \quad (7)$$

$$\hat{\delta} = (D^T D)^{-1} D^T (y - Z\hat{\beta}). \quad (8)$$

It turns out that  $D^T D$  in (8) is a diagonal matrix. As a result, the price index formula for a particular element  $\hat{\delta}_t$  of  $\hat{\delta}$  reduces to the following:

$$\hat{\delta}_t = \sum_{h=1}^{H_t} \left( \frac{\ln P_{th}}{H_t} \right) - \sum_{c=1}^C \left[ \hat{\beta}_c \left( \frac{\sum_{h=1}^{H_t} z_{cth}}{H_t} \right) \right].$$

Taking exponents of both sides, we obtain the following price index in a comparison between periods  $s$  and  $t$ :<sup>9</sup>

$$\frac{P_t}{P_s} = \frac{\left( \prod_{h=1}^{H_t} p_{th} \right)^{1/H_t}}{\left( \prod_{h=1}^{H_s} p_{sh} \right)^{1/H_s}} \bigg/ \frac{\exp \left( \sum_{c=1}^C \hat{\beta}_c \bar{z}_{ct} \right)}{\exp \left( \sum_{c=1}^C \hat{\beta}_c \bar{z}_{cs} \right)}, \quad (9)$$

<sup>8</sup>This section draws extensively on the findings of Hill and Melser (2009).

<sup>9</sup>This formula can be found in Triplett (2004; p. 51), Diewert, Silver and Heravi (2007; p. 7) and Hill and Melser (2009).

where

$$\bar{z}_{cs} = \sum_{h=1}^{H_s} z_{csh}/H_s, \quad \bar{z}_{ct} = \sum_{h=1}^{H_t} z_{cth}/H_t.$$

The term  $(\prod_{h=1}^{H_t} p_{th})^{1/H_t}/(\prod_{h=1}^{H_s} p_{sh})^{1/H_s}$  in the numerator of (9) compares the average price of a house in the two periods. The quality adjustment is provided by the term  $\exp(\sum_{c=1}^C \hat{\beta}_c \bar{z}_{ct})/\exp(\sum_{c=1}^C \hat{\beta}_c \bar{z}_{cs})$  in the denominator of (9). This is a quantity index that compares the price of the average house in the two periods using the time-dummy average characteristic prices.

The denominator in (9) can now be rewritten as follows:

$$\frac{\exp\left(\sum_{c=1}^C \hat{\beta}_c \bar{z}_{ct}\right)}{\exp\left(\sum_{c=1}^C \hat{\beta}_c \bar{z}_{cs}\right)} = \left[ \frac{\exp\left(\sum_{c=1}^C \hat{\beta}_c \bar{z}_{ct}\right)}{\exp\left(\sum_{c=1}^C \hat{\beta}_c \bar{z}_c\right)} \right] \bigg/ \left[ \frac{\exp\left(\sum_{c=1}^C \hat{\beta}_c \bar{z}_{cs}\right)}{\exp\left(\sum_{c=1}^C \hat{\beta}_c \bar{z}_c\right)} \right] = \frac{\tilde{Q}_{X,t}^L}{\tilde{Q}_{X,s}^L}, \quad (10)$$

where  $\tilde{Q}_{X,t}^L$  denotes a Laspeyres-type quantity index that compares the cost of the average house sold in period  $t$  to that sold in the average period  $X$ . Substituting (10) into (9), we obtain that

$$\frac{P_t}{P_s} = \frac{\left(\prod_{h=1}^{H_t} p_{th}\right)^{1/H_t}}{\left(\prod_{h=1}^{H_s} p_{sh}\right)^{1/H_s}} \bigg/ \frac{\tilde{Q}_{X,t}^L}{\tilde{Q}_{X,s}^L}. \quad (11)$$

The fact that the time-dummy method uses a Laspeyres-type quantity index to make its quality adjustments raises the possibility that there may be a problem of substitution bias. This would not be substitution bias quite in the normal sense, since the average household does not buy houses every period. Nevertheless, something mathematically akin to substitution bias may occur. For this to happen, however, requires a triple coincidence of events. First, the shadow prices of the characteristics need to change over time. This is almost certainly the case. For example, Hill and Melsler (2009) are able to clearly reject the null hypothesis of constant shadow prices. Second, even if the shadow prices change over time, they also need to interact with changes in the observed quantities of these same characteristics in systematic ways (i.e., characteristics that have risen in price the most in percentage terms experience the biggest percentage falls or rises in quantities). Third, substitution bias in the price indexes  $P_s$  and  $P_t$  does not necessarily imply bias in the ratio  $P_t/P_s$ , which is what we really care about. To put it another way, what matters is not the bias in each period's price index relative to the artificial average period  $X$ , but the bias in the price indexes relative to each other. If all the price indexes  $P_t$  are say too large relative to the average period  $X$ , then these biases will at least partially offset each

other. This, however, is unlikely. The more different a period  $t$ 's estimated characteristic shadow price vector  $\hat{\beta}_t$  is from the average estimated characteristic shadow price vector  $\beta$  (measured using an appropriate dissimilarity metric) the larger the substitution bias for period  $t$  is likely to be (assuming some kind of systematic difference exists between changes in prices and quantities).<sup>10</sup> There is no reason to expect all the period shadow prices to be equidistant from the average shadow price vector once there are three or more periods in the comparison. Ultimately, the issue of whether substitution bias is a problem for the time-dummy method must be settled empirically. Hill and Melser (2009) have begun this process. Preliminary results suggest that this problem seems to manifest itself more in spatial rather than temporal comparisons.

A second concern with the time-dummy method is that when a new period is added to a data set, the price indexes for all periods change (i.e., the property of temporal fixity as discussed in Hill 2004 is violated). This indeed is one of the same criticisms leveled against the repeat-sales method.

One simple way of avoiding both these problems is to estimate the time-dummy model only over adjacent periods. Triplett (2004) refers to this method as the adjacent-period method. The overall price index is then constructed by chaining together the results obtained between adjacent period. This deals with the substitution bias problem since by construction the shadow price vectors of the two periods should be approximately equidistant from the average shadow price vector. Temporal fixity is also satisfied, since a new period when added to the data set is only compared directly with the period preceding it.

### 4.1.3 Usage

Although it is the original hedonic method and is widely used in other contexts, such as for constructing quality-adjusted price indexes for computers and cars, the time-dummy method has not been used that much in a housing context. Some applications of the method can be found in the academic literature. Examples include Follain and Malpezzi (1980), Palmquist (1980), Mark and Goldberg (1984), and Haughwout, Orr and Bedoll (2008). However, it has generally been eschewed by index providers. One exception is the Statistical Office of the Republic of Slovenia which has used it to construct an experimental house price index over the period 2003-2006. An important reason for its lack of popularity

---

<sup>10</sup>See Diewert (2002) for a discussion of dissimilarity metrics.

is its violation of temporal fixity, which means that all results need to be recomputed when a new period is added to the data set (except when used in the adjacent-period form). This is not in itself necessarily an insurmountable problem, as has been demonstrated by maintained repeat-sales indexes such as the Standard and Poor's/Case-Shiller (SPCS) Home Price Indexes which suffer from the same weakness. The adjacent-period variant on the time-dummy method is more popular. Both RPData-Rismark (for Australian cities) and Informations und Ausbildungszentrum für Immobilien (for Zurich in Switzerland) use this method.<sup>11</sup>

## 4.2 Imputation methods

### 4.2.1 Description of the method

Imputation methods make use of standard price index formulas.<sup>12</sup> The two best known of these formulas are Laspeyres and Paasche. Laspeyres and Paasche price index formulas measure the change in the price of a given basket of goods (in our case dwellings) over time. This requires the price of each good in the basket to be available each period. In a housing context, therefore, it is not possible to compute Laspeyres and Paasche price indexes based on actual transaction prices, since each dwelling sells only at infrequent and irregular intervals. However, it may still be possible to compute such indexes if we are willing to replace actual transaction prices with imputed prices. This provides the underlying rationale of imputation methods. The estimated hedonic model is used to impute prices for dwellings. In this way it is possible to ensure that a price is available for both periods for each dwelling included in the price index formula. The price index can then be calculated in the standard way.

The price index formulas I focus on here are Paasche, Laspeyres, Fisher, Geometric-Paasche, Geometric-Laspeyres and Törnqvist. Let  $P_{st}$  denote a price index between periods  $s$  and  $t$ . The price of dwelling  $h$  in period  $t$  is denoted by  $p_{th}$ . In our context, each dwelling is unique and hence the quantity of each dwelling is 1. Hence the formulas look a bit

---

<sup>11</sup>The index methodology underlying the ZWEX is explained in a document downloadable from the following site:

[http://www.zkb.ch/etc/ml/repository/textdokumente/finanzieren/indexmethode\\_zwex.pdf](http://www.zkb.ch/etc/ml/repository/textdokumente/finanzieren/indexmethode_zwex.pdf).File.pdf.

<sup>12</sup>This section draws extensively from material in Hill and Melser (2008).



different from their standard formulations in the price index literature.

$$\text{Paasche : } P_{st}^P = \left\{ \sum_{h=1}^{H_t} w_{th} [p_{th}/p_{sh}]^{-1} \right\}^{-1} = \sum_{h=1}^{H_t} p_{th} / \sum_{h=1}^{H_t} p_{sh} \quad (12)$$

$$\text{Laspeyres : } P_{st}^L = \sum_{h=1}^{H_s} \{w_{sh} [p_{th}/p_{sh}]\} = \sum_{h=1}^{H_s} p_{th} / \sum_{h=1}^{H_s} p_{sh} \quad (13)$$

$$\text{Fisher : } P_{st}^F = \sqrt{P_{st}^P \times P_{st}^L} \quad (14)$$

$$\text{Geometric Paasche : } P_{st}^{GP} = \prod_{h=1}^{H_t} [(p_{th}/p_{sh})^{w_{th}}] \quad (15)$$

$$\text{Geometric Laspeyres : } P_{st}^{GL} = \prod_{h=1}^{H_s} [(p_{th}/p_{sh})^{w_{sh}}] \quad (16)$$

$$\text{Törnqvist : } P_{st}^T = \sqrt{P_{st}^{GP} \times P_{st}^{GL}} \quad (17)$$

The terms  $w_{th}$  are expenditure weights defined as follows:

$$w_{th} = p_{th} / \sum_{m=1}^{H_t} p_{tm},$$

while  $H_s$  and  $H_t$  denote the total number of dwellings sold respectively in periods  $s$  and  $t$ . The exponents in the geometric Paasche and geometric Laspeyres formulas could alternatively be set to  $1/H_t$  and  $1/H_s$  respectively, depending on whether one wants to give equal weight to all price relatives, or give greater weight to the price relatives of more expensive dwellings.

These price index formulas, in general, all give different answers. This is what is meant by the price index problem. It is a problem that has attracted some of the greatest minds in the economics profession over the best part of two centuries, such as Marshall, Edgeworth, Keynes, Fisher, Hicks and Samuelson. Fisher (1922) for example considers in excess of 100 different formulas. The price index problem has been attacked from two main directions, usually referred to as the economic and axiomatic approaches. The economic approach views quantities as utility maximizing responses to prices. This approach has culminated in the work of Diewert (1976), who proposed the concept of a superlative price index (a class of indexes that attain a second order approximation to the underlying cost of living index). Each of the indexes outlined above can be derived from a particular functional form for the cost or utility function. Diewert's contribution was to show that some of the indexes are based upon more flexible representations of the cost function than others. The Fisher and Törnqvist indexes are superlative indexes as they allow for flexible substitution behavior. An alternative approach to justifying the choice of index number formula is the

axiomatic approach, which proposes a series of axioms that a price index should satisfy, and then discriminates between price index formulas on the basis of their performance relative to these axioms (see Eichhorn and Voeller 1976, Balk 1995 and Diewert 2007b). Fortunately, the axiomatic approach also tends to favor the Fisher and Törnqvist indexes as these satisfy what are generally considered to be the most important axioms.

This literature, however, assumes that there is no matching problem. That is, it is assumed that in all periods price and quantity data are available for all commodity headings included in the comparison. Once this assumption is relaxed, the price index problem becomes more complex.

The use of the hedonic imputation method adds a new dimension to the price index problem. This is because we have some discretion as to which prices are imputed. If a product is unavailable in a particular period, we have no choice but to impute it. If the product is available, we may nevertheless still prefer to use an imputed price over the actual price. This might seem counterintuitive. However, it turns out that replacing real prices with imputations can sometimes reduce the omitted variables bias and help ensure that like is compared with like.

It should be noted that imputations methods do not require any particular functional form for the hedonic model. In fact the hedonic model could even be nonparametric. Nonparametric methods are discussed in a later section.

