EUROPEAN COMMISSION
EUROSTAT

Directorate G: Business Statistics
Unit G-6: Price statistics; Purchasing Power Parities

HARMONIZED INDICES OF CONSUMER PRICES (HICPs)

DRAFT
HANDBOOK ON RESIDENTIAL PROPERTY PRICE INDICES

This present version is a second version draft and has been posted so that interested parties may make comments. The work is carried out by Statistics Netherlands with the involvement of numerous authors under a contract with Eurostat. The present version therefore reflects these authors' personal opinions and does not necessarily reflect the views of Eurostat or its staff.

All comments should be sent to the following e-mail address: ESTAT-hicp-methods@ec.europa.eu and also copied to Professor Bert Balk at the following e-mail address: bbalk@rsm.nl.

Draft Version 3
April 2011

Handbook on Residential Property Price Indices

Table of Contents

Chapter 2: Introduction

Residential property is both a source of wealth and, insofar as property owners live in or on their property, an important determining factor in their cost of living. The price of a house is something different from the cost of dwelling services it provides, though the two concepts are obviously interlinked.

Monitoring the development of house prices is considered important, especially in times of economic turbulence. Yet the way house price development is measured varies per country, and even within a country there are sometimes two or more competing methods in use. This situation is of course not favourable for the design of consistent policy measures based on solid international comparisons.

Against this background it is understandable that it was proposed that a handbook be prepared on housing, or broader residential property, price indexes.[1] The primary goals of the handbook are

- to provide guidance for those wishing to set up residential property price indexes or modify existing indexes in view of international harmonisation;
- to provide a discussion and comparison of the various targets and their corresponding conceptual frameworks;
- to provide an inventory of existing practices.

The contents of the handbook is briefly outlined below.

Chapter 3 reviews a number of areas where residential property price indexes (RPPIs) play a role. The following applications are considered:

- as a macro-economic indicator of economic activity;
- for use in monetary policy and inflation targeting;
- as a component of wealth;
- as a financial stability or soundness indicator to measure risk exposure;
- as a deflator in the Sytem of National Accounts;
- as an input into citizens' decision making on whether to buy or sell residential property;
- as an input into the Consumer Price Index;
- for use in making interarea and international comparisons.

[1] Actually this was one of the conclusions of the OECD-IMF Workshop on Real Estate Price Indexes (Paris, 6-7 November 2006).

In *Chapter 4* on the uses of a RPPI the focus will be to fill in gaps in the System of National Accounts and in the compilation of a Consumer Price Index. It is likely that if appropriate RPPIs can be constructed to fill in these gaps, then the resulting family of RPPIs will meet the needs of most users.

Broadly speaking, two separate types of RPPI can be distinguished: a constant quality price index for the *stock* of residential housing at a particular moment in time and a constant quality price index for residential property *sales* that took place during a particular period of time. The construction of these two types of index will be different; most particularly, the weighting associated with the two types will differ.

The chapter continues by summarizing the four main approaches to constructing a RPPI. In the final sections a number of miscellaneous topics are addressed, such as the frequency of an RPPI, the consistency of monthly with quarterly estimates and the consistency of quarterly with annual estimates, revision policies, and seasonal adjustment.

Chapters 5-8 review in depth the main methods for compiling RPPIs. The simplest house price indexes are based on some measure of central tendency of the distribution of transaction prices in a period, in particular the mean or the median. Since house price distributions are generally positively skewed (predominantly reflecting the heterogeneous nature of housing, the positive skew in income distributions, and the zero lower bound on transaction prices), the median rather than the mean is often used. As no data on housing characteristics are required to calculate the median, a price index that tracks changes in the price of the median house sold from one period to the next can be easily constructed. Another attraction of median indexes is that they are easy to understand.

The main drawback of simple median based indexes is that they provide very noisy estimates of price change. The set of houses actually traded in a period, or a sample thereof, is typically small and not necessarily representative of the total stock of houses. Changes in the mix of properties sold will therefore affect the sample median price much more than the median price of the housing stock. A perhaps bigger problem than short-term noise is systematic error, or bias. A simple median index will be subject to bias when the quality of the housing stock changes over time. Bias can also arise if certain types of houses are sold more frequently than other types of houses and at the same time exhibit different price changes.

A general technique for reducing sample selection bias is (post-) stratification. This technique, which is also known as mix adjustment, is discussed in *Chapter 5*.

Chapter 6 continues by reviewing the hedonic regression approach. This approach recognizes that heterogeneous goods can be described by their attributes or characteristics. That is, each good is essentially a bundle of performance characteristics. In the housing context, this bundle may contain attributes of both the structure and the location of the properties. Although there is no market for characteristics, since they cannot be sold separately, the demand and supply for the properties implicitly determine the characteristics' marginal contributions to the prices of the properties. Regression techniques can be used to estimate those marginal contributions or implicit prices.

This chapter reviews, in a non-technical way, the main models used as well as the methods to form RPPIs from estimation of such models. The overall evaluation of the hedonic regression method is that it is probably the best method that could be used in order to construct constant quality RPPIs for various types of residential property. However, it is also the most data-intensive method.

The repeat sales method, reviewed in *Chapter 7*, utilizes information on the same properties which have been sold more than once. Because only "matched models" are used, there is no change in the quality mix to control for. In its basic form, the only information required is price, sales date and address of the property. So the repeat sales method is much less data-intensive than hedonic methods. Also, the repeat sales method will automatically control for micro location (address), something which hedonic methods are unable to do.

The matched model methodology, where prices of exactly the same item are compared over time, is the natural starting point for the construction of any price index. Because of the low incidence of transactions, and because the quality of houses continually changes, the standard matched model methodology cannot be applied straightforwardly. The repeat sales method attempts to deal with this issue by looking only at properties that have been sold more than once over a sample period. This, however, can lead to a relatively low number of observations and to sample selection bias. To overcome such problems, assessed values of the properties could be used.

In many countries, official government assessments are available for all properties, because such data are needed for taxation. If the assessments pertain to some reference date, an RPPI can be constructed by relating actual sale prices to assessed values. This constitutes a variant of the matched model methodology, the distinct feature being that compositional change over time is accounted for. In this case, there is no need to use econometric techniques. The various assessment-based methods, and in particular the sale-price appraisal ratio (SPAR) method, are reviewed in *Chapter 8*.

Chapters 5-8 all end with empirical examples tested on a real-life data set in order to illustrate the methods discussed and to provide additional background material. The data set covers 14 quarters of residential property sales for a relatively small town in the Netherlands. As will become clear in Chapters 5-8, most methods are unable to decompose an RPPI into a land and a structures component. *Chapter 9* discusses how hedonic regressions method can be used to obtain such a decomposition and considers how to construct an RPPI for the *stock* of housing when hedonic regression methods are exploited. Using the real-life data set, this chapter also suggests ways to overcome several practical problems that are encountered, such as a high correlation between the size of the structure and the size of the land.

In practice, because of the high cost of undertaking purpose-designed surveys of house prices, the approaches adopted by statistical agencies and others to construct RPPIs have been mainly a function of the house price data sets generated by the legal and other processes associated with buying a house. The indexes so constructed can vary according to the point in the house purchasing process at which the price is measured, for instance whether the final transaction price or the earlier valuation used for securing a loan is taken. Also, the amount of detailed information available on the characteristics of the properties sold will affect index compilation methods, often acting as a constraint on the techniques available to quality adjust for houses of different sizes and locations. Thus data sets have historically acted as a constraint on index construction.

Chapter 10 qualitatively examines the different data sources that can be used for constructing RPPIs, such as newspapers, real estate agents, mortgage companies, property registers, and tax offices. In the final section attention is paid to the situation in many developing countries where data are scarce and ownership is a fuzzy notion.

Chapter 11 catalogues the availability of RPPIs in different countries and also presents some case studies. It relies on meta-data gathered by various organisations, including the European Central Bank and the Bank for International Settlements, and more recently a fact-finding exercise conducted by Eurostat in connection with the inclusion of owner-occupied housing costs in the European Union’s Harmonised Index of Consumer Prices, which was extended to cover some non-EU countries.

Chapter 12 brings together the RPPI construction methods that were outlined in Chapters 5-8 and provides additional empirical examples. The chapter first demonstrates the working of the methods on simple examples using small data sets. In the final section, an extensive, realistic data set is used to compute the various indexes and an evaluation of the methods is provided. In all, some 15 RPPIs are presented in this last section.

Chapter 13 concludes by providing recommendations.

Chapter 3: Uses of Residential Property Price Indexes

3.1 Introduction

There are many areas of society where individuals or organisations use residential property price indexes (RPPIs) directly or indirectly either to influence practical decision making or to inform the formulation and conduct of economic policy. Different uses can have a significant impact on the preferred coverage of the index and also on the appropriate methodology for its construction.

From an individual household's perspective, real estate often represents the single largest investment in their portfolio. It also accounts for the largest share of wealth in most nations' balance sheet. Changes in house prices can have far-reaching implications for individuals. For example, changes in housing equity and household debt levels can permeate through to the overall economy. In fact, consumer spending is often affected by changes in house prices as a result of wealth effects and its effect on consumer confidence. House prices influence home improvement and renovations expenditures, which in many countries are higher than overall spending on new house construction. House prices play a key role in the measurement of the affordability of home-ownership, a key housing policy objective in some countries. House price changes also influence the decision to build new houses (the supply side) as well as the decision to become a homeowner (the demand side).[1] Investors turn to house price indexes to not only measure wealth but also to help in assessing current and future rates of return.[2]

From a broader perspective, analysts, policymakers, and financial institutions follow trends in house prices to expand their understanding of real estate and credit market conditions as well as to monitor the impact on economic activity, and financial stability and soundness.[3] For instance, mortgage lenders will use information on house price inflation to gauge default risk. Central banks often rely on movements in house price indexes to monitor households' borrowing capacity and debt burden[4] and their effects on aggregate consumption.[5]

In this context it should be emphasised that the different uses of residential property price indexes may require different conceptual bases and methodology, although in practice, other factors sometimes come into play, such as data availability.[6] In general, no single indicator of

---

[1] See Duffy (2009).

[2] Residential construction investment accounts for about 5% of GDP in the euro area.

[3] See Case and Wachter (2005).

[4] See Finocchiaro and von Heideken(2007).

[5] See Case *et al.* (2001), Phang(2004) and Belsky and Prakken (2004).

[6] See Fenwick (2006) and also Chapter 10.

house price change can satisfy every purpose. For instance, the price dynamics of the housing market for monitoring house inflation, as experienced by purchasers, may best be estimated by collecting information on current transaction prices and using this information to construct a *price index for the sales of housing units*. In contrast, to estimate an economy's (real) stock of wealth, information on the sample of transacted dwellings must ideally be supplemented by information on the stock of non-transacted dwellings in order to construct a *price index for the housing stock*. This may be done by re-weighting to reflect the different mix of houses in the housing stock compared with transactions but the adequacy of this method depends on whether the dwellings that are actually transacted can act as a proxy for the ones that have not been subject to a change of ownership. If the prices of houses that have not changed ownership are not available and there is either no or very limited information on their numbers and characteristics, the user needs to be assured that the profile of the transactions is representative of the overall housing stock. In practice, the latter condition may not be fully met as different sectors of the housing market can be influenced by different factors and a sparseness of transactions may lead to unreliable or non-existent price data for some of these different strata.

The (price determining) attributes of individual houses often change over time. These changes include improvements to the dwelling in the form of renovations to kitchens and bathrooms, replacement windows with insulated glazing, or the installation of energy efficient heating or air-conditioning systems, and also extensions of the structure which reflect the trend in many countries towards larger houses. Improvements and extensions will be partially offset by depreciation of the structures. Irrespective of the purpose of the index, an RPPI should be adjusted for all of those changes. To put it differently, the index should represent changes in the prices of properties that are comparable in quality over time.

But the need for quality adjustment extends beyond controlling for home improvements and depreciation. The quality mix of dwellings that are sold in one month is likely to be different from that in the next month because, say, more larger houses are sold. Such quality mix changes may have a cyclical pattern as sales of larger houses decline as an economy enters a recession. Quality mix changes, just like quality changes of individual dwellings, should not be measured as price changes – measurement techniques are required to adjust the price changes for quality mix changes. A short overview of the various methods that are available to solve the problems of quality (mix) change will be provided in Chapter 4. A detailed discussion of these methods will follow in Chapters 5-9.

3.2 A Review of the Different Uses of Residential Property Price Indexes

Residential property price indexes have a number of important uses:

- as a macro-economic indicator of economic growth;

- for use in monetary policy and inflation targeting;
- as a component of wealth;
- as a financial stability or soundness indicator to measure risk exposure;
- as a deflator in the System of National Accounts;
- as an input into an individual citizen's decision making on whether to buy (or sell) a residential property;
- as an input into the Consumer Price Index (CPI), which in turn is used for wage bargaining and indexation purposes;
- for use in making inter-area and international comparisons.

Each use is considered in turn.

3.2.1 As a Macro-Economic Indicator of Economic Growth

Rising house prices are often associated with periods of economic expansion while falling house prices often correspond with a slowing economy. Goodhart and Hofmann (2006) show that for 16 industrialised countries there exists a strong correlation between house prices and economic activity. In fact the six major banking crises in advanced countries since the mid 1970s were all associated with the bursting of a housing bubble (Reinhart and Rogoff, 2009).[7] In the main, house prices are treated as a leading indicator although there is some debate about whether house price change is a leading, lagging or coincident economic indicator.

What is clear is that rising house prices are often associated with economic growth through at least three channels:

- Higher (relative) house prices tend to *stimulate increased construction activity*, which in turn leads to higher employment and higher incomes for a wide range of workers involved in the housing market, such as real estate agents, construction workers and professionals in the financial and the legal professions. Expectations of higher future returns on property investment lead builders to start new construction and this is accompanied by higher market demand in property-related sectors from owner-occupiers and property investors.[8] In addition, building activity will tend to increase from more home renovations.

[7] Claessens, Kose and Terrones (2008; 25) find that "... recessions associated with house price busts are on average over a quarter longer than those without busts. Moreover, output declines (and corresponding cumulative losses) are typically much larger in recessions with busts, 2.2 (3.7) percent versus 1.5 (2.3) percent in those without busts. These sizeable differences also extend to the other macroeconomic variables, including consumption, investment and the unemployment rate."

[8] See Zhu (2005).

- Higher house prices tend to lead to *increased sales of existing housing units* and this in turn can lead to additional tax revenues in the form of property transfer taxes generated from the higher volume and value of property sales. These *increased tax revenues* can lead to increased government spending which in turn provides additional economic stimulus.
- Rising real estate prices will lead to improvements in the household sector's balance sheet (*the wealth effect*) and this in turn will generally lead to increased household spending on consumption and investment.[9] According to a report by the U.S. Congressional Budget Office (2007), when house prices surged in the 1990s and 2000s, consumer spending grew faster than incomes. This household wealth effect generally leads to increases in spending by consumers on home renovations and repairs in addition to increased spending on other goods and services.