Let  $\hat{p}_{th}(z_{sh})$  denote the estimated price in period  $t$  of a dwelling sold in period  $s$ . This price is imputed by substituting the characteristics of house  $h$  sold in period  $s$  into the estimated hedonic model of period  $t$  as follows:

$$\hat{p}_{th}(z_{sh}) = \exp\left(\sum_{c=1}^C \hat{\beta}_{ct} z_{csh}\right).$$

Again, however, a bias correction is required, since the focus of our attention is  $\hat{p}_{th}(z_{sh})$  and not  $\ln \hat{p}_{th}(z_{sh})$ . The correction now entails adding half the variance of the error term from the hedonic regression equation (see Malpezzi, Chun and Green 1998 and Coulson 2008). The corrected imputed price index denoted here by  $\hat{p}_{th}^*(z_{sh})$  is calculated as follows:

$$\hat{p}_{th}^*(z_{sh}) = \exp\left(\sum_{c=1}^C \hat{\beta}_{ct} z_{csh} + \sigma_t^2/2\right).$$

Malpezzi, Chun and Green (1998) provide comparisons of corrected and uncorrected imputed prices for average dwellings (see the characteristics approach below) and find that

they differ by between 5 and 10 percent. Given the earlier findings of Syed, Hill and Melser (2008) this suggests that the correction may matter more in an imputation setting than in a time-dummy setting.

To illustrate how the hedonic imputation method complicates the price index problem, consider first the case of the Laspeyres price index. Four different varieties of the Laspeyres price index are obtained (denoted here by L1, L2, L3 and L4) depending on how exactly the hedonic imputation method is implemented. For the case of L2 and L4, the  $H_s$  dwellings available in period  $s$  are ordered as follows:  $h = 1, \dots, H_{st}$  indexes the dwellings sold in both periods  $s$  and  $t$ , while  $h = H_{st} + 1, \dots, H_s$  indexes the dwellings sold in period  $s$  but not in period  $t$ .

$$\begin{aligned}
\text{L1 : } P_{st}^{L1} &= \sum_{h=1}^{H_s} \{w_{sh} [\hat{p}_{th}(z_{sh})/p_{sh}]\} = \sum_{h=1}^{H_s} \hat{p}_{th}(z_{sh}) / \sum_{h=1}^{H_s} p_{sh} \\
\text{L2 : } P_{st}^{L2} &= \sum_{h=1}^{H_{st}} [w_{sh} (p_{th}/p_{sh})] + \sum_{h=H_{st}+1}^{H_s} \{w_{sh} [\hat{p}_{th}(z_{sh})/p_{sh}]\} \\
&= \left[ \sum_{h=1}^{H_{st}} p_{th} + \sum_{h=H_{st}+1}^{H_s} \hat{p}_{th}(z_{sh}) \right] / \sum_{h=1}^{H_s} p_{sh} \\
\text{L3 : } P_{st}^{L3} &= \sum_{h=1}^{H_s} \{\hat{w}_{sh} [\hat{p}_{th}(z_{sh})/\hat{p}_{sh}(z_{sh})]\} = \sum_{h=1}^{H_s} \hat{p}_{th}(z_{sh}) / \sum_{h=1}^{H_s} \hat{p}_{sh}(z_{sh}) \\
\text{L4 : } P_{st}^{L4} &= \sum_{h=1}^{H_{st}} [w_{sh} (p_{th}/p_{sh})] + \sum_{h=H_{st}+1}^{H_s} \{\hat{w}_{sh} [\hat{p}_{th}(z_{sh})/\hat{p}_{sh}(z_{sh})]\} \\
&= \left[ \sum_{h=1}^{H_{st}} p_{th} + \sum_{h=H_{st}+1}^{H_s} \hat{p}_{th}(z_{sh}) \right] / \left[ \sum_{h=1}^{H_{st}} p_{sh} + \sum_{h=H_{st}+1}^{H_s} \hat{p}_{sh}(z_{sh}) \right]
\end{aligned}$$

L2 uses the minimum number of imputations. The other three varieties (to differing extents) sometimes throw away real price observations and replace them with imputed prices. More specifically, L1 inputs all prices in period  $t$ . L2 only imputes prices in period  $t$  for dwellings that did not sell in both periods  $s$  and  $t$ . L3 imputes all prices for both periods  $s$  and  $t$ , while L4 imputes all prices except for dwellings that sell in both periods. Also, L3 imputes all expenditure shares and L4 does likewise except for repeat sales. The imputed expenditure shares in L3 and L4 are calculated as follows:

$$\hat{w}_{sh} = \hat{p}_{sh}(z_{sh}) / \sum_{m=1}^{H_s} \hat{p}_{sm}(z_{sm}).$$

In the hedonic literature L1 is referred to as a single imputation method, and L3 as a double imputation method (see Silver and Heravi 2001, Pakes 2003, de Haan 2004a, and

Hill and Melser 2008). Actually, in the literature it is typically not made clear whether or not the double imputation method imputes expenditure shares as well. Hence we can distinguish between two double imputation methods, one that imputes expenditure shares and one that does not.

Varieties of Paasche, Fisher, geometric Paasche, geometric Laspeyres and Törnqvist can be derived in an analogous manner. Here we illustrate this point for the first and third varieties only.

$$\begin{aligned}
\text{P1 : } P_{st}^{P1} &= \left\{ \sum_{h=1}^{H_t} w_{th} [p_{th}/\hat{p}_{sh}(z_{th})]^{-1} \right\}^{-1} = \sum_{h=1}^{H_t} p_{th} / \sum_{h=1}^{H_t} \hat{p}_{sh}(z_{th}) \\
\text{F1 : } P_{st}^{F1} &= \sqrt{P_{st}^{P1} \times P_{st}^{L1}} \\
\text{GP1 : } P_{st}^{GP1} &= \prod_{h=1}^{H_t} [(p_{th}/\hat{p}_{sh}(z_{th}))^{w_{th}}] \\
\text{GL1 : } P_{st}^{GL1} &= \prod_{h=1}^{H_s} [(\hat{p}_{th}(z_{sh})/p_{sh})^{w_{sh}}] \\
\text{T1 : } P_{st}^{T1} &= \sqrt{P_{st}^{GP1} \times P_{st}^{GL1}} \\
&= \sqrt{\prod_{h=1}^{H_t} [(p_{th}/\hat{p}_{sh}(z_{th}))^{w_{th}}] \times \prod_{h=1}^{H_s} [(\hat{p}_{th}(z_{sh})/p_{sh})^{w_{sh}}]} \\
\text{GP1' : } P_{st}^{GP1'} &= \prod_{h=1}^{H_t} [(p_{th}/\hat{p}_{sh}(z_{th}))^{1/H_t}] \\
\text{GL1' : } P_{st}^{GL1'} &= \prod_{h=1}^{H_s} [(\hat{p}_{th}(z_{sh})/p_{sh})^{1/H_s}] \\
\text{T1' : } P_{st}^{T1'} &= \sqrt{P_{st}^{GP1'} \times P_{st}^{GL1'}} \\
&= \sqrt{\prod_{h=1}^{H_t} [(p_{th}/\hat{p}_{sh}(z_{th}))^{1/H_t}] \times \prod_{h=1}^{H_s} [(\hat{p}_{th}(z_{sh})/p_{sh})^{1/H_s}]} \\
\text{P3 : } P_{st}^{P3} &= \left\{ \sum_{h=1}^{H_t} \hat{w}_{th} [\hat{p}_{th}(z_{th})/\hat{p}_{sh}(z_{th})]^{-1} \right\}^{-1} = \sum_{h=1}^{H_t} \hat{p}_{th}(z_{th}) / \sum_{h=1}^{H_t} \hat{p}_{sh}(z_{th}) \\
\text{F3 : } P_{st}^{F3} &= \sqrt{P_{st}^{P3} \times P_{st}^{L3}} \\
\text{GP3 : } P_{st}^{GP3} &= \prod_{h=1}^{H_t} [(\hat{p}_{th}(z_{th})/\hat{p}_{sh}(z_{th}))^{\hat{w}_{th}}] \\
\text{GL3 : } P_{st}^{GL3} &= \prod_{h=1}^{H_s} [(\hat{p}_{th}(z_{sh})/\hat{p}_{sh}(z_{sh}))^{\hat{w}_{sh}}] \\
\text{T3 : } P_{st}^{T3} &= \sqrt{P_{st}^{GP3} \times P_{st}^{GL3}}
\end{aligned}$$

$$\begin{aligned}
&= \sqrt{\prod_{h=1}^{H_t} \left[ (\hat{p}_{th}(z_{th})/\hat{p}_{sh}(z_{th}))^{\hat{w}_{th}} \right] \times \prod_{h=1}^{H_s} \left[ (\hat{p}_{th}(z_{sh})/\hat{p}_{sh}(z_{sh}))^{\hat{w}_{sh}} \right]} \\
\text{GP3}' : P_{st}^{GP3} &= \prod_{h=1}^{H_t} \left[ (\hat{p}_{th}(z_{th})/\hat{p}_{sh}(z_{th}))^{1/H_t} \right] \\
\text{GL3}' : P_{st}^{GL3} &= \prod_{h=1}^{H_s} \left[ (\hat{p}_{th}(z_{sh})/\hat{p}_{sh}(z_{sh}))^{1/H_s} \right] \\
\text{T3}' : P_{st}^{T3} &= \sqrt{P_{st}^{GP3} \times P_{st}^{GL3}} \\
&= \sqrt{\prod_{h=1}^{H_t} \left[ (\hat{p}_{th}(z_{th})/\hat{p}_{sh}(z_{th}))^{1/H_t} \right] \times \prod_{h=1}^{H_s} \left[ (\hat{p}_{th}(z_{sh})/\hat{p}_{sh}(z_{sh}))^{1/H_s} \right]}
\end{aligned}$$

For geometric Paasche, geometric Laspeyres and Törnqvist an alternative to weighting by either actual or imputed expenditure shares is to give all dwellings equal weight. This is the approach followed by  $GP1'$ ,  $GL1'$ ,  $T1'$ ,  $GP3'$ ,  $GL3'$  and  $T3'$ . These indexes could also be viewed as variants on the Jevons price index formula. In principle, equal weighting could also be applied to Laspeyres, Paasche, and Fisher. Laspeyres in this case reduces to a variant on the Carli price index formula, and Paasche to the harmonic mean equivalent of Carli. However, if equal weighting is desired, it is probably preferable to use Jevons since it has superior axiomatic properties.

Which variety is best? To simplify matters this question will be addressed for the case of the Laspeyres index. The arguments carry forward equally well to other price index formulas. Our focus is on minimizing omitted variables bias and generating economically meaningful results.

The four varieties of Laspeyres indexes differ in their treatment of the price relatives and expenditure shares. Focusing specifically on the price relatives, a single imputation Laspeyres index uses the price relatives  $\hat{p}_{th}(z_{sh})/p_{sh}$ , while a double imputation index uses  $\hat{p}_{th}(z_{sh})/\hat{p}_{sh}(z_{sh})$ . There has been some debate in the literature on which approach is best. The discussion focuses primarily on the case of computers. Silver and Heravi (2001), Pakes (2003), de Haan (2004a) and Hill and Melser (2008) all argue in favor of double imputation on the grounds that it can reduce omitted variables bias.

For example, consider the case of a dwelling for which  $\hat{p}_{sh}(z_{sh}) > p_{sh}$ . This means either that the buyer got a bargain or that the dwelling performs poorly on its omitted variables. Assuming that the latter is correct, it follows that  $\hat{p}_{th}(z_{sh})$  will overstate the true price of a house with characteristics vector  $z_{sh}$  in period  $t$ . It follows that the price relative  $\hat{p}_{th}(z_{sh})/p_{sh}$  will have an upward bias. In contrast, the biases in  $\hat{p}_{sh}(z_{sh})$  and  $\hat{p}_{th}(z_{sh})$  will

partially offset each other in the price relative  $\hat{p}_{th}(z_{sh})/\hat{p}_{sh}(z_{sh})$ , thus tending to generate a more accurate overall estimate.

The use of double imputation is particularly beneficial in cases such as housing where there is likely to be a serious omitted variables problem. This leads us to prefer varieties 3 and 4 over varieties 1 and 2. Even using double imputation, omitted variables bias will still be a problem when either the average quantities of some of the omitted characteristics or their shadow prices change over time (see Benkard and Bajari 2005). For example, if the overall quality of the omitted characteristics is improving over time, or their shadow prices are rising, the double imputed price relatives  $\hat{p}_{th}(z_{sh})/\hat{p}_{sh}(z_{sh})$  will tend to be too large. In a housing context, it is not clear how big a problem this is. Certainly, in comparisons between chronologically adjacent periods omitted variables should be less of a problem than in comparisons over longer time horizons.