Of course, the above stimulative effects of increasing house prices go into reverse when (real) house prices fall. It is therefore important that the public and economic policy makers have at their disposal accurate and timely information on movements in real estate prices.

Asset prices, including real estate prices, are a key indicator for more fully understanding the dynamics of the economy.[10] According to Plosser (2007), asset prices contain important information about the current and future state of the economy and can play an important role in the deliberations of central bankers as they seek to achieve their objectives of price stability and sustainable output growth.

3.2.2 House Prices and Monetary Policy and Inflation Targeting

In addition to the above general interest in monitoring property prices, many central banks have *inflation targets* which can directly involve indexes of property prices. For instance, central banks in some countries utilize a Monetary Conditions Index (MCI) as a day-to-day operating target for the conduct of monetary policy. In an expanded version of this index, as that suggested by Jarociński and Smets (2008) and Goodhart and Hofmann (2007), the MCI would include some measure of house prices because of the important role this variable plays in the inflationary process and for economic performance. Other central banks who have an inflation target based on the Consumer Price Index (CPI) will indirectly take into account the movement in house prices when setting interest rates, depending in part on the treatment of Owner Occupied Housing (OOH) in their country's CPI. This issue is discussed further in section 3.2.7 and in Chapter 4.

[9] See Campbell and Cocco (2007).

[10] See Turvey (1989) and Goodhart (2001).

It can be argued that in the future, residential property prices are likely to play an increasing role in the conduct of monetary policy. Over recent years an inflation target has been used by a growing number of countries to define and operate their monetary policy frameworks. The IMF (2007) provides a list of 28 countries classified as inflation "targeters" according to their "exchange rate arrangements" (without specifying the target or inflation measure). Carare and Stone (2003) extend this analysis further by classifying countries that use an inflation target for monetary policy, into fully-fledged inflation "targeters", eclectic "targeters" and inflation targeting lite regimes, using the clarity and credibility[11] of the commitment to the inflation target to classify individual countries. The authors then identify 42 medium and large country central banks who have some form of floating exchange rate mechanism (i.e. not adopting a fixed exchange rate) leaving their degree of commitment to an inflation target as the defining monetary objective. They estimated that by 2001 some 7 industrial and 11 emerging markets operated fully-fledged inflation targeting, that is "have a medium to high level of credibility, clearly commit to their inflation target, and institutionalize this commitment in the form of a transparent monetary framework that fosters accountability of the central bank to the target". The number of countries operating fully-fledged inflation targeting has been increasing.

3.2.3 As a Component of Wealth

House prices are an input into the measurement of aggregate wealth in the economy. Existing dwelling units are part of the Balance Sheet accounts in the SNA and thus it is necessary to have a price index for this asset class in order to form estimates of real household wealth. As was mentioned in section 3.2.1 above, rising house prices will generate a wealth effect that can lead to increases in consumption and increased household borrowing.

More generally, individuals will have an indirect stake in real estate asset prices, including residential property, through pension funds and other direct investments in housing properties.

3.2.4 As a Financial Stability or Soundness Indicator to Measure Risk Exposure

Financial Soundness Indicators (FSIs) are indicators of the current health and soundness of the financial system and institutions of a country and of their corporate and household components. They include both aggregated individual institution data and indicators that are representative of the markets in which the financial institutions operate, including statistics on real estate prices. FSIs are calculated and disseminated for the purpose of supporting national and international surveillance of financial systems. The IMF developed FSIs with a view to monitoring and strengthening the global financial system and to increasing stability following

[11] Clarity is gauged by the public announcement of the inflation target and by the institutional arrangements for accountability. Credibility is measured indirectly using as a proxy the actual inflation outturn and by market ratings of long-term local currency government debt.

the financial market crises of the late 1990s, and as a way of combating the subsequent growing number of banking crises that have occurred globally. The compilation guide for financial soundness indicators provides some advice on compiling house price indices whilst acknowledging the relative absence of international experience and guidance and the absence of a comprehensive framework for constructing such indexes. More recently, the October 2009 Report to the G-20 Finance Ministers and Central Bank Governors on the *Financial Crisis and Information Gaps*[12] mentions that information on dwellings and their associated price changes are critical ingredients for financial stability policy analysis.

Sharp falls in real estate prices have a detrimental impact on the health and soundness of the financial sector and on the financial situation of individuals and of individual households, by affecting credit ratings, the value of collateral and the debt to equity ratio.

It should come as no surprise that the relationship between real estate cycles and economic cycles has been well documented and that the role of real estate prices in debt finance and financial crises has long been recognised. This has lead to the use of residential property price indexes as indicators of financial stability, particularly in countries where real estate accounts for a significant proportion of national and household wealth and where the propensity of home ownership is relatively high.

The use of trends in residential property prices, and real estate prices more generally, as an indicator of financial soundness, has been supported by in-depth analytical studies. Included amongst the vast amount of material published on this subject is a paper by Nabarro and Key (2003) who present a model for real estate and lending cycles, supported by case studies. Their paper traces the cyclical evolution from initial indicators provided by the rental market, to property prices and through to balance sheets of borrowers and lenders, and draws attention to a number of relevant indicators of the real estate market. It describes what the authors call "the dangerous interdependence between real estate cycles and financial systems". Whilst the authors acknowledge the highly unpredictable nature of the real estate cycle and its different characteristics and properties from one cycle to the next, they discuss the linkages between real estate cycles and debt finance to identify areas where improved information could support effective counteracting strategies and policies. They explain how a reliable and cost-effective system of performance measurement and monitoring can be developed and implemented and suggest how such a system can provide a mechanism for analytical decision making, designed to impact upon the behaviour of the real estate sector.

Information on residential property and other property prices needs to be supplemented by relevant and timely detailed analyses, and by other information such as the proportion of houses being purchased with cash rather than being financed through a loan. The average ratio

[12] Available at: http://www.imf.org/external/np/seminars/eng/2010/infogaps/index.htm.

of loan to property price, and how this is distributed, provides an indication of the exposure of the borrower and the lender, as does the price to earnings ratio and, to a certain extent, the volume of transactions.[13] Similarly, a more detailed analysis of the types of houses being sold by region will show whether activity in the housing market is concentrated in particular segments of the market such as high-end properties or in certain geographical locations such as the capital city or large urban areas.

3.2.5 As a Deflator in the System of National Accounts

National statistical agencies use house price indexes in at least two ways. First, the structures component of a price index for newly-built houses is often used to deflate current price values for residential construction; see Bover and Izquierdo (2003). Second, house price indexes may be included in the construction of the CPI, depending on the choice of its conceptual basis. This second use is considered below in section 3.2.7 and in more detail in Chapter 4.

Price indexes and deflators are seemingly different entities within a wider group of statistics relating to prices.[14] It is pertinent to note against this background that two of the most recent and widely available references on the compilation and use of national accounts deflators, SNA (1993) and the Eurostat (2001) *Handbook on Price and Volume Measures in National Accounts,* pre-date the CPI Manual (2004) and PPI Manual (2004).

The CPI and PPI Manuals were developed in parallel and take advantage of the latest research into index number theory and practice, which is not fully reflected in the official literature on national accounts.[15] The two Manuals are essentially based on the same underlying economic principles and statistical theory. They provide a comprehensive and coherent overview of the conceptual and theoretical issues associated with consumer and producer price indexes and translate these into available options for practical measurement. The CPI Manual also acted as a catalyst for the new ILO Resolution on Consumer Price Indices, which was passed in 2003.