The difference between varieties 3 and 4 is that 4 only uses double imputation when a dwelling is not available in both periods. Variety 3 by contrast always uses double imputation, even when there are no missing prices. For the case of computers this would be hard to justify, since a particular model is the same irrespective of when it is sold. Housing, however, is another matter. There is no guarantee even for a repeat sale that we are comparing like with like since the characteristics (observed and unobserved) of a dwelling may change over time due to renovations. This is particularly a risk here since the hedonic imputation method is typically used to compare adjacent periods, and dwellings that sell in two consecutive periods are more likely to have been renovated. The best way of ensuring that like is compared with like is to double impute prices for all dwelling (even repeat sales). Hence, for the case of housing I recommend using variety 3.

There remains the question of which price index formula should be used? The choice of formula should be between Fisher and Törnqvist since these formulas treat both periods symmetrically, are superlative, and have desirable economic and axiomatic properties (see Diewert 1976, 2007b and Balk 1995). From an economic and axiomatic perspective neither Fisher nor Törnqvist clearly dominates the other. When used in conjunction with the hedonic imputation method, we must however also consider the functional form of the hedonic regression.

I recommend using T3 or T3'. Following Hill and Melsner (2008), for the semi-log case,

GL3' can be re-expressed as follows:

$$\begin{aligned}
P_{st}^{GL3'} &= \prod_{h=1}^{H_s} \left\{ [\hat{p}_{th}(z_{sh})/\hat{p}_{sh}(z_{sh})]^{1/H_s} \right\} \\
&= \prod_{h=1}^{H_s} \left\{ \exp \left[ \frac{1}{H_s} \sum_{c=1}^C (\hat{\beta}_{ct} - \hat{\beta}_{cs}) z_{csh} \right] \right\} \\
&= \exp \left[ \frac{1}{H_s} \sum_{h=1}^{H_s} \sum_{c=1}^C (\hat{\beta}_{ct} - \hat{\beta}_{cs}) z_{csh} \right] \\
&= \exp \left[ \frac{1}{H_s} \sum_{c=1}^C (\hat{\beta}_{ct} - \hat{\beta}_{cs}) \sum_{h=1}^{H_s} z_{csh} \right] \\
&= \exp \left[ \sum_{c=1}^C (\hat{\beta}_{ct} - \hat{\beta}_{cs}) \bar{z}_{cs} \right] \\
&= \prod_{c=1}^C \exp \left[ (\hat{\beta}_{ct} - \hat{\beta}_{cs}) \bar{z}_{cs} \right],
\end{aligned}$$

where we have defined

$$\bar{z}_{cs} = \frac{1}{H_s} \sum_{h=1}^{H_s} z_{csh}.$$

Rearranging GP3' in a similar manner we obtain that

$$P_{st}^{GP3'} = \exp \left[ \sum_{c=1}^C (\hat{\beta}_{ct} - \hat{\beta}_{cs}) \bar{z}_{ct} \right].$$

Taking the geometric mean of GP3' and GL3', we obtain the following expression for T3':

$$P_{st}^{T3'} = \exp \left[ \frac{1}{2} \sum_{c=1}^C (\hat{\beta}_{ct} - \hat{\beta}_{cs}) (\bar{z}_{cs} + \bar{z}_{ct}) \right].$$

The fact that T3' (and T3) can be decomposed multiplicatively by its characteristics has advantages when interpreting the results. It means that the contribution of each characteristic to the overall price index can be easily discerned. That is, we can decompose the price index as follows:

$$P_{st}^{T3'} = P_{st}^1 \times P_{st}^2 \times \dots \times P_{st}^C,$$

where  $P_{st}^c$  measures the multiplicative contribution of characteristic  $c$  to the differences in house prices between periods  $s$  and  $t$ . Of particular interest is the ratio  $P_{st}^c/P_{st}^{T3'}$ . If this ratio exceeds 1, it implies that characteristic  $c$  is exerting upward pressure on the overall price index, while, when less than 1,  $c$  is exerting downward pressure on the index.

One of the key insights here is that the choice of price index formula and functional form for the hedonic regression model should not be decoupled. In particular, the semi-log

model has a natural affinity with the Törnqvist price index. Hill and Melser (2008) likewise show that the linear model has a natural affinity with the Fisher index. Also, as is shown in the next section, in certain cases imputations methods are dual to characteristics methods. These duality results could also influence the choice of price index formula and functional form for the hedonic model.

It is informative at this point to consider an empirical example. Tables 2 and 3 depict semi-log regression equations estimated for Sydney in 2005 and 2006 using the Australian Property Monitors data set described previously.

**TABLE 2: A Semi-Log Regression for 2005**

Variable	Estimate	Standard Error	t statistic
Intercept	13.19824	0.00619	1303.62
Bed1	-0.06431	0.04297	-1.50
Bed3	0.01570	0.01724	0.91
Bed4	0.07743	0.01269	6.10
Bed5	0.09960	0.01046	9.52
Bath2	0.30875	0.00810	38.10
Bath3	0.60458	0.01255	48.19
Bath4	0.97436	0.02648	36.79
LandArea	0.01480	0.01160	1.28

**TABLE 3: A Semi-Log Regression for 2006**

Variable	Estimate	Standard Error	t statistic
Intercept	13.18259	0.00994	1325.71
Bed1	-0.08396	0.04528	-1.85
Bed3	0.00323	0.01662	0.19
Bed4	0.07717	0.01224	6.31
Bed5	0.10944	0.01016	10.77
Bath2	0.33398	0.00779	42.86
Bath3	0.64746	0.01199	54.02
Bath4	0.96480	0.02458	39.24
LandArea	0.00379	0.01105	0.34



The sample size in Table 2 is 19,849, while in Table 3 it is 23,502. The regression results presented in Tables 2 and 3 have been slightly modified for expositional purposes.

Suppose now we consider a house that sold in 2005. One such specimen has the following characteristics vector:

$$p_{05,h} = \$810,000,$$

$$\text{Number of Bedrooms} = 4,$$

$$\text{Number of Bathrooms} = 2,$$

$$\text{Land Area} = 0.522.$$

Sticking these characteristics into the estimated hedonic models for 2005 and 2006, the following imputed prices are obtained:

$$\hat{p}_{05,h} = \exp(13.19824 + 0.07743 + 0.30875 + 0.01480 \times 0.522) = \$799,823,$$

$$\hat{p}_{06,h} = \exp(13.18259 + 0.07717 + 0.33398 + 0.00379 \times 0.522) = \$802,686.$$

A double imputation Laspeyres price index comparison between years 2005 and 2006 requires the price in 2005 and 2006 of every dwelling sold in 2005 to be imputed in the way outlined above. The overall price index is then obtained by summing all the imputed prices in 2005, summing all the imputed prices in 2006, and then dividing the latter by the former.

So far I have focused exclusively on bilateral comparisons. Imputation methods generalize to three or more periods in the same way as the adjacent-period method. That is, a Fisher or Törnqvist index should be calculated between adjacent periods, and these bilateral indexes are then chained chronologically. For example, a comparison between periods 1 and 3 is made by multiplying a comparison between periods 1 and 2 by a comparison between periods 2 and 3.

#### 4.2.2 Strengths and weaknesses

Imputation methods are very flexible, in that they allow the characteristic shadow prices to evolve over time. The use of double imputation should also help to reduce omitted variables bias. The fact that the imputation method links in well with the much older price index literature and is dual to the characteristics method in some cases may be beneficial in certain contexts, such as the treatment of housing in the CPI, where it may be desirable to use a consistent approach across expenditure categories. It can also be easily combined with nonparametric estimation of the hedonic model.

Perhaps the biggest criticism of the imputations approach is that estimating a separate hedonic model for each period prevents the exploitation of interactions between the equations. From an econometric perspective it may be preferable to estimate the equations as a system of seemingly unrelated regressions (see Zellner 1962). Using this type of approach, however, will tend to lead to violations of temporal fixity as all the results will change when a new period is added to the data set. One further disadvantage of the imputation method is that it does not generate standard errors on the price indexes.

### 4.2.3 Usage

The imputations method has also not been much used. It should be remembered, however, that a number of characteristics methods (see below) can also be described as imputation methods. The only index provider to use an imputation method explicitly as far as I am aware is RPData-Rismark. The RPData-Rismark method is quite complex and entails imputing prices from a semiparametric generalized additive hedonic model.

## 4.3 Characteristics methods

### 4.3.1 Description of the method

Characteristics methods, like imputation methods, generally require the hedonic model to be estimated separately for each period.<sup>13</sup> They also use standard price index formulas. The key difference is that a characteristics price index is defined in characteristics space. The most common characteristics method, at least in a housing context, constructs an average dwelling for each period, and then imputes the price of this hypothetical dwelling (which for example may have two and a half bedrooms) as a function of its characteristics using the shadow prices derived from the hedonic model. The average dwelling may be either the arithmetic mean or median. A price index is obtained by taking the ratio of the imputed price of the same average dwelling in two different periods. By construction it follows that characteristics methods use double imputation.

Taking the semi-log hedonic model as our point of reference, a price index between periods  $s$  and  $t$  can be calculated using the average dwelling from either period (see Dulberger

---

<sup>13</sup>This section also draws extensively on Hill and Melser (2008).

1989 and Diewert 2001). In this way we obtain Laspeyres and Paasche-type indexes, which we refer to here as L5 and P5.<sup>14</sup>

$$\text{L5 : } P_{st}^{L5} = \hat{p}_t(\bar{z}_s)/\hat{p}_s(\bar{z}_s) = \exp \left[ \sum_{c=1}^C (\hat{\beta}_{ct} - \hat{\beta}_{cs}) \bar{z}_{cs} \right], \text{ where } \bar{z}_{cs} = \frac{1}{H_s} \sum_{h=1}^{H_s} z_{csh}, \quad (18)$$

$$\text{P5 : } P_{st}^{P5} = \tilde{P}_{st}^P = \hat{p}_t(\bar{z}_t)/\hat{p}_s(\bar{z}_t) = \exp \left[ \sum_{c=1}^C (\hat{\beta}_{ct} - \hat{\beta}_{cs}) \bar{z}_{ct} \right], \text{ where } \bar{z}_{ct} = \frac{1}{H_t} \sum_{h=1}^{H_t} z_{cth}. \quad (19)$$

$H_s$  and  $H_t$  are again the total number of dwellings sold respectively in periods  $s$  and  $t$ . Alternatively,  $\bar{z}_{cs}$  and  $\bar{z}_{ct}$  in (18) and (19) could be defined as the characteristic vectors of the median dwelling in each period.

The price index  $P_{st}^{L5}$  is analogous to a Laspeyres index in the sense that it uses the earlier period  $s$  as the reference, while  $P_{st}^{P5}$  is analogous to a Paasche index in that it uses the later period  $t$  as the reference. Taking the geometric mean we obtain a price index  $P_{st}^{F5}$  that is analogous to a Fisher index.

$$\text{F5 : } P_{st}^{F5} = \sqrt{P_{st}^{L5} \times P_{st}^{P5}} = \exp \left[ \frac{1}{2} \sum_{c=1}^C (\hat{\beta}_{ct} - \hat{\beta}_{cs}) (\bar{z}_{cs} + \bar{z}_{ct}) \right]. \quad (20)$$

What is particularly interesting here is that F5 on closer inspection can be seen to be identical to T3' from the previous section. In other words, an hedonic imputation index defined on the semi-log model and which uses the Törnqvist (or perhaps more precisely a Jevons type) price index formula is identical to a characteristics index defined on the average house of both periods and calculated using a Fisher-type formula.

The hedonic regression equations in Tables 2 and 3 above can be used to illustrate the characteristics approach. In the Sydney data set, the average characteristic vectors for 2005 and 2006 are as follows:

- AverageBeds in 05 = 3.3256,
- AverageBaths in 05 = 1.6551,
- AverageLandArea in 05 = 0.6048,
- AverageBeds in 06 = 3.3474,
- AverageBaths in 06 = 1.6589,
- AverageLandArea in 06 = 0.6032.

Using these data, prices can now be imputed in either year for the average dwelling of either year. The use of dummy variables for bedrooms and bathrooms does not lend

---

<sup>14</sup>Again a correction is required for bias when deriving the imputed prices. The correction is exactly the same as for imputation methods as discussed in Malpezzi, Chun and Green (1998).

itself naturally to the characteristics approach since the average dwelling inevitably has a fractional number of bedrooms and bathrooms. This problem, however, can be finessed by taking an appropriately weighted linear combination of the coefficients of the two bounding dummy variables. For example, using Table 2, the price of the average 2005 dwelling in 2005 can be imputed as follows:

$$\hat{p}_{05}(\bar{z}_{05}) = \exp[13.19824 + 0.01570 \times (1 - 0.3256) + 0.07743 \times 0.3256 + 0.30875 \times 0.6551 + 0.01480 \times 0.6048] = \$690,556.$$

Imputed prices in 2006 for the average house of 2005 and in 2005 and 2006 for the average house of 2006 can likewise be calculated as follows (where imputed prices in 2006 use the estimated coefficients from Table 3):

$$\hat{p}_{06}(\bar{z}_{05}) = \exp[13.18259 + 0.00323 \times (1 - 0.3256) + 0.07717 \times 0.3256 + 0.33398 \times 0.6551 + 0.00379 \times 0.6048] = \$680,768,$$

$$\hat{p}_{05}(\bar{z}_{06}) = \exp[13.19824 + 0.01570 \times (1 - 0.3474) + 0.07743 \times 0.3474 + 0.30875 \times 0.6589 + 0.01480 \times 0.6032] = \$692,281,$$

$$\hat{p}_{06}(\bar{z}_{06}) = \exp[13.18259 + 0.00323 \times (1 - 0.3474) + 0.07717 \times 0.3474 + 0.33398 \times 0.6589 + 0.00379 \times 0.6032] = \$682,737.$$

A Laspeyres-type price index comparing 2005 with 2006 is now given by  $\hat{p}_{06}(\bar{z}_{05})/\hat{p}_{05}(\bar{z}_{05}) = 680,768/690,556 = 0.9858$ . A Paasche-type price index is given by  $\hat{p}_{06}(\bar{z}_{06})/\hat{p}_{05}(\bar{z}_{06}) = 682,737/692,281 = 0.9862$ . The difference between the two sets of results in this case is minimal. By comparison, according to the time-dummy method, the corresponding price index is 0.9822. All three formulas agree that house prices in Sydney fell slightly from 2005 to 2006.

Another approach to constructing characteristic price indexes is to first compute the quality adjustment factor and then use this to deflate a quality-unadjusted price index in a manner somewhat analogous to that used by the time-dummy method as outlined in (9). The quality-adjustment factor, with the characteristic shadow prices derived from the semi-log hedonic model, could take the form of Laspeyres, Paasche or Fisher-type quantity indexes, referred to here as QL, QP, QF.

$$\text{QL} : Q_{st}^L = \exp \left[ \sum_{c=1}^C (\bar{z}_{ct} - \bar{z}_{cs}) \hat{\beta}_{cs} \right], \quad \text{where } \bar{z}_{cs} = \frac{1}{H_s} \sum_{h=1}^{H_s} z_{csh}, \quad (21)$$

$$\text{QP} : Q_{st}^P = \exp \left[ \sum_{c=1}^C (\bar{z}_{ct} - \bar{z}_{cs}) \hat{\beta}_{ct} \right], \quad \text{where } \bar{z}_{ct} = \frac{1}{H_t} \sum_{h=1}^{H_t} z_{cth}. \quad (22)$$

$$\text{QF} : Q_{st}^F = \sqrt{Q_{st}^L \times Q_{st}^P} = \exp \left[ \frac{1}{2} \sum_{c=1}^C (\bar{z}_{ct} - \bar{z}_{cs}) (\hat{\beta}_{cs} + \hat{\beta}_{ct}) \right]. \quad (23)$$

The quality-adjusted price indexes are now obtained implicitly as follows:

$$\tilde{P}_{st}^L = \frac{\prod_{h=1}^{H_t} (p_h)^{1/H_t}}{\prod_{h=1}^{H_s} (p_s)^{1/H_s}} / Q_{st}^P, \quad (24)$$

$$\tilde{P}_{st}^P = \frac{\prod_{h=1}^{H_t} (p_h)^{1/H_t}}{\prod_{h=1}^{H_s} (p_s)^{1/H_s}} / Q_{st}^L, \quad (25)$$

$$\tilde{P}_{st}^F = \frac{\prod_{h=1}^{H_t} (p_h)^{1/H_t}}{\prod_{h=1}^{H_s} (p_s)^{1/H_s}} / Q_{st}^F. \quad (26)$$

An intriguing feature of the implicit Paasche price index in (25) is that it only requires the hedonic model to be estimated in the base period. Presumably, the base would need updating every five years or so.

### 4.3.2 Strengths and weaknesses

The strengths and weaknesses of characteristics methods are much the same as those of imputation methods that use double imputation. The duality results discussed above, however, can break down for more complicated versions of the hedonic specification such as those that include adjustments for spatial dependence. All things considered, I think imputation methods are more reliable since they impute prices for actual rather than hypothetical dwellings.

### 4.3.3 Usage

The characteristics method in its various guises has proved to be by far the most popular for computing price indexes.

The New House Price Index computed by the Census Bureau in the US, the Halifax and Nationwide indexes in the UK, and the permanent tsb index in Ireland are calculated using the Laspeyres version L5 of the characteristics method in (18) with a semi-log functional form for the hedonic equation (see US Census Bureau undated, and Fleming and Nellis 1985).

An alternative approach is used to construct the CLG index in the UK (see Communities and Local Government 2003, 2004). The CLG index considers 100,000 different combinations of characteristics, which it refers to as cells. Using the characteristic shadow

prices obtained from the hedonic model for that period, it imputes a price for each of these cells. The overall price index is obtained by taking a weighted average of these imputed cell prices. The derivation of these weights is an important feature of the method. Unfortunately, the available documentation does not explain how these weights are calculated. One interesting feature of the method that follows from its use of imputed characteristic prices is that it does not matter if some of the cells are empty in a particular period.

Statistics Finland uses a weighted variant on the implicit Paasche price index in (25) (see Saarnio 2006) to compute house price indices for Finland. Statistics Norway uses the same method as Statistics Finland except that it calculates its hedonic model using the previous five years of data and chains the index on an annual basis (see Thomassen 2007). Statistics Sweden also uses a variant on the implicit Laspeyres price index in (25) (see Ribe 2009), although the exact details of the method are not provided. Closely related to these Nordic methods is the Conseil Supérieur du Notariat (CSN) and INSEE method used to compute hedonic indexes for regions in France (see Gouriéroux C. and A. Laferrère 2009).

Characteristics methods have also proved useful for making comparisons of prices across regions (see Gillingham 1975, Moulton 1995, Wallace 1996, and Chowhan and Prud'homme 2004).

#### **4.4 Nonparametric methods**

The use of nonparametric methods for estimating hedonic models of the housing market has become popular in recent years (one of the earliest examples is Meese and Wallace 1991). The main advantage of nonparametric methods is that they do not need to assume a functional form for the hedonic model, and hence avoid problems of misspecification. They tend to work particularly well for large data sets.

Comparisons of parametric and nonparametric models almost invariably find that the latter outperform the former in terms of mean-square error for out of sample predictions (see Pace 1993, Bao and Wan 2004, and Martins-Filho and Bin 2005).

The greater flexibility of the nonparametric models, however, can come at a price. A parametric hedonic model provides characteristic shadow prices while a nonparametric model may not. Nevertheless, it is often possible to provide equivalent information by other means in a nonparametric model (see for example Wallace 1996 and Kagie and van

Wezel 2007). For example, shadow prices at a particular point may be obtainable from the slope of the estimated nonparametric surface.

Most articles that construct nonparametric hedonic models of the housing market (for example Pace 1993, Bao and Wan 2004, Martins-Filho and Bin 2005) do not attempt to construct hedonic price indexes. Notable exceptions are Meese and Wallace (1991) and Wallace (1996). Both these studies derive characteristic shadow prices from the slope of the nonparametric surface and then feed them into the Fisher price index formula in (20). Another interesting case is Clapp (2004), who estimates prices for particular points on the nonparametric surface. In my opinion, however, prices at particular points on the surface do not constitute price indexes. A temporal price index should track changes in house prices for a specific region over time, while a spatial price index should compare prices across two or more regions at the same point in time.

There has been surprisingly little discussion of how to use nonparametric methods to construct price indexes. Probably the most natural way is by combining a nonparametric hedonic model with an imputation method. It makes no difference to the imputation method whether the hedonic model is parametric or nonparametric. All that is required is that it is possible to impute a price for any vector of characteristics, and that these imputed prices should be time dependent. I can find only two examples in the literature that combine a nonparametric hedonic model with the imputation method. Kagie and van Wezel (2007) estimate their hedonic model using a decision tree method called boosting, while Hill, Melser and Reid (2010) use a multidimensional thin-plate spline. In both cases they use the Fisher price index formula, although Kagie and van Wezel use single imputation (i.e., F1) while Hill, Melser and Reid use double imputation (i.e., F3). Also, RPData-Rismark compute an imputation index in combination with a semiparametric hedonic model. The exact details of the method, however, are not publicly available.

A nonparametric hedonic model can also be combined with the characteristics method (see Meese and Wallace 1991, and Wallace 1996). However, this requires the derivation of characteristic shadow prices from the hedonic model, which is not necessarily straightforward. Also, imputing prices based on average values of characteristics (such as location) is not necessarily always meaningful.

The fact that there are few attempts in the housing literature to construct price indexes from nonparametric hedonic models perhaps should be attributed more to a lack of interest in this topic amongst researchers in the nonparametric field rather than to the inherent

difficulty of the exercise. In my opinion the combination of a nonparametric hedonic model and the imputation method has a lot of potential.

## 4.5 Repeat sales and hedonic hybrid methods

Attempts have been made to combine the repeat-sales and hedonic approaches. Case and Quigley (1991) and Quigley (1995) use samples of single-sale and repeat-sale dwellings to jointly estimate price indexes using generalized least squares. Hill, Knight and Sirmans (1997) undertake a similar although more general exercise using maximum likelihood. A rather different perspective is provided by Shiller (1993) who proposes a repeat-sales method that allows the price path for dwellings to depend on their quality measured using a hedonic method (see also Clapp and Giaccotto 1998).

Focusing on the Hill-Knight-Sirmans version of the hybrid model, they estimate the following two equations:

$$\ln p_{th} = \sum_{c=1} \beta_c z_{cth} + aA_{th} + \sum_{\tau=1}^T \delta_\tau d_{\tau h} + \epsilon_{th}, \quad (27)$$

$$\ln p_{t+s_h, h} - \ln p_{th} = as_h + \delta_\tau f_{\tau h} + \varepsilon_{th}, \quad (28)$$

where  $A_{th}$  denotes the age of dwelling  $h$  from period  $t$  and  $s_h$  the time being sales of dwelling  $h$ . Equation (27) is a hedonic model and (28) is a repeat-sales model. The hybrid model stacks the two equations and estimates them jointly. Hill-Knight-Sirmans use a characteristics method (L5) to derive the price indexes. Estimating (27) and (28) jointly, however, implies imposing the restriction that the  $\delta_t$  parameter in period  $t$  is the same for single and repeat sales. Unobservable differences in the single and repeat-sales data sets could make this restriction problematic. Meese and Wallace (1997) make this same point in the context of the Case-Quigley model.

Ultimately, the rationale for hybrid methods is to try and combine the best features of each approach. Advocates of these hybrid methods presumably believe that repeat-sales price relatives (adjusted for estimated depreciation) are more reliable indicators of quality-adjusted price changes than are say double imputation hedonic price relatives. Conversely, repeat-sales methods throw away all the single sales data. By combining the two approaches, no data are discarded while repeat sales are still allowed to play a prominent role in the index construction methodology. Even putting aside the concerns raised by Meese and Wallace, I have difficulty accepting the assumption that a repeat-sales price relative



should be preferred to a double imputation hedonic price relative. We must trade off the risk of omitted variables bias in the double imputation price relative against the risk that changes have been made to a repeat-sales dwelling. I am reminded of Heraclitus, who said that “you cannot step into the same river twice”. I have similar concerns about repeat sales. If repeat-sales price relatives are not deemed more reliable than double imputation price relatives, there is no reason to prefer hybrid methods to hedonic methods.

## 5 Geospatial Data and Spatial Dependence

As has been already noted, one of the most important determinants of house prices is location. The simplest way of accounting for location is through the inclusion of neighborhood dummy variables. The increased availability of longitude and latitude (geospatial) data at the level of individual dwellings, however, has stimulated a burst of research exploring ways of using such geospatial data in the construction of hedonic price indexes (starting particularly with Anselin 1988).