3.2.6 As an Input into an Individual Citizen's Decision Making on Whether to Buy (or sell) a Residential Property

---

[13] Past observation suggests that when price-to-earnings ratios get to an unsustainable high level, the adjustment is initially seen in a reduction in the volume of housing turnover rather than in transaction prices.

[14] However, the underlying theory of deflators and (direct) price indexes is the same; see Chapter 16 in SNA (1993). Samuelson and Swamy (1974) note the following: "Although most attention in the literature is devoted to price indices…. Once somehow estimated, price indices are in fact used, if at all, primarily to deflate nominal or monetary totals in order to arrive at estimates of underlying "real" magnitudes".

[15] However, the CPI and PPI Manuals are consistent with the material in Chapter 16 of SNA (1993) and also with the 2008 *System of National Accounts*. The newer Manuals delve deeper into the problems associated with the construction of price indexes, particularly at lower levels of aggregation.

The buying or selling of a dwelling is typically the largest financial transaction a household will enter into during his or her life. Changes in house prices are therefore likely to influence substantially whether a household purchases a property and also the budget plans and savings decisions of the prospective house buyers and sellers. The purchase of a house is considered by many owner-occupiers both as a means of obtaining shelter services and as a capital investment, the latter potentially providing an opportunity for significant capital gains in the longer-term. Current price levels and trends, together with expectations about future trends in house prices and mortgage interest rates,[16] will influence an individual's decision on whether to purchase now or postpone the purchase. The opportunity cost associated with the sums of money involved will also come into play as prospective purchasers evaluate the alternative choices available to them. For instance, prospective purchasers will often take into account the impact of changes in house prices on market rents.

More generally, individuals also have an indirect stake in real estate asset prices through pension funds and other investments for which house prices will likely have an effect. For instance, the portfolios of some pension funds include apartment blocks whose rents provide an income and where a capital gain is expected to materialise from an increase in the property value.

3.2.7 As an Input into the Construction of a Consumer Price Index (CPI)

House prices will directly affect measured inflation when the CPI includes owner-occupier housing costs and the method of measurement draws on house prices as one of the inputs. Measured inflation is indirectly affected if house prices influence market rents, which constitute another element of a CPI, and where additionally imputed rents are used as a proxy for owner occupier housing costs. Renting and buying can be substitutes and the level of house prices will have an impact on the rate of return obtained by a landlord on his or her investment and also on the rent charged.

The treatment of Owner Occupied Housing (OOH) is one of the most difficult challenges faced by compilers of consumer price indexes. There are a number of alternative conceptual treatments and the choice between them can have a significant impact on the overall index, affecting both the weight attributed to OOH (and by implication to an RPPI) and the measured rate of inflation. In essence there are four possible main approaches to including OOH in a CPI: the acquisitions approach, the payments or money outlays approach, the user cost approach and the rental equivalence approach. The first three approaches require the construction of a housing price index. These various approaches to the treatment of OOH are reviewed in more detail in section 4.3 of Chapter 4.

[16] Interest rate policy will have an impact both on inflation and on net disposable income after the payment of interest.

3.2.8 International and Inter-area Comparisons

House price indexes are also used in conjunction with (comparable) benchmark data on house price levels across regions or countries to generate inter-area or international comparisons of living cost differentials. The problems that arise in attempting to price the services of OOH in a national context also arise in the context of inter-area and international comparisons. In the latter context, however, the problems are somewhat more difficult than making, say, national comparisons over time because inter-area/international comparisons require comparable types of housing across the regions/countries being compared (or comparable information on the characteristics of housing units across the regions if a hedonic regression technique is used) in order to construct a constant quality price index.

The European Central Bank (ECB) – in co-operation with the central banks of the individual countries of the euro-zone and the European Union – has an interest in comparative measurement of changes in residential property prices across different euro-area countries and for the euro-area as a whole. The raw data used here come from various national sources and have primarily been collected and documented by the Bank for International Settlements (BIS).[17] Since 2001, the ECB has compiled an aggregate for the euro-area by weighting together changes in prices for houses and flats for the euro-area countries.[18] The national methodologies associated with the figures available for each individual country and for the euro-area aggregate, have improved over recent years but perhaps fall short of the standards applied to other economic statistics and price indicators for the euro-area.[19] The BIS has also brought together residential property price statistics for the non-euro area countries of the European Union and has in many cases been confronted with the even more pronounced issues concerning data comparability and quality.

Such comparisons can be confounded by methodological and coverage differences and also by differences in the frequency and timeliness of the data. Some of these differences arise from the different sources of data used to compile national indexes. Chapter 10 explores these data sources in more detail and Chapter 11 gives an inventory of the different methods used by countries to compile their indexes of residential property prices. It can be observed that a notable proportion of countries, including some developed countries, do not have reliable residential property price indexes.

[17] The BIS data set of residential property price statistics is available at: http://www.bis.org/statistics/pp.htm.

[18] See box "Preliminary evidence on developments in euro area residential property prices" in the October-2001 issue of the ECB Monthly Bulletin.

[19] See Eiglsperger (2010), page 233.

Chapter 4: Elements for a Conceptual Framework

4.1 Introduction

What makes a residential property price index (RPPI) so difficult to construct? This question was addressed in Chapter 2 of this Handbook but it will be useful to remind readers about the main problems, which are as follows:

- The compilation of price indexes is typically based on *matching* the prices of identical items over time. However, in the housing context, each property has a unique location and usually a unique set of structural characteristics. Thus, the matched model methodology will be difficult or impossible to apply.
- Transactions are sporadic.
- The desired index number concept may not be clear, or put another way, there are *several distinct purposes* for which a RPPI is required and, broadly speaking, different purposes require different indexes.
- For some purposes, notably the construction of balance sheets in the national accounts and the estimation of user costs of owner occupied housing, a decomposition of a property price into land and structures components is required but it is not clear how to accomplish such a decomposition.

The first two difficulties are well recognized in the housing measurement literature as the following quotations indicate:

"The price of housing is harder to measure than that of most other goods and assets because of three key distinguishing characteristics. First, and most importantly, dwellings are heterogeneous. No two dwellings are identical, if only because they cannot occupy quite the same location. This means that sampled house prices may be a poor indicator of all house prices because we cannot always reliably predict the sales price of a given dwelling from the price of another." Robert Wood (2005; 213).

"The *fundamental problem* that price statisticians face when attempting to construct a real estate price index is that *exact matching of properties over time is not possible* for two reasons: (i) the property depreciates over time (*the depreciation problem*) and (ii) the property may have had major repairs, additions or remodeling done to it between the two time periods under consideration (*the renovations problem*). Because of the above two problems, constructing constant quality real estate price indexes cannot be a straightforward matter; some form of imputation or indirect estimation will be required." Erwin Diewert (2009b; 92).

These quotations indicate that the construction of a RPPI will be much more difficult than the construction of a "normal" price index based on a matched model methodology. It should be recognized at the outset that, because of the difficulties resulting from the uniqueness of each dwelling unit, it will not be possible to construct a "perfect" RPPI; it will only be possible to construct an approximation to the theoretically ideal index for each purpose.

The question of what is the purpose of a Residential Property Price Index has been addressed in Chapter 3 where the many uses of a RPPI have been considered. In this chapter, the focus on the uses of a RPPI will be to fill in gaps in the System of National Accounts (section 4.2) and in the construction of a Consumer Price Index (section 4.3). It is likely that if appropriate RPPIs can be constructed to fill in these gaps, then the resulting family of RPPIs will meet the needs of most users.

Broadly speaking, two separate RPPIs can be distinguished: a constant quality price index for the *stock* of residential housing at a particular moment in time and a constant quality price index for residential property *sales* that took place during a particular period of time. The construction of these two types of index will be different; e.g., the weighting associated with the two types will differ.