If location is an important determinant of price, then it follows that house prices are spatially dependent. Spatial dependence is likely to exist since many of the price determining factors are shared by neighborhoods but are difficult to document explicitly (see Basu and Thibodeau 1998). Neighborhoods tend to develop at the same time resulting in dwellings having similar structural characteristics, and dwellings in a neighborhood share the same locational amenities. The existence of positive spatial correlation in the residuals of hedonic models that fail to account for spatial dependence has been reported in many empirical studies (see for example Can 1990, Pace and Gilley 1997, Basu and Thibodeau 1998, Pace, Barry, Clapp and Rodriguez 1998, and Bourassa, Hoesli and Peng 2003).

The presence of spatial correlation implies that the Gauss-Markov assumptions are violated. At the very least this implies inefficient estimators and incorrect standard errors where available (see Anselin 1988 and Basu and Thibodeau 1998). It can potentially also cause bias in the estimators themselves. If the spatially correlated omitted variables are correlated with the included variables, then the estimated coefficients can be biased even in large samples.

The main benefit of explicitly accounting for spatial dependence in a hedonic model is that it should help offset the locational omitted variables problem. This is because while

many of the price determining factors shared by neighborhoods are difficult to document explicitly, their influence is contained in the prices of neighboring dwellings.

Spatial dependence can be captured either in the regressors or the error term. This distinction is clearly demonstrated by the mixed autoregressive-regressive and spatial errors models discussed in LeSage (1998). Both these models require as a first step the construction of a spatial weights matrix. The dimensions of such a matrix could be very large (e.g., 100,000 by 100,000 if there are 100,000 dwellings). A popular way of easing the computational burden associated with such a large matrix is to use the Delaunay triangulation algorithm to reduce  $W$  to a matrix of zeros and ones (mostly the former). With longitude and latitude of each dwelling as inputs, the Delaunay algorithm creates a set of triangles in two-dimensional Cartesian space such that no points are contained in any triangle's circumcircle. The edges of each triangle satisfy the 'empty circle' property. That is, the circumcircle of a triangle formed by three points is empty if it does not contain any other vertices apart from the three that define it. Two dwellings are categorized as neighbors if there exists a triangle on which they are both vertices. More densely populated areas contain more triangles. If dwellings  $j$  and  $k$  are neighbors then positions  $jk$  and  $kj$  in the spatial weights matrix  $W$  each receive a weight of 1. If  $j$  and  $k$  are not neighbors, positions  $jk$  and  $kj$  receive weights of 0.<sup>15</sup> Also, all the terms on the lead diagonal  $jj$  are set to zero. The main attraction of this approach is that a spatial matrix containing only binary numbers is easier to work with (see Kelejian and Robinson 1995 for a discussion of alternative ways of constructing a spatial weights matrix).

Once we have defined the spatial weights matrix  $W$ , spatial dependence can be captured in a simultaneous autoregressive (SAR) hedonic model. The SAR model can be written as follows:

$$y = X\beta + u,$$

$$u = \lambda Wu + \varepsilon,$$

where  $\varepsilon$  is an  $H \times 1$  vector of random errors. The parameters to be estimated are the  $C \times 1$  vector  $\beta$  and the scalar  $\lambda$ . The parameter  $\lambda$  measures the average locational influence of the neighboring observations on each observations. For example,  $\lambda = 0.30$  means that 30 per cent of the variation of  $u_h$  is explained by locational influences of dwelling  $h$ 's neighbors. One attraction of this approach to modeling spatial dependence is that it only requires the estimation of one additional parameter (namely  $\lambda$ ). The parameters ( $\beta$  and

---

<sup>15</sup>Matlab 6.5 has an in-built Delaunay triangle algorithm routine.

$\lambda$ ) can be estimated using maximum likelihood (see Anselin (1988)). This method is used for example by Pace and Gilley (1997) and Hill, Melser and Syed (2009). In principle, with appropriate modifications, it can be combined with any of the time-dummy, imputation or characteristics methods. The model can also be extended to capture spatial-temporal dependence. The spatial-temporal autoregressive (STAR) model is used by Pace et al. (1998) and Nappi-Choulet and Maury (2009). Bayesian versions of the STAR model have also been developed by Gelfand et al. (1998).

An alternative approach is to calculate the error covariance matrix directly using geo-statistical methods. The hedonic model in this case is as follows:

$$y = X\beta + u,$$

where  $E(uu') = \Omega$ , with at least some non-zero off-diagonal terms. Once the covariance matrix  $\Omega$  has been computed, the model can be estimated by generalized least squares (see for example Basu and Thibodeau 1998, Dubin 1988, 1998, Gillen, Thibodeau and Wachter 2001, and Bourassa, Cantoni and Hoesli 2007). The model is extended to a Bayesian setting by Gelfand et al. (2004).

A further alternative is what LeSage (1998) refers to as a mixed autoregressive-regressive model. This model captures spatial dependence in the regressors as follows:

$$y = \rho W y + X\beta + \varepsilon,$$

where now the parameters to be estimated are  $\beta$  and  $\rho$ . The model can again be estimated by maximum likelihood.<sup>16</sup> Variants on this method are used by Can (1992) and Can and Megbolugbe (1997).

A more direct approach is to include location as an explanatory variable either parametrically (see Fik, Ling and Mulligan 2003) or nonparametrically (see Colwell 1998, Clapp 2003, Bao and Wan 2004, and Hill, Melser and Reid 2009).

Given the importance of location as a determinant of house prices, it is perhaps surprising that RPData-Rismark is the only hedonic index provider, as far as I know, that

---

<sup>16</sup>A useful source for advice on estimation on spatial econometric models in a housing context is the Spatial Econometrics Toolbox for MATLAB developed by J. P. LeSage available at <http://www.spatial-econometrics.com>. The details on the toolbox and the associated computational issues can be found in LeSage (1999). A second toolbox for MATLAB which implements various spatial auto-regressive models has been developed by R. K. Pace and R. Barry <http://www.spatial-statistics.com>.

makes use of geospatial data. The lack of indexes exploiting geospatial data may be attributable to the fact that the vast majority of hedonic indexes are calculated using the characteristics method. While it is not unreasonable to price an artificial house with the average number of bedrooms, bathrooms land area, etc., the use of an average location is more problematic. For example, for coastal cities with harbors, such as Sydney, the average location could be underwater. More generally, there is no reason necessarily to expect the average location to have the average level of locational amenities. The apparent strong preference for characteristics methods among index providers, therefore, may be impeding the use of geospatial data. For this reason, I think imputation methods such as the one used by RPData-Rismark should be preferred to characteristics methods.

## 6 Estimation of Hedonic Models

At this stage it may be useful to run through the list of decisions that must be made by the provider of an hedonic index. First, one must decide on the basic methodology. The index could be constructed using the time-dummy, imputation or characteristics methods. For the former, the next task is to choose a functional form for the hedonic model. For the latter two methods, it is necessary to choose both a price index formula and a functional form (which may even be nonparametric).

One must also decide on the list of explanatory characteristics, and on whether to include interactions between particular combinations of characteristics (e.g., bedrooms and bathrooms) or transformations of particular characteristics, such as land area. Also, discrete variables such as bedrooms and bathrooms can be included either as standard variables or as dummy variables. For example, an explanatory variable could be the number of bedrooms or separate dummy variables can be included for two bedrooms, three bedrooms, four bedrooms, etc. The latter approach has the advantage of greater flexibility in that it allows the effect of an extra bedroom to differ depending on the initial number of bedrooms. However, this increased flexibility comes at the price of less degrees of freedom. For data sets with large numbers of observations, which is increasingly the norm, the dummy variable approach is probably preferable.

Similarly, estimating a separate hedonic model for each period, as the imputation and characteristics methods do, also acts to significantly reduce the degrees of freedom. While modern data sets are often large enough to allow separate estimation of the model for

each period, from an econometric point of view this is clearly inefficient. Some variant on Zellner's (1962) seemingly unrelated regression (SUR) method would be a useful addition to the imputation and characteristics methods.

According to Sirmans et al. (2006), the nine characteristics that appear most often in hedonic regressions for housing (and all of which are of the physical variety) are floor space, land area, age, bedrooms, bathrooms, garage, swimming pool, fireplace, and air conditioning. To these presumably can be added dwelling-type dummy variables (e.g., house, townhouse, apartment, etc). The choice of explanatory characteristics is often determined largely by data availability. For data sets with relatively few characteristics (see for example Hill, Melser and Syed 2009) the omitted variables problem may be a particular concern. Such a situation strengthens the case for using either an imputation or characteristics method, since the use of double imputation helps offset the omitted variables problem. Conversely, the case for using the time-dummy method is stronger when the number of dwellings in the data set is small. In such cases, estimating separate hedonic models for each period could generate erratic coefficients.

In most cases, the index provider has a prior expectation as to the expected sign on each coefficient in the hedonic model. At first glance, one might expect the estimated coefficients on all nine characteristics above to be positive, with the exception of age. On further reflection the situation is often more complicated. The impact of age in particular on house prices can be quite complex. Newer dwellings typically command a price premium. However, older dwellings may also command a premium in the same way as antique furniture. For example, in Graz, Austria, apartments built before 1914 are much sought after. In other words, the impact of age on house prices may be nonmonotonic. Imposing a monotonic relationship between house prices and age in the hedonic equation therefore may do more harm than excluding age completely from the model. A better approach in this case might be to assume a quadratic relationship between age and house prices or to specify dummy variables for dwellings in various age ranges (e.g., 0-10, 11-20, 21-40, etc).

The sign of the bedrooms coefficient may also be unclear when floor space is also included. This is because the inclusion of an extra bedroom while holding floor space fixed is not necessarily desirable, since it has an associated opportunity cost of less space for everything else in the dwelling. The same tradeoff may also exist between floor space and land area, since more floor space holding land area fixed implies a smaller garden.

An index provider with access to locational characteristics, such as postcodes or longi-

tudes and latitudes for each dwelling faces further dilemmas. The latter can be used to construct numerous locational characteristics, such as the distance to the nearest school, train station, hospital, shopping center, airport, park, beach or to the city center. The relationship between distance to a particular amenity and house prices, like age, is sometimes nonmonotonic. For example one might ideally want to live reasonably near but not too near a shopping center, train station or hospital. Hence as with age, a quadratic specification for the relationship between distance to an amenity and house prices (or the specification of dummy variables for various distance ranges) may in some cases be appropriate.

More generally, I am not convinced that there is much benefit to including locational characteristics over and above postcode dummies and longitude and latitude (where the latter are used to explicitly model spatial dependence). This is an issue though that could benefit from further empirical research.

Given that sufficient attention has been paid to the complexities that can arise from interactions between characteristics, how concerned should one be if some of the signs of the estimated coefficients do not accord with prior expectations? Pakes (2003) created some controversy by arguing that at least in monopolistically competitive markets subject to mark-ups one should not interpret the estimated coefficients as shadow prices, and hence that one should not have much in the way of prior expectations regarding the signs of these coefficients. In my opinion, Pakes's arguments do not carry through to the housing market, where the vast majority of buyers and sellers are private households, most dwellings on the market have not been produced in that particular period, and where the fundamental building block of value – namely the land itself – has not been produced at all. Hence I would argue that unexpected signs on some of the coefficients, at least in a housing context, are a cause for concern. Having said that, it is a fact of life in empirical research that the results do not work out exactly as one had hoped.

The semi-log functional form has been widely used in the hedonic literature. Given a semi-log specification, probably the two most likely causes of unexpected coefficient signs are insufficient sample size, and poor quality data.

One particular data problem often encountered is missing observations for some characteristics. For example, the bedroom count may be missing for a certain percentage of the dwellings in the data set. This problem can be dealt with by deleting from the model any characteristics that are particularly prone to having missing observations, or alternatively by omitting all dwellings that have an incomplete list of characteristics. Neither of these

solutions is particularly appealing. Both throw away potentially useful data. Omitting dwellings with incomplete characteristics lists may also cause sample selection bias. For example, Hill, Melser and Syed (2009) find the incomplete characteristics list problem in their data set is more severe in poorer suburbs of Sydney. Deletion of dwellings with incomplete characteristics lists may therefore impart bias to the index if the price paths of richer and poorer suburbs differ in a systematic way over time. An alternative solution is to simply set all missing observations to zero or some other default value. This, however, may also create distortions and perhaps bias. In my opinion, the problem of missing observations should be dealt with in one of two ways. The first approach can only be used in combination with the imputation method. This approach requires a number of hedonic models to be estimated for each period each with varying combinations of explanatory variables. The imputed price for each dwelling is then calculated using the hedonic model that includes all the characteristics that are available for that particular dwelling and excludes any characteristics that are missing for that particular dwelling. In this way, all the available and relevant information is used when imputing the price of each dwelling. While this method is computationally somewhat laborious, it is well within the capabilities of modern computers. The second approach is to impute values for the missing observations prior to estimating the hedonic model. That is, rather than imputing the price of a dwelling from its list of characteristics, one must first impute say the number of bathrooms in a dwelling from its price and its other characteristics. This approach risks introducing a circularity into the estimation method if prices are used to impute characteristic values which are then in turn used to impute prices. The multiple imputation (MI) method developed by Rubin (1976, 1987) and others is an example of such a method. MI imputes say 10 sets of values for the missing observations, then estimates the regression model for each set of imputations and only then averages the results across the 10 sets of results. The imputations exploits correlations between the variables to impute missing data points. The MI method is applied to housing data by Syed, Hill and Melser (2008). It is important that the assumptions underlying the imputation methodology are appropriate to the particular context. Otherwise the process of imputation of missing observations may act merely to hide rather than reduce the problem of sample selection bias.