In section 4.4, four main approaches to constructing a RPPI will be briefly discussed. Details on these methods will be presented in Chapters 5 to 8.

A variety of miscellaneous topics will be addressed in the final four sections of this chapter. These topics include the frequency of the RPPI and user needs (section 4.5), the consistency of monthly with quarterly estimates and the consistency of quarterly with annual estimates (section 4.6), revision policies (section 4.7), and seasonal adjustment (section 4.8).

4.2 Residential Property Price Indexes and the System of National Accounts

The *System of National Accounts (SNA) 1993* and its recent updating, the *System of National Accounts 2008*,[1] provide a comprehensive accounting framework for an economy. The SNA partitions the value flows in the economy into various meaningful categories and provides a reconciliation of the flow accounts with the corresponding stock accounts. It is furthermore recommended to decompose the values in the cells of these accounts into price and volume (or quantity) components.

There are three places in the SNA where residential property price indexes are required to convert nominal values into volumes or real values:

- the *stock* of residential properties that exist at a particular location in the country at a particular point in time;
- the *sales* of residential properties that were sold in a particular location in the country over a particular time period, and
- the structures part of the *sales* of *new* residential properties that were sold in a particular location in the country over a particular time period.

[1] See Eurostat, IMF, OECD, UN and the World Bank (1993) and (2009).

The stock of residential properties in a country is a component of national wealth. Hence, a price index is required for residential properties so that balance sheet estimates of *real wealth by component* can be formed.[2] Balance sheet estimates of national wealth typically distinguish between the structures component of residential property and the land component, so if it is desired to form estimates of the country's real stock of residential structures and the real stock of residential land, it will be necessary to decompose residential property values into *land and structures components* and to construct price indexes for each of these components.

It may not be immediately evident why a price index for the sales of residential properties is required for national income accounting purposes. It is required to estimate the *real output* of the residential real estate industry; i.e., the industry that provides services to help facilitate the sales of residential properties. A bit of algebra will help to explain why having a price index for the sales of residential properties is essential for this purpose.

Suppose that the value of real estate agent commissions is V_C^t for some class of property transactions in period *t* and suppose that the corresponding value of sales for the same group of properties (including the commissions) is V_S^t. Further, suppose that a constant quality price index for this type of sale has been constructed and the period *t* value of this price index is P_S^t.[3] An estimate of volume of sales for this class of real estate transactions in period *t*, say Q_S^t, can be formed by simple division:

$$Q_S^t \equiv V_S^t / P_S^t . \tag{4.1}$$

The real estate industry can be treated as a retailing or wholesaling industry; i.e., it is a margin industry that can be thought of as buying a property at the pre-commission price and selling it at the post-commission price. The value of the service is equal to the commission revenue, V_C^t, and the quantity of the service is proportional to the volume of the sales, Q_S^t. Thus set the volume of the real estate services, Q_C^t, equal to Q_S^t:

[2] A price index for the stock of residential properties is also of some use to central bankers who are typically interested in whether property price bubbles are forming in their respective countries; see Chapter 3 above.

[3] Instead of using a purchaser's price index, it is also possible to use a seller's price index. When constructing a constant quality price index for housing, should the price determining characteristics of the seller or those of the purchaser be used to do quality adjustment? It could be argued that the quality determining characteristics of the purchaser should be used in order to quality adjust prices for residential properties, since if the purchaser does not see enough value in the price of a property, it will not be purchased. This suggests that a purchaser's constant quality price index should be constructed as opposed to a seller's constant quality price index. However, one could also argue that if the selling price of a property (regarded as a function of the characteristics of the property) is not high enough, then producers of new housing units will not build a new unit and thus it is the price determining characteristics of the seller that should count, at least in the context of valuing new housing units. Rosen (1974) discusses these issues. In terms of Rosen's analysis of the determinants of the hedonic surface, for the case of new housing units, it is likely that his Case 1 analysis is relevant, where cost conditions are identical across firms and thus the hedonic surface is determined by the supply side of the market; see Rosen (1974; 50-51).

$$Q_C^t \equiv Q_S^t . \tag{4.2}$$

Finally, the price index in period *t* for the subsector of the real estate industry associated with the property sales, with value V_S^t in period *t*, is set equal to the value of the corresponding commissions, V_C^t, divided by the corresponding volume, Q_C^t:

$$\begin{aligned} P_C^t &\equiv V_C^t / Q_C^t && (4.3) \\ &= V_C^t / [V_S^t / P_S^t] && \text{using (4.1) and (4.2)} \\ &= [V_C^t / V_S^t] P_S^t \\ &= m_C^t P_S^t , \end{aligned}$$

where $m_C^t = V_C^t / V_S^t$ is the period *t margin rate* for this class of real estate transactions; i.e., m_C^t is the ratio of commissions in period *t* to the corresponding purchaser's total value of the real estate transactions. Thus the period *t* price index for the output of this segment of the real estate industry is the product of the margin rate m_C^t times the constant quality price index for the properties sold in period *t*, P_S^t. This explains why constant quality price indexes for sales of residential properties are useful for national income accounting purposes.

The third value cell in the System of National accounts that requires a housing price deflator is the *value of new housing produced* in various locations in the country over a reference time period. This value flow is part of the *gross investment* in the country. When a new property is produced in the reference period and if there were no improvements made to the underlying land that the new structure occupies, then the portion of the sale price that can be attributed to the site land should be deducted from the sale price and the residual amount is part of gross investment and also part of the construction industry's output. Thus, a RPPI for the *structures component of the sales of new residential properties* is required in the national accounts. It will be necessary to decompose sales of new residential properties into *land and structures components* and to construct a constant quality price index for the structures component in order to serve the needs of the System of National Accounts.

Recall the above discussion about modeling the output of the real estate industry. The fact that a sale of a new property will have various transactions costs (such as real estate commissions) associated with the sale leads to some complexities in the System of National Accounts that have not yet been definitively resolved. From the viewpoint of the construction industry, these transactions costs are not part of the revenues that accrue to the construction sector, so these costs should not be included in the value of the output of the construction industry. However, from the viewpoint of the sector that purchases the new housing unit, these transactions costs are a real cost and there must be an accounting for these costs. There are a number of ways that *transactions costs* associated with the purchase of a new housing unit could be treated (from the viewpoint of the purchaser):

- simply attribute all of the costs to the period of purchase and treat the transactions costs as an expenditure by the purchaser[4] (which is an acquisitions approach to these costs);
- include transactions costs as part of the structures component of the value of the purchase so that these costs would be amortized over time using the same depreciation rate that was being used to depreciate the structure or
- separately amortize the transactions costs according to average length of time a residential property of the type under consideration is being held before it is resold.

Conceptually, the last treatment seems preferable[5] but the first and second treatments will lead to a simpler set of accounts. These issues need to be studied further by national accountants with the input of the broader economics community.

4.3 Residential Property Price Indexes and the Consumer Price Index

Pricing the services of Owner Occupied Housing (OOH) in a Consumer Price Index (CPI) is extensively dealt with in the *Consumer Price Index Manual*.[6] There is no universal consensus on the treatment of OOH in a CPI but the manual suggests four possible approaches.[7]

- The *money outlays or payments approach*. In this approach, the out of pocket expenses of home ownership are simply added up. These costs include expenditures on maintenance and repair, mortgage interest costs, insurance premiums, property taxes and condominium charges (if the housing unit is a condominium). Two important types of implicit cost and one important implicit benefit of home ownership are not included. The two omitted costs are depreciation and the opportunity cost of the funds that are tied up in the homeowner's equity in the house; the implicit omitted benefit is any (net) capital gains that may accrue to the owner during the time period under consideration.[8] The money outlays approach is

---

[4] The price index that could be used to convert the nominal value of transaction charges into a real amount (or volume) is a composite purchase price index for the type of property under consideration which includes both the land and structures components.