A further problem faced by an index provider prior to estimation of the hedonic model is the problem of outliers. For example, suppose a particular dwelling has 20 bedrooms. This may be because it is in fact a hotel, and hence should be excluded if the objective is to construct a price index for residential housing. Alternatively, the entry of 20 bedrooms

may be a typo and should be 2 rather than 20. In housing data sets, the vast majority of outliers tend to be of the latter variety, and are probably best dealt with by simple deletion or replacement by imputed values. The problem here is deciding where to draw the line. For example, should the line be drawn at 7 bedrooms, 8 bedrooms, or 9 bedrooms or at some higher or lower level? Similar decisions need to be made for bathrooms, land area, and floor space and perhaps for other characteristics as well.

## 7 Criticisms of Hedonic Price Indexes

### 7.1 Omitted variables bias

The omitted variables problem is likely to be much more severe for housing than say for computers. It is probably impossible to quantify all the factors that influence the price of a dwelling, while the list of relevant characteristics for a computer is presumably relatively short and more quantifiable (e.g., RAM, ROM, speed, screen type and size, manufacturer, weight).

In a housing context a distinction can be drawn between omitted variables that relate to the physical characteristics of a dwelling (such as land area) and those that relate to its location (such as distance to the city center or the local crime rate). As noted above, according to Sirmans et al. (2006), the nine characteristics that appear most often in hedonic regressions for housing (and all of which are of the physical variety) are square footage, land area, age, bedrooms, bathrooms, garage, swimming pool, fireplace, and air conditioning.

Examples of likely omitted variables in hedonic models of the housing market include the following: state of maintenance of a dwelling, traffic noise, the amount of sunlight received, the functionality of the layout of the rooms, the presence or otherwise of damp or water damage, the quality of the building materials and workmanship, and the general ambience.

The significance of the omitted variables problem depends on the objective of the exercise. For example, omitted variables are a much larger problem for automated valuation (appraisal) models or studies that focus specifically on the characteristic shadow prices (e.g., for clean air) than for price indexes. There is a reasonable expectation that the omitted variables bias will mostly offset itself across dwellings in a hedonic price index (see



Malpezzi 2002), particularly when double imputation is used. Furthermore, the extent of the locational omitted variables problem can be reduced by including postcode dummies and modeling spatial dependence using geospatial data.

## **7.2 Functional form misspecification**

Given the hedonic model is a reduced form it is not possible to determine the appropriate functional form by theoretical reasoning alone (see Rosen 1974). The most popular functional form is the semi-log model. Malpezzi (2002) lists five advantages of the semi-log model. These are that it allows the value added of an incremental increase in a particular characteristic to vary proportionally with the size and quality of a dwelling, the simple and appealing interpretation of the estimated coefficient parameters, computational simplicity, mitigation of heteroscedasticity, and the ease with which it can be extended to include nonparametric terms such as splines.

Interest in Box-Cox (1964) transformations, which nest semi-log as a special case, waxed and waned in the 1980s (see for example Halverson and Pollakowski 1981). As Coulson (2008) points out, rather than turning to more flexible functional forms, researchers concerned about the restrictiveness of the semi-log model and the resulting risk of functional form misspecification have tended to turn instead to semiparametric or nonparametric models (an early notable example is Meese and Wallace 1991). Nonparametric methods can approximate any relationship without prespecification of functional form. Given the availability now of various nonparametric approaches in the hedonic literature, functional form misspecification (beyond the omission of relevant explanatory variables) is no longer as big a concern as it used to be.

## **7.3 Data mining and lack of transparency and reproducibility**

Shiller (2008) makes the following criticism of hedonic methods:

The problem is that there are too many possible hedonic variables that might be included, and if there are  $n$  possible hedonic variables, then there are  $n$ -factorial possible lists of independent variables in a hedonic regression, often a very large number. One could strategically vary the list of included variables until one

found the results one wanted. Looking at different hedonic indices for the same city, I remember seeing substantial differences, which must be due to choices the constructors made. Thus, the indices have the appearance of hypotheses rather than objective facts (Shiller 2008; p. 10).

It is undoubtedly true that two researchers given the same data set will end up constructing different hedonic indexes. Shiller in fact far understates the number of choices that must be made by an index provider. As noted above, the index provider in addition to choosing which hedonic explanatory variables to include must also decide on the following:

- whether to include interaction terms between explanatory variables,
- a functional form (which may be parametric, semiparametric or nonparametric),
- one of the time-dummy, imputation and characteristics approaches,
- for the latter two approaches an index number formula,
- the treatment of spatial dependence,
- the treatment of missing observations (e.g., the bathroom count could be missing for some dwellings),
- the treatment of outliers.

The flexibility of the hedonic approach can be viewed both as an advantage and disadvantage. Shiller highlights the disadvantages of flexibility. One notable advantage is that the index provider can tailor the approach used to the data set and the needs of users. For example, the case for using double imputation rises when only a small list of characteristics are available, while the case for using the time-dummy method rises when the sample of dwellings is small.

There is also the matter of how sensitive the hedonic index is to all of these choices. Sirmans et al. (2006) find that the estimated shadow prices of characteristics are surprisingly stable across semi-log models estimated by different researchers using different data sets. It would be a useful exercise to see how sensitive hedonic house price indexes really are to the various choices faced by the provider.

Shiller also raises the specter of strategic manipulation. While it is true that academic researchers in the hedonic field seeking new results for publication in an academic journal have an incentive to use the choices available strategically, the same is unlikely to be true for index providers. An index provider typically chooses a methodology (hopefully makes it publicly available) and then sticks with it. Hence I feel that Shiller's concerns over strategic

manipulation of hedonic indexes are overstated, at least for index providers.

Finally, it is worth remembering that the provider of a repeat-sales index (Shiller's preferred method) also has to make a number of choices (e.g., the treatment of outliers, of repeat sales at very short time intervals, and the relative weighting given to shorter and longer time interval repeat sales). Hence, two researchers given the same data set would probably likewise end up constructing different repeat-sales indexes as well. Admittedly, the differences would probably be smaller, although this need not necessarily be the case. Leventis (2008) for example documents how the differing treatment of repeat-sales with longer time intervals between sales generates significant differences between the SPCS and OFHEO repeat-sales indexes. This issue also warrants further investigation.

## 7.4 Sample selection bias

A hedonic sample may be subject to two types of sample selection bias. First the population of house sales may not be representative of the overall housing stock. This is a problem that applies to all index construction methods that rely on transaction as opposed to appraisal prices (particularly repeat-sales indexes). In fact, whether this is actually a problem may depend on how the index is used. For example, from the perspective of monetary policy it is not clear whether it is desirable to try and adjust a hedonic index to make it more representative of the housing stock.

If characteristics information are available on all dwellings (not only those transacted) then it is possible if desired to correct for this type of sample selection bias. Gatzlaff and Haurin (1998) propose a censored regression method. The probability of a sale is estimated using property, owner, and macroeconomic factors. These probabilities are used to calculate a selection bias correction variable. Gatzlaff and Haurin find that their corrected price indexes are more volatile than the uncorrected indexes. LeSage and Pace (2004) and Reid (2008) follow a different approach and impute prices in each period for all properties for which characteristics data are available. LeSage and Pace impute prices using the multiple imputation approach of Rubin (1976, 1987). Reid, by contrast, imputes prices directly from the estimated hedonic model, and then constructs price indexes using the double imputation method applied to all dwellings in the data set irrespective of whether they were traded in either of the two periods being compared. He refers to this as the augmented double imputation method. These methods should all generate price indexes that are more

representative of the overall housing stock.

The second type of sample selection bias arises when certain types of housing transactions are more likely to be recorded than others. Hill, Melser and Syed (2009) find two such trends in their data set for Sydney, Australia (obtained from Australian Property Monitors). First, the coverage improves over time. For example, a transaction in 2006 is more likely to be included than a transaction in 2001. Second, the coverage seems to be better in richer suburbs of Sydney, both in terms of the percentage of total sales included and the list of characteristics provided. This latter type of sample selection bias in particular could translate into index bias if the price path across richer and poorer suburbs differs in a systematic way over time.

Incomplete lists of characteristics for some dwellings (e.g., the bathroom count is missing) can also cause problems, particularly if these dwellings are simply deleted. As well as wasting data, such wholesale deletion can cause sample selection bias. The relevance of this problem may differ considerably from one data set to the next. For an example where this problem is severe, see Hill, Melser and Syed (2009). Ways of addressing this problem are discussed in section 6 above.

## 8 Conclusion

Hedonic indexes seem to be gradually replacing repeat-sales indexes as the method of choice for constructing quality-adjusted house price indexes. This trend can be attributed to the inherent weaknesses of the repeat-sales method (especially its deletion of single-sales data) and a combination of the increasing availability of detailed data sets of house prices and characteristics including geospatial data, increases in computing power and the development of more sophisticated hedonic models that in particular take account of spatial dependence in the data. The hedonic approach provides a rich and flexible structure that allows index providers to tailor the method to the available data and the needs of users. As Hoffman and Lorenz (2006) say, “the future will belong to hedonic indices.”

## References

- Academetrics (2009), *House Price Indices - Fact or Fiction*, www.academetrics.co.uk.
- Ahnert H. and G. Kenny (2004), *Quality Adjustment of European Price Statistics and the Role for Hedonics*, Occasional Paper Series No. 15, European Central Bank.
- Anselin L. (1988), *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwar Academic Publishers.
- Arguea N. M. and C. Hsiao (1993), "Econometric Issues of Estimating Hedonic Price Functions: With an Application to the US Automobile Market," *Journal of Econometrics* 56(1-2), 243-267.
- Australian Bureau of Statistics (2006), *A Guide to House Price Indexes* (cat. no. 6464.0). Canberra: Australian Bureau of Statistics.
- Balk B. M. (1995), "Axiomatic Price Index Theory: A Survey," *International Statistical Review* 63(1), 69-93.
- Bao H. X. H. and A. T. K. Wan (2004), "On the Use of Spline Smoothing in Estimating Hedonic Housing Price Models: Empirical Evidence Using Hong Kong Data," *Real Estate Economics* 32(3), 487-507.
- Basu S. and T. Thibodeau (1998), "Analysis of Spatial Autocorrelation in House Prices," *Journal of Real Estate Finance and Economics* 17(1), 61-85.
- Bayer P. N. Keohane and C. Timmins (2009), "Migration and Hedonic Valuation: The Case of Air Quality," *Journal of Environmental Economics and Management* 58(1), 1-14.
- Benkard C. L. and P. Bajari (2005), "Hedonic Price Indexes With Unobserved Product Characteristics, and Application to Personal Computers," *Journal of Business and Economic Statistics* 23(1), 61-75.
- Berndt E. R., A. Bir, S. H. Busch, R. G. Frank, S. T. Normand (2002), "The medical treatment of depression, 1991-1996: productive inefficiency, expected outcome variations, and price indexes," *Journal of Health Economics* 21(3), 373-396.
- Berndt E. R., Z. Griliches, and N. J. Rappaport (1995), Econometric Estimates of Price Indexes for Personal Computers in the 1990s, *Journal of Econometrics* 68, 243-268.
- Berry B. J. L. and R. Bednarz (1975), "A Hedonic Model of Prices and Assessments for Single Family Homes in Chicago: Does the Assessor Follow the Market or the Market

- Follow the Assessor?" *Land Economics* 51(1), 21-40.
- Boskin M., E. Dullberger, R. Gordon, Z. Griliches, and D. Jorgenson (1996), *Toward a More Accurate Measure of the Cost of Living*. Final report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index.
- Bourassa S. C., E. Cantoni and M. Hoesli (2007), "Spatial Dependence, Housing Submarkets and House Price Prediction," *Journal of Real Estate Finance and Economics* 35, 143-160.
- Bourassa S. C., M. Hoesli and V. C. Peng (2003), "Do Housing Submarkets Really Matter?" *Journal of Housing Economics* 12(1), 12-18.
- Box G. E. P. and D. Cox (1964), "An Analysis of Transformations," *Journal of the American Statistical Association, Society Series B* 26, 211-252.
- Calhoun C. A. (1996), OFHEO House Price Indexes: HPI Technical Description, Office of Federal Housing Enterprise Oversight, Washington, D.C.
- Can A. (1992), "Specification and Estimation of Hedonic Housing Price Models," *Regional Science and Urban Economics* 22, 453-474.
- Can A. and I. Megbolugbe (1997), "Spatial Dependence and House Price Index Construction," *Journal of Real Estate Finance and Economics* 14, 203-222.
- Cartwright D. W. (1986), "Improved Deflation of Purchases of Computers," *Survey of Current Business* 66 (March), 7-9.
- Case B., H. O. Pollakowski and S. M. Wachter (1991), "On Choosing among House Price Index Methodologies," *American Real Estate and Urban Economics Association Journal* 19(3), 286-307
- Case B. and J. M. Quigley (1991), "The Dynamics of Real Estate Prices," *Review of Economics and Statistics* 73(1), 50-58.
- Case K. E. and J. M. Quigley (2008), "How Housing Booms Unwind: Income Effects, Wealth Effects, and Feedbacks through Financial Markets," *European Journal of Housing Policy* 8(2), 161-180.
- Case K. E., J. M. Quigley and R. J. Shiller (2005), "Comparing Wealth Effects: The Stock Market versus the Housing Market," *The B.E. Journal of Macroeconomics* 5(1) (Advances), Article 1.
- Case K. E. and R. J. Shiller (1987). "Prices of Single Family Real Estate Prices," *New England Economic Review*, Sep, 45-56.