[5] This is the treatment used by the Australian Bureau of Statistics. An unresolved issue is the choice of price deflator in order to form real amortization charges. That is, should a structures price index be used or should a composite structures and land price be used? In the case of real estate the commissions are generally proportional to the overall price of the property (the sum of the land and structures components) so it would be appropriate to use a composite property price index for the deflation of this component of transactions costs. Government transactions taxes or stamp duties may impose different rates on the land and structures components of the sale and so working out an appropriate real price for this component of transactions costs may be rather complicated. Again, it may be acceptable to avoid all of these complexities and just use a composite purchase price index to do the deflation.

[6] See ILO, IMF, OECD, Eurostat, UN and World Bank (2004), Chapter 23.

[7] Diewert (2002) (2009a) (2009b) provides more discussion of alternative methods.

[8] The money outlays concept is explained in some detail in Baldwin, Nakamura and Prud'homme (2010).

useful if an analyst wishes to focus on the disposable income of households. However, it is not particularly useful as a measure of household consumption services (because of the omission of the costs and benefits mentioned above).

- The *(net) acquisitions approach*. In this approach, the services of OOH are ignored in the CPI except when a new housing unit is introduced into the market place. The purchase price of the new dwelling unit is charged to the period of purchase so that a purchase of a new house is treated in the same manner as the purchase of a nondurable good or service. A variant of this approach is to decompose the selling price of the newly built residential property into land and structures components and to use just the structures component as the price which will enter into the CPI.
- The *rental equivalence approach*. In this approach, a price is imputed for the shelter cost of the owner occupied housing unit (both for new and existing units), which is equal to the price at which the unit could be rented.[9]
- The *user cost approach.* In this approach, the financial opportunity cost of owning the house and using its services during the reference period is calculated.

Since the *Consumer Price Index Manual* was written, a fifth concept for pricing the services of OOH has been suggested:[10]

- The *opportunity cost approach*. In this approach, the price for the services of an owned dwelling unit is set equal to the *maximum* of its rental equivalence and user cost prices.

In all of the above approaches except the money outlays and rental equivalence approaches, there is a need for constant quality price indexes for either newly-built dwelling units or for the existing stock of dwelling units. The user cost and opportunity cost approaches to pricing the services of a residential housing unit are not completely straightforward. The Appendix to this chapter outlines in more detail how these approaches work.

To summarize this section and section 4.2, RPPIs are needed in the construction of a CPI and to deflate several value flows and stock holdings in the System of National Accounts. For CPI and national accounts purposes, it will be useful or necessary to have a decomposition of the price indexes into structures and land components. More specifically, it would be useful to be able to construct the following set of RPPIs:[11]

[9] This is also part of household final consumption in the National Accounts.

[10] See Diewert (2009b), Diewert and Nakamura (2009) and Diewert, Nakamura and Nakamura (2009).

[11] Fenwick (2005) (2006) argued that it would be useful to develop a coherent conceptual framework for a *family* of real estate price indexes. "It can be seen that user needs will vary and that in some instances, more than one measure of house price or real estate inflation may be required. It can also be seen that coherence between different measures and with other economic statistics is important and that achieving this will be especially difficult as statisticians are unlikely to have an ideal set of price indicators available to them." David Fenwick (2006; 8).

- a price index for the total *stock* of residential housing at a particular moment in time;
- a price index for the *owner occupied stock* of residential housing (which is a subindex of the index in the bullet point above);
- a price index for residential property *sales* (both newly-built and existing dwelling units) that took place during a given period of time;
- a price index for the *sales of newly-built residential properties* during a given period of time;
- a price index for the *structures component of newly-built residential properties* that were sold during a given period of time.

4.4 Main Methods

To measure pure price change, RPPIs must be adjusted for quality changes. In other words, to construct a *constant quality* RPPI, it will be necessary to somehow control for variations in the amounts of the price determining characteristics of the properties. The most important characteristics are:

- the *area of the structure* (in squared feet or in meters squared);
- the *area of the land* that the structure sits on (in squared feet or meters squared);
- the *location* of the property;
- the *age* of the structure;
- the *type of structure*; the structure can sit entirely in the lot without sharing any walls with an adjacent structure (detached dwelling unit) or share one wall with a neighbouring unit (semi-detached dwelling unit), or the dwelling unit can be a single apartment or unit in a multifamily residence (apartment or condominium building);
- the *materials used* in the construction of the house (primarily wood, brick, concrete or traditional materials; i.e., a shack or shanty), and
- *other price determining characteristics* such as the number of bedrooms, the number of bathrooms, an external garage, a swimming pool, air conditioning, distance to amenities, etc.

Four main methods have been suggested in the literature to control for changes in the amounts of the property characteristics: stratification or mix adjustment, repeat sales methods, hedonic regression methods, and the use of property assessment information. Below, a brief overview of the four methods will be given. More details can be found in Chapters 5-8.

Stratification of the property sales according to some of the price determining characteristics is a straightforward and computational simple way to adjust for changes in the quality mix of the samples in different time periods. By determining a number of reasonably homogeneous strata or cells, the average selling price within each cell can be used as a (proxy to a) constant quality price for that type of property. Regular index number theory can then be applied to aggregate up the average prices by cell into an overall index. Stratification methods are also referred to as *mix adjustment* methods:

"House price observations are grouped into sets or 'cells' of observations on houses with similar location and physical attributes. [.....] The mean prices in each cell are weighted together to give a 'mix adjusted' price. A change in the composition of the sample will alter the number of observations in each cell. But if the cells are defined sufficiently precisely, so that all elements of the cell have similar prices and price trends, then such compositional changes will not systematically affect the mix adjusted house price. Robert Wood (2005; 214).

The *repeat sales method* addresses the quality mix problem by looking at properties that were sold twice or more in the data set at hand. Restricting the comparison to such repeat sales ensures that all price relatives compares like with like, provided that the quality of the houses remains unchanged. The standard repeat sales method is based on a regression model where the repeat sales data pertaining to all periods are pooled. A potential drawback is "revision": when new data is added to the sample and the model is re-estimated, previously computed index numbers will change. An advantage of repeat sales methods is that, as properties are matched at the address level, micro location is held constant in the comparisons.

One problem with the repeat sales method is that it does not account for quality changes of the individual houses; a dwelling unit can be renovated over time and it experiences depreciation over time so its quality changes over time. *Hedonic regression methods* can in principle adjust for such quality changes in addition to changes in the quality mix of the samples. They utilize information on the relevant property characteristics to estimate quality adjusted price indexes using regression techniques, though it may prove difficult to sufficiently control for location. There are different ways to estimate hedonic price indexes. The *time dummy variable method* has been prominent in the real estate literature. This method models the price of a property as a function of its characteristics and a set of dummy variables which indicate the time periods. Because the data of all periods are pooled, this method suffers from revision, similar to the repeat sales method. Another drawback of the time dummy method is that it places perhaps unwarranted restrictions on variations in the price of land and structures across time. These difficulties with the time dummy variant of the hedonic regression approach can be avoided by using another variant of the method known as the *hedonic imputation method*.