- Case K. E. and R. J. Shiller (1989), The Efficiency of the Market for Single-Family Homes, *American Economic Review* 79, 125-137.
- Case K. E., R. J. Shiller and A. N. Weiss (1993) "Index-Based Futures and Options Markets in Real Estate," *Journal of Portfolio Management*, Winter, 83-92.
- Cecchetti S. (2007), "Housing in Inflation Measurement," <http://www.voxeu.org/index.php?q=node/248>
- Chanel O., L. Grard-Varet and V. Ginsburgh (1996), "The relevance of hedonic price indices: The case of paintings," *Journal of Cultural Economics* 20(1), 1-24.
- Chinloy P. T. (1977), "Hedonic Price and Depreciation Indexes for Residential Housing: A Longitudinal Approach," *Journal of Urban Economics* 4(4), 469-482.
- Chowhan J. and M. Prudhomme (2004), "City Comparisons of Shelter Costs in Canada: A Hedonic Approach," Research Paper, Catalogue No. 62F0014MIE, Series No. 17, Prices Division, Statistics Canada, Ottawa.
- Clapham E., P. Englund, J. M. Quigley and C. L. Redfearn (2006), "Revisiting the Past and Settling the Score: Index Revision for House Price Derivatives," *Real Estate Economics* 34(2), 275-302.
- Clapp J. M. (2003), "A Semiparametric Method for Valuing Residential Locations: Application to Automated Valuation," *Journal of Real Estate Finance and Economics* 27(3), 303-320.
- Clapp J. M. (2004), "A Semiparametric Method for Estimating Local House Price Indices," *Real Estate Economics* 32(1), 127-160
- Clapp J. M. and C. Giaccotto (1992), "Estimating Price Trends for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods," *Journal of Real Estate Finance and Economics* 5(4), 357-374.
- Clapp J. M. and C. Giaccotto (1998), "Price Indices Based on the Hedonic Repeat-Sales Method: Application to the Housing Market," *Journal of Real Estate Finance and Economics* 16(1), 5-26.
- Clapp J. M. and C. Giaccotto (1999), "Revisions in Repeat-Sales Price Indexes: Here Today, Gone Tomorrow?" *Real Estate Economics* 27(1), 79-104.
- Cockburn I. M. and A. H. Anis (2001), "Hedonic Analysis of Arthritis Drugs," in D. M. Cutler and E. R. Berndt (eds.), *Medical Care Output and Productivity*. Cambridge, MA: MIT Press, 439-462.

- Cole R., Y. C. Chen, J. A. Barquin-Stolleman, E. Dulberger, N. Helvacian and J. H. Hodge (1986), "Quality-Adjusted Price Indexes for Computer Processors and Selected Peripheral Equipment," *Survey of Current Business* 66 (January), 41-50.
- Collins, A. A. Scorcu and R. Zanola (2009), "Reconsidering Hedonic Art Price Indexes," *Economics Letters* 104(2), 57-60.
- 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.
- Communities and Local Government (2003), Experimental Monthly House Price Index Methodology - full report.  
<http://www.communities.gov.uk/documents/housing/pdf/141407.pdf>
- Communities and Local Government (2004), The "Hedonic" model used in the ODPM's House Price Indices.  
<http://www.communities.gov.uk/documents/housing/pdf/141410.pdf>
- Coulson E. (2008), *Monograph on Hedonic Estimation and Housing Markets*, Department of Economics, Penn State University.
- Court A. T. (1939), "Hedonic Price Indexes with Automotive Examples," in *The Dynamics of Automobile Demand*. The General Motors Corporation, New York, 99-117.
- de Haan J. (2004a), "Direct and Indirect Time Dummy Approaches to Hedonic Price Measurement," *Journal of Economic and Social Measurement* 29(4), 427-443.
- de Leeuw F. (1993), "A Price Index for New Multifamily Housing," *Survey of Current Business* 73, (February), 33-42.
- de Vries P., J. de Haan, E. van der Wal and G. Marién (2009), "A House Price Index Based on the SPAR Method," *Journal of Housing Economics*, forthcoming.
- Denk R. (2006), "Home Ownership: The Engine of Wealth Accumulation," National Association of Home Builders, Report Posted on July 17, 2006.
- Diewert W. E. (1976), "Exact and superlative index numbers," *Journal of Econometrics* 4, 115-145.
- Diewert W. E. (2001), "Hedonic Regressions: A Consumer Theory Approach," Discussion Paper 01-12, Department of Economics, University of British Columbia.
- Diewert W. E. (2002), "Similarity and Dissimilarity Indexes: An Axiomatic Approach," Discussion Paper 02-10, Department of Economics, University of British Columbia.



- Diewert W. E. (2003), "Hedonic Regressions: A Review of Some Unresolved Issues," Mimeo.
- Diewert W. E. (2007a), "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.
- Diewert W. E. (2007b), "Index Numbers," Discussion Paper 07-02, Department of Economics, University of British Columbia.
- Diewert W. E., J. de Haan and R. Hendriks (2010), "The Decomposition of a House Price Index into Land and Structure Components: A Hedonic Regression Approach," Discussion Paper 10-01, Department of Economics, University of British Columbia.
- Diewert W. E. and A. Nakamura (2009), "Accounting for Housing in a CPI," Working Paper No. 09-4. Research Department, Federal reserve bank of Philadelphia.
- Diewert W. E., A. Nakamura and L. I. Nakamura (2009), "The Housing Bubble and a New Approach to Accounting for Housing in a CPI," *Journal of Housing Economics*, forthcoming.
- Diewert W. E., S. Heravi and M. Silver (2007), "Hedonic Imputation Indexes versus Time Dummy Hedonic Indexes," *IMF Working Paper*, WP/07/234.
- Dubin R. A. (1988) "Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms," *Review of Economics and Statistics* 70, 466-474.
- Dubin R. A. (1998), "Predicting House Prices Using Multiple Listings Data," *Journal of Real Estate Finance and Economics* 17(1), 35-59.
- Duffy D. (2009), "Measuring House Price Change," Working Paper No. 291, Economic and Social Research Institute (ESRI), Dublin.
- Dulberger E. R. (1989), "The Application of a Hedonic Model to a Quality-Adjusted Price Index for Computer Processors," in D. W. Jorgenson and R. Landau (eds.), *Technology and Capital Formation*. Cambridge, MA: MIT Press, 37-75.
- Eichhorn W. and J. Voeller (1976), *Theory of the Price Index*, Lecture Notes in Economics and Mathematical Systems, Volume 140. Berlin: Springer-Verlag.
- Eurostat (In Progress), *Handbook on Residential Property Price Indices (RPPI)*, B. Balk (ed.), <<[http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/methodology/owner\\_occupied\\_housing\\_hpi/rppi\\_handbook](http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/methodology/owner_occupied_housing_hpi/rppi_handbook)>>.

- Ferri M. G. (1977), "An Application of Hedonic Index Methods to Monthly Changes in Housing Prices: 1965-1975," *American Real Estate and Urban Economics Association Journal* 5(4), 455-462.
- Fisher I. (1922), *The Making of Index Numbers*. Boston: Houghton-Mifflin.
- Fleming M. C. and J. G. Nellis (1985), "The Application of Hedonic Indexing Methods: A Study of House Prices in the United Kingdom," *Statistical Journal of the United Nations Economic Commission for Europe* 3, 249-270.
- Follain J. R. and S. Malpezzi (1980), "Estimates of Housing Inflation for Thirty-Nine SMSAs: An Alternative to the Consumer Price Index," *Annals of Regional Science* 14(3), 41-56.
- Garner T. I. and R. Verbrugge (2009), "Reconciling User Costs and Rental Equivalence: Evidence from the US Consumer Expenditure Survey," *Journal of Housing Economics*, forthcoming.
- Gatzlaff D. H. and D. R. Haurin (1998), "Sample Selection and Biases in Local House Value Indices," *Journal of Urban Economics* 43(2), 199-222.
- Gelfand A. E., S. K. Ghosh, J. R. Knight and C. F. Sirmans (1998), "Spatio-Temporal Modeling of Residential Sales Data," *Journal of Business and Economic Statistics* 16(3), 312-321.
- Gelfand A. E., M. D. Ecker, J. R. Knight and C. F. Sirmans (2004), "The Dynamics of Location in Home Price," *Journal of Real Estate Finance and Economics* 29(2), 149-166.
- Gibbons S. (2004), "The Costs of Urban Property Crime," *Economic Journal* 114, F441-F463.
- Gibbons S. and S. Machin (2003), "Valuing English Primary Schools," *Journal of Urban Economics* 53, 197-219.
- Giles D. E. A. (1982), "The Interpretation of Dummy Variables in Semi-Logarithmic Equations: Unbiased Estimation," *Economics Letters* 10(1-2), 77-79.
- Gillen K., T. Thibodeau and S. Wachter (2001), "Anisotropic Autocorrelation in House Prices," *Journal of Real Estate Finance and Economics* 23(1), 5-30.
- Gillingham R. F. (1975), "Place to Place Rent Comparisons Using Hedonic Quality Adjustment Techniques," *Annals of Economic and Social Measurement* 4, 153-174.

- Goetzmann W. (1992). "The Accuracy of Real Estate Indexes: Repeat Sales Estimators," *Journal of Real Estate Finance and Economics* 5, 5-53.
- Goldberger A. S. (1968), "The Interpretation and Estimation of Cobb-Douglas Functions," *Econometrica* 35, 464-472.
- Goodman A. C. (1978), "Hedonic Prices, Price Indices and Housing Markets," *Journal of Urban Economics* 5, 471-484.
- Gouriéroux C. and A. Laferrère (2009), "Managing Hedonic Housing Price Indexes: The French Experience," *Journal of Housing Economics*, forthcoming.
- Griliches Z. (1961), "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in *The Price Statistics of the Federal Government*, G. Stigler (chairman). Washington D.C.: Government Printing Office.
- Griliches Z. (1971), "Introduction: Hedonic Price Indexes Revisited," in Z. Griliches (ed), *Price Indexes and Quality Change*, 3-15. Cambridge MA: Harvard University Press.
- Halifax Financial Services (2007), "Household Wealth Has More Than Doubled in the Last Ten Years," Press Release.
- Halvorsen R. and H. O. Pollakowski (1981), "Choice of Functional Form for Hedonic Price Equations," *Journal of Urban Economics* 10, 37-49.
- Haughwout A., J. Orr and D. Bedoll (2008), "The Price of Land in the New York Metropolitan Area," Federal Reserve Bank of New York *Current Issues in Economics and Finance* 14 (3), April/May 1-7.
- Hill R. C., J. R. Knight and C. F. Sirmans (1997), "Estimating Capital Asset Price Indexes," *Review of Economics and Statistics* 79(2), 226-233.
- Hill R. J. (2004), "Constructing Price Indexes Across Space and Time: The Case of the European Union," *American Economic Review* 94(5), 1379-1410.
- Hill R. J. and D. Melser (2008), "Hedonic Imputation and the Price Index Problem: An Application to Housing," *Economic Inquiry* 46(4), 593-609.
- Hill R. J. and D. Melser (2009), "Hedonic Time-Dummy and Region-Dummy Price Indexes for Housing: The Problem of Substitution Bias," Mimeo.
- Hill R. J. D. Melser and I. Syed (2009), "Measuring a Boom and Bust: The Sydney Housing Market 2001-2006," *Journal of Housing Economics* 18(3), forthcoming.
- Hill R. J., D. Melser and B. Reid (2010), "Hedonic Imputation with Geospatial Data: An Application of Splines to the Housing Market," Mimeo.