Many countries tax real estate property and are likely to have some sort of official valuation office that provides periodic appraisals of all taxable real estate property. *Assessment-based methods* combine selling prices with appraisals to compute price relatives (sale price appraisal ratios) and control for quality mix changes. The Sale Price Appraisal Ratio (SPAR) method is based on the matched model methodology. In contrast to the repeat sales method, it relies on all (single and repeat) sales data and there is no revision of previously computed figures. Of course the method can only be used in countries where accurate assessed values of the properties are available.

If the reference period is a year, then all methods will tend to generate similar estimates of the trend in residential property price changes for an entire country. However, as will be seen in the examples presented in Chapters 5-8 and Chapter 12, different methods do generate small but significant differences in trends while for shorter periods they can lead to rather different estimates of price change. The various methods could also measure turning points, e.g., from booms to busts, differently.

As hedonic methods assume that information on the characteristics of the properties sold is known, the samples can be stratified and, if a sufficient number of observations is available, separate indexes can be estimated for the strata. In other words, hedonic regression methods can provide a set of constant quality price indexes for various types of property. Obviously, if data on some price determining characteristics is available, then repeat sales and assessment-based methods can also be combined with stratification.

Stratification can also be used to approximate a stock based RPPI. In this case the stratum weights will be based on census data pertaining to the value of the owner occupied housing stock. The stratum price indexes will still be based on sample data of properties sold. Within each stratum, the properties traded are now treated as a (preferably random) sample from the stock. Since census data is typically collected with long time intervals, up to ten years, stock value weights can usually only be updated very infrequently.

As was discussed in Sections 4.2 and 4.3, for various purposes it is necessary to decompose the overall price of a property into (additive) components that reflect the price of the structure and the price of the land the structure sits on. In Chapter 9 it is shown how hedonic regression techniques can be used to accomplish this decomposition.

4.5 The Frequency of the RPPI and User Needs

For inflation monitoring purposes, central banks would like to have an RPPI on a monthly or quarterly basis. For National Accounts purposes quarterly indexes will suffice, while for CPI purposes monthly indexes are generally required. Given that the number of observations for a monthly price index will only be approximately one third of the number for a quarterly index, statistical agencies will have to carefully evaluate the tradeoff between publication frequency, timeliness and accuracy. The use of monthly data may lead to rather noisy figures, whatever method used to compile an RPPI. To mitigate the noise, a moving average could be computed but this creates new problems, as will be explained below.[12]

It is useful to outline some of the tradeoffs that statistical agencies may face when attempting to construct house price indexes that meet the needs of users. Before examining the tradeoffs, it will be necessary to review the user needs for a *family of residential property price indexes*. The following list of user needs is borrowed from the list compiled by Emily Carless (2011) from the National Statistician's Office of the UK Statistics Authority. The family of RPPIs should:[13]

- be based on the price paid for transacted properties;
- be stratified by region;
- be stratified by type of housing (e.g., detached, row, high rise, type of construction, etc.);
- be computed on a monthly basis;
- aggregate up to a consistent national index;
- be accurate and timely with minimal revisions.

The fifth requirement, that the various subindexes aggregate up to a consistent national index is not too difficult to meet. Whether the first requirement, that the price indexes be based on transaction prices, can be met, depends on availability of the data. In many countries, actual selling prices are used to compile RPPIs, but not all statistical agencies may have access to transaction data. Even if transaction data are available, there can be a time lag involved (as will be discussed in Chapter 10), so that in practice the first requirement could be at variance with the sixth requirement saying that the indexes should be timely.

There is also a tension in the other requirements: having many strata and asking for monthly indexes means that not many sales will take place in each stratum, leading to rather volatile (high variance) subindexes, which are necessarily inaccurate. Although, as mentioned above, volatility can be reduced by taking moving averages of the monthly indexes,[14] these moving averages will not give timely signals of price change. That is, the resulting average index will be centered in the middle of the time period for the moving average and hence will not be

[12] Nevertheless, moving averages are, for example, used in Iceland. It may also be necessary to use slightly out of date information in a monthly CPI context; see Gudnason and Jónsdóttir (2006; 4).

[13] In addition to the requirements listed, Carless noted that users desire a clear explanation of the methods used to construct the statistics and indicators of the quality of the measures. Also, some users want seasonally adjusted series in addition to the unadjusted series.

[14] The volatility may also be mitigated by combining some of the strata but then users will lose some of the desired geographical or type of housing coverage. In addition, the strata that are combined may not be subject to the same general price trends and thus there is the possibility of some unit value bias due to the aggregation of the strata.

accurate until some months have passed.[15] In particular, this could give a misleading picture of upswings and downturns in the housing market. So in general, *it will be impossible to meet all of the user needs listed above*. Statistical agencies will have to make compromises of some sort in their attempts to meet the different user needs.

4.6 Consistency of Monthly with Quarterly Estimates

How can monthly estimates of real estate price changes be made consistent with quarterly estimates? The answer to this question is reasonably straightforward if the same average price or unit value methodology is applied to the quarterly data as is applied to the monthly data. Suppose that a monthly sales RPPI is constructed using the stratification (or mix adjustment) method. As will be explained in Chapter 5 more thoroughly, the monthly price for a particular cell is the average transaction price or unit value and the corresponding quantity is the total number of properties traded. The quarterly RPPI for that cell would start out by calculating a quarterly unit value, and the corresponding quantity is the quarterly total number of stratum transactions. A bit of algebra will make the relationship between the quarterly cell price and quantity data to the corresponding monthly data clear.[16]

Suppose that there are T quarters of monthly data. Denote the value of quarterly transactions in a particular cell in the stratification scheme by V^t for $t = 1,...,T$. Within each quarter t, the value of first month transactions is denoted by V_1^t, of second month transactions by V_2^t and of third month transactions by V_3^t. The quarter *t monthly unit value prices* are denoted by P_1^t, P_2^t and P_3^t and the corresponding *monthly number of transactions* are denoted by Q_1^t, Q_2^t and Q_3^t. Note that V_m^t equals $P_m^t Q_m^t$ for $m = 1,2,3$ and $t = 1,...,T$. The value of transactions for quarter t, V^t, is equal to the sum of the monthly transactions within the quarter:

$$V^t = V_1^t + V_2^t + V_3^t = P_1^t Q_1^t + P_2^t Q_2^t + P_3^t Q_3^t \qquad t = 1,...,T. \tag{4.4}$$

The quarterly quantity series, Q^t, is the sum of the monthly transactions within the quarter and the quarterly price series, P^t, is the quarterly unit value for the cell under consideration; i.e.:

$$Q^t = Q_1^t + Q_2^t + Q_3^t; \qquad t = 1,...,T \tag{4.5}$$

$$\begin{aligned} P^t &= V^t / Q^t \qquad t = 1,...,T \qquad (4.6) \\ &= [P_1^t Q_1^t + P_2^t Q_2^t + P_3^t Q_3^t]/[Q_1^t + Q_2^t + Q_3^t] \\ &= s_1^t P_1^t + s_2^t Q_2^t + s_3^t Q_3^t, \end{aligned}$$

[15] This number is equal to half the window length of the moving average.

[16] The same type of analysis can be applied to the relationship between an annual (mix adjustment) sales RPPI and the corresponding quarterly estimates.

where the *month m share of transactions in quarter t*, s_m^t, is defined as

$$s_m^t = Q_m^t / Q^t \qquad m = 1,2,3\,;\; t = 1,...,T\,. \qquad (4.7)$$

Thus, the quarterly price level for the cell under consideration, P^t, is equal to a transaction share weighted average of the monthly price levels P_m^t for the months m in quarter t.