- Hoffman J. and A. Lorenz (2006), "Real Estate Price Indices for Germany: Past, Present and Future," Paper prepared for the OECD-IMF Workshop on real estate price indexes, Paris, 6-7 November 2006. Revised draft of 30 November 2006.
- Hulten C. R. (2003), "Price Hedonics: A Critical Review," *Federal Reserve Bank of New York Economic Policy Review*, September, 5-15.
- Jansen S. J. T., P. de Vries , H. C. C. H. Coolen , C. J. M. Lamain and P. J. Boelhouwer (2008), "Developing a House Price Index for The Netherlands: A Practical Application of Weighted Repeat Sales," *Journal of Real Estate Finance and Economics* 37(2), 163-186.
- Johnson D. S., S. B. Reed and K. J. Stewart (2006), "Price measurement in the United States: a decade after the Boskin Report," *Monthly Labor Review*, May, 10-19.
- Kagie M. and M. van Wetzel (2008), "Hedonic Price Models and Indices Based on Boosting Applied to the Dutch Housing Market," *Intelligent Systems in Accounting, Finance and Management* 15, 85106.
- Kain J. F. and J. M. Quigley (1970), "Measuring the Value of Housing Quality," *Journal of the American Statistical Association* 65, 532-48.
- Kawaguchi Y. (2007), *Property Derivatives Study Report*. Report of the Property Derivatives Study Group chaired by Y. Kawaguchi and established by Nomura Research Institute, Ltd., as the Secretariat under Consignment by the Ministry of Land, Infrastructure and Transport, Japan.
- Kelejian H. H. and D. P. Robinson (1995), "Spatial Correlation: A Suggestive Alternative to the Autoregressive Model," In L. Anselin and R. J. G. M. Florax (eds.), *New Directions in Spatial Econometrics*. Berlin: Springer.
- Kennedy P. E. (1981), "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations," *American Economic Review* 71 (4), 801.
- Kiel K. and J. Zabel (2000), "Estimating the Demand for Air Quality in Four Cities in the United States," *Land Economics* 78, 174-194.
- Lancaster K. J. (1966), A New Approach to Consumer Theory, *Journal of Political Economy* 74, 13257.
- LeSage J. P. (1998), *Spatial Econometrics*, Department of Economics, University of Toledo, Ohio.

- LeSage and Pace (2004), "Models for Spatially Dependent Missing Data," *Journal of Real Estate Finance and Economics* 29(2), 233-254.
- Leventis A. (2008), "Revisiting the Difference between the OFHEO and S&P/Case-Shiller House Price Indexes: New Explanations," Office of Federal Housing Enterprise Oversight.
- Linneman P. (1980), "Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing Market," *Journal of Urban Economics* 8, 47-68.
- Maclennan D. (1977), "Some Thoughts on the Nature and Purpose of Hedonic Price Functions," *Urban Studies* 14, 59-71.
- Malpezzi S. (2003), "Hedonic Pricing Models: A Selective and Applied Review," in A. OSullivan and K. Gibb (eds.) *Housing Economics: Essays in Honor of Duncan Maclennan*, 67-89. Blackwell: Malder MA.
- Malpezzi S., G. H. Chun, R. K. Green (1998) "New Place-to-Place Housing Price Indexes for U.S. Metropolitan Areas, and Their Determinants," *Real Estate Economics* 26(2), 235-274.
- Mark J. H. and M. A. Goldberg (1984), "Alternative Housing Price Indices: An Evaluation," *American Real Estate and Urban Economics Association Journal* 12(1), 30-49, 1984.
- Martins-Filho C and O. Bin (2005), "Estimation of Hedonic Price Functions Via Additive Nonparametric Regression," *Empirical Economics* 30(1), 93114.
- McMillen D. (2004), "Airport Expansions and Property Values: The Case of Chicago OHare Airport," *Journal of Urban Economics* 55, 627640.
- Meese R. A. and N. E. Wallace (1991), "Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indices," *American Real Estate and Urban Economics Association Journal* 19(3), 308-332.
- Meese R. A. and N. E. Wallace (1997), "The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches," *Journal of Real Estate Finance and Economics* 14, 5173.
- Moulton B. R. (1995), "Interarea Indexes of the Cost of Shelter Using Hedonic Quality Adjustment Techniques," *Journal of Econometrics* 68, 181204.
- Nappi-Choulet I. and T. Maury (2009), "A Spatiotemporal Autoregressive Price Index for the Paris Office Property Market," *Real Estate Economics* 37(2), 305340.

- Naroff J., D. Hellman and D. Skinner (2006), "Estimates of the Impact of Crime on Property Values," *Growth and Change* 11, 24-30.
- Oates W. (1969), "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis," *Journal of Political Economy* 77(6), 957-971.
- Pace R. K. (1993), "Nonparametric Methods with Applications to Hedonic Models," *Journal of Real Estate Finance and Economics* 7(3), 185-204.
- Pace R. K., R. Barry, J. M. Clapp and M. Rodriguez (1998), "Spatiotemporal autoregressive models of neighborhood effects," *Journal of Real Estate Finance and Economics* 17(1), 15-33.
- Pace R. K. and O. Gilley (1997), "Using Spatial Configuration of the Data to Improve Estimation," *Journal of Real Estate Finance and Economics* 14(3), 333-340.
- Pakes A. (2003), "A Reconsideration of Hedonic Price Indices with an Application to PC's," *American Economic Review* 93(5), 1578-1596.
- Palmquist R. B. (1980), "Alternative Techniques for Developing Real Estate Price Indexes," *Review of Economics and Statistics* 62(3), 442-448.
- Pavlin B. (2006), Dwelling price index in Slovenia pilot study of hedonic approaches, Paper Presented at the OECD-IMF Workshop on Real Estate Price Indexes, Paris, 6-7 November 2006.
- Poole R., F. Ptacek and R. Verbrugge (2005), "Treatment of Owner-Occupied Housing in the CPI." Paper presented to the Federal Economic Statistics Advisory Committee (FESAC) on Dec 9, 2005.
- Quigley J. (1995), "A Simple Hybrid Model for Estimating Real Estate Price Indexes," *Journal of Housing Economics* 4(1), 1-12.
- Prasad N. and A. Richards (2006), "Measuring Housing Price Growth – Using Stratification to Improve Median-based Measures," Research Discussion Paper 2006-04, Reserve Bank of Australia.
- Reid B. (2007). Hedonic Imputation House Price Indexes: Bias and Other Issues. Honours Thesis, School of Economics, University of New South Wales, Sydney, Australia.
- Ribe M. (2009), *House Prices in a Swedish CPI Perspective*, Statistics Sweden. Paper Presented at the 11th Ottawa Group Meeting in Neuchâtel, 27-29 May, 2009.

- Ridker R. G. and J. A. Henning (1967), "The Determinants of Residential Property Values with Special Reference to Air Pollution," *Review of Economics and Statistics* May, 246-257.
- Rosen S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy* 82(1), 34-55.
- Rouwendal J. and J. W. van der Straaten (2008), "The Costs and Benefits of Providing Open Space in Cities," Tinbergen Institute, Tinbergen Institute Discussion Papers 08-001/3.
- Rubin D. B. (1976), "Inference and Missing Data," *Biometrika* 63, 581-592.
- Rubin D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Saarnio M. (2006), "Housing Price Statistics at Statistics Finland," Paper presented at the OECD-IMF Workshop on Real Estate Price Indexes, Paris, 6-7 November 2006.
- Schreyer P. (2009), "User Costs and Bubbles in Land Markets," *Journal of Housing Economics*, forthcoming.
- Schultze C. S., and C. Mackie (2002), At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes, in *Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of-Living Indexes*. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC, National Academy Press.
- Shiller R. J. (1993), "Measuring Asset Values for Cash Settlement in Derivative Markets: Hedonic Repeated Measures Indices and Perpetual Futures," *Journal of Finance* 48, 911-931.
- Shiller R. J. (2007), "Understanding Recent Trends in House Prices and Home Ownership," Cowles Foundation Discussion Paper No. 1630, Yale University.
- Shiller R. J. (2008), "Derivatives Markets for Home Prices," Cowles Foundation Discussion Paper No. 1648, Yale University.
- Shimizu C., K. G. Nishimura and T. Watanabe (2010), "Housing Prices in Tokyo: A Comparison of Hedonic and Repeat-Sales Measures," Research Center for Price Dynamics Institute of Economic Research, Hitotsubashi University, Japan.
- 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.

- Sirmans S., L. MacDonald, D. Macpherson and E. Zietz (2006), "The Value of Housing Characteristics: A Meta Analysis," *The Journal of Real Estate Finance and Economics* 33, 215-240.
- Song Y. and G. Knaap (2004), "Measuring the Effects of Mixed Land Uses on Housing Values," *Regional Science and Urban Economics* 34(6), Special Issue, 663-680.
- Standard and Poor's (2008), *S&P/Case-Shiller Home Price Indices Index Methodology*. New York: Standard and Poor's.
- Steele M. and R. Goy (1997), "Short Holds, the Distribution of First and Second Sales, and Bias in the Repeat-Sales Price Index," *Journal of Real Estate Finance and Economics* 14(1-2), 133-154.
- Stone R. (1954), *The Measurement of Consumer Behaviour and Expenditure in the United Kingdom, 1920-1938*, Vol.1, Studies in the National Income and Expenditure of the United Kingdom (assisted by D. R. Rowe, W. J. Corlett, R. Hurstfield, and M. Potter). Cambridge: Cambridge University Press.
- Syed I., R. J. Hill and D. Melsner (2008), *Flexible Spatial and Temporal Hedonic Price Indexes for Housing in the Presence of Missing Data*, Discussion Paper 2008/14, School of Economics, University of New South Wales, Sydney, Australia.
- Syz J. (2008), *Property Derivatives: Pricing, Hedging and Applications*. Chichester, England: John Wiley & Sons.
- Syz J., P. Vanini and M. Salvi (2008), "Property Derivatives and Index-Linked Mortgages," *Journal of Real Estate Finance and Economics* 36, 2335.
- Thomassen A. (2007), *Price Index for New Multidwelling Houses: Sources and Methods*, Statistics Norway/Department of Industry Statistics/Construction and Service Statistics, Document 2007/9.
- Triplett J. E. (1969), "Automobiles and hedonic Quality Adjustment," *Journal of Political Economy* 77(3), 408-417.
- 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.
- van Garderen K. J. and C. Shah (2002), "Exact Interpretation of Dummy Variables in Semilogarithmic Equations" *Econometrics Journal* 5, 149-159.



- Verbrugge R. (2008), "The Puzzling Divergence of Rents and User Costs," *Review of Income and Wealth* 54(4), 671-699.
- von Auer L. and J. E. Brennan (2007), "Bias and Inefficiency in Quality-Adjusted Hedonic Regression Analysis," *Applied Economics* 39(1-3), 95-107.
- Wallace N. E. (1996), "Hedonic-Based Price Indexes for Housing: Theory, Estimation, and Index Construction," *Federal Reserve Bank of San Francisco Economic Review* 3, 34-48.
- Wasshausen D. and B. R. Moulton (2006), "The Role of Hedonic Methods in Measuring Real GDP in the United States," BEA Papers.
- Waugh F. V. (1928), "Quality Factors Influencing Vegetable Prices," *Journal of Farm Economics* 10, April, 185-196.
- Wenzlick R. (1952), "As I See the Fluctuations in the Selling Prices of Single-Family Residences," *The Real Estate Analyst* 21, 541-548.
- Wyngarden H. (1927), "An Index of Local Real Estate Prices," *Michigan Business Studies* 1(2), Ann Arbor, University of Michigan.
- US Census Bureau (undated), Description of Price Index for Sales Price of New One-Family Houses Sold. <http://www.census.gov/const/C25/newresindextext.html>
- Zellner, Arnold (1962), "An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias," *Journal of the American Statistical Association* 57, 348-368.