For RPPI construction methods other than straightforward stratification (hedonic regression, repeat sales, use of appraisal data), the relationship between the quarterly estimates of price change and the corresponding monthly estimates will be more complex. However, in the end, these methods will generate a price index, say P^t for period t, that is associated with a certain group of transactions (or stocks). Generally, the corresponding period t value associated with these stocks, say V^t, will be available and thus a corresponding period t volume, $Q^t = V^t/P^t$, can be defined, so the above algebra can be applied.

4.7 Revision Policies

It would seem that a RPPI for the sales of properties could be constructed without a need for revisions but it turns out that it is not always easy to assemble timely data on property sales. In section 4.6 it was seen that the construction of a stock type RPPI is dependent on census information on housing, which is often subject to long delays. Moreover, when a new housing census becomes available, it is generally desirable to use this information to retrospectively adjust the stock type RPPI back to the previous housing census date. Thus it will generally be desirable to allow stock RPPIs to be revised. This should not create any major problems for National Accounts purposes, since revisions are routine in this context.

The area where revisions do cause problems is in the context of producing a CPI. In general, CPIs are not revised. The treatment of owner occupied housing in a CPI requires a stock type RPPI if either the acquisitions, user cost or opportunity cost approach is used. It may then be necessary to use preliminary information to compile the RPPI. When additional data becomes available, a revised CPI could be published as an *analytical series* so that analysts could form some rough estimates of the possible bias in using the unadjusted CPI based on a preliminary estimate of the RPPI for owner occupied housing.

4.8 Seasonal Adjustment

Although the situation may differ somewhat across countries, in general there are substantial seasonal fluctuations in the *quantities* of properties traded over the year. For the construction of a RPPI, the question is whether seasonality in quantities leads to seasonality in *prices*. The empirical evidence is somewhat mixed. Meese and Wallace (1991) find limited seasonality in prices in their econometric study. Prasad and Richards (2008) report that median prices in

Australian cities are seasonal, but this seasonality vanishes after controlling for compositional change through stratification. At aggregate levels, and particularly at the nation-wide level, it seems therefore unlikely that RPPI series exhibit strong seasonal fluctuations. However, at lower levels of aggregation it would be useful to check whether any seasonality in prices is present and adjust for this if seasonally adjusted series are required. As noted in section 4.5, users do want seasonally adjusted series made available to them (in addition to the unadjusted series) if there is evidence of seasonality in prices.

In section 5.6 of Chapter 5, a numerical example is worked out which shows how seasonality can be treated using simple index number techniques. Standard seasonal adjustment methods could also be used.

Appendix: The Role of House Price Indexes in the Construction of User Costs

This Appendix shows how user costs and opportunity costs can be constructed. Section 4.A.1 discusses how user costs are constructed for durable goods in general. Section 4.A.2 brings in the additional difficulties due to the fact that properties are unique goods and are a mixture of land and structures components. Section 4.A.3 discusses the opportunity cost approach to pricing the services of Owner Occupied Housing (OOH).

4.A.1 The Construction of User Costs for Durable Goods in General

In this section, the elements of user cost theory for a durable consumer good are laid out. The essence of durability is that it provides some sort of service to the purchaser over many time periods. For many purposes (including the valuation of household consumption expenditures on owner occupied housing services) it is not appropriate to charge the entire purchase cost of a durable good to the initial period of purchase; the purchase cost should be spread over its useful life. The question is: how exactly should this intertemporal cost allocation be done?

As mentioned in section 4.3 of the main text, there are two main approaches to pricing the *services* of an owner occupied dwelling unit:[17] the *rental equivalence approach* and the *user cost approach*. The user cost approach is important in its own right (when only few dwelling units in a country are rented, it is not possible to value the services of owner occupied housing using rental equivalence) but it also is important as a way to explain how landlords might set their rents for rental dwelling units. However, pricing the services of houses is more difficult than pricing the services of, say, a standard model automobile because housing services are much more complex.[18] Therefore, in this section the problems of pricing the services of an *ordinary consumer durable good* (that is available in the marketplace in the same form over many periods) will be studied before tackling the complexities of housing in the following section.

The user cost approach to the treatment of durable goods is in some ways very simple: it calculates the cost of purchasing the durable at the beginning of the period, using the services of the durable during the period and then netting off from these costs the benefit that could be obtained by selling the durable at the end of the period, taking into account the interest forgone in tying up one's capital in purchasing the durable. However, there are several details that are somewhat controversial. These details involve the treatment of depreciation, interest and capital gains or holding gains.

[17] The acquisitions approach implicitly allocates all of the services of a newly purchased housing unit to the period of purchase but the System of National Accounts does not recognize this approach as a valid approach to pricing the services of OOH. For other durable goods, the SNA does recognize the acquisitions approach as a valid approach for pricing the services of a durable good.

[18] In particular, housing services provide the joint services of the structure and the land that the structure sits on and houses are generally unique goods.

Another complicating factor of the user cost approach is that it involves making distinctions between current period purchases within the period under consideration and the holdings of physical stocks of the durable at the beginning and the end of the accounting period. Normally in the System of National Accounts, all prices and quantity purchases are thought of as taking place at a single point in time, say in the middle of the period under study, and consumption is thought of as taking place within the period as well. Thus, in this case where the commodity is entirely consumed within the purchasing period, there is no need to consider the valuation of stocks of consumer durables that households may have at their disposal. The complexity involved in accounting for stocks and flows are unfamiliar to many price statisticians and so it may be useful to describe these problems in some detail.

To determine the net cost of using a particular durable good during say period 0, assume that one unit of the durable good is purchased at the beginning of period 0 at the price P^0. The "used" or "second-hand" durable good can be sold at the end of period 0 at the price $P_S{}^1$. It might seem that a reasonable net cost for the use of one unit of the consumer durable during period 0 would be its initial purchase price P^0 less its end of period 0 "scrap value" or market opportunity selling price, $P_S{}^1$. However, money received at the end of the period is not as valuable as money received at the beginning of the period. To convert the end of period value into its beginning of the period equivalent value, it is necessary to *discount* the term $P_S{}^1$ by the term $1+r^0$ where r^0 is the beginning of period 0 nominal interest rate that the household (or purchaser) faces. Hence, the *period 0 user cost* u^0 for the consumer durable[19] is defined as

$$u^0 \equiv P^0 - P_S{}^1/(1+r^0). \qquad (4.A1)$$

There is another way to interpret the user cost formula (4.A1): the consumer purchases the durable at the beginning of period 0 at the price P^0 and charges himself or herself the rental price u^0. The remainder of the purchase price, I^0, defined as

$$I^0 \equiv P^0 - u^0 \qquad (4.A2)$$

can be regarded as an *investment*, which is to yield the appropriate opportunity cost of capital r^0 the consumer faces. At the end of period 0, this rate of return could be realized provided that I^0, r^0 and the selling price of the durable at the end of the period $P_S{}^1$ satisfy

$$I^0(1+r^0) = P_S{}^1. \qquad (4.A3)$$

Given $P_S{}^1$ and r^0, (4.A3) determines I^0, which in turn, given P^0, determines the user cost u^0 via (4.A2).[20]

[19] This approach to the derivation of a user cost formula was used by Diewert (1974) who in turn based it on an approach due to Hicks (1946; 326). Note that later, this user cost will be interpreted as a beginning of the period user cost since all costs are discounted to the beginning of the period.

[20] This derivation for the user cost of a consumer durable was also made by Diewert (1974; 504).

The above made clear that the user cost approach to pricing the services of a durable good for a period has an investment aspect to it. Note that the user cost approach is also a *financial opportunity cost approach*; i.e., the opportunity cost of the financial capital that is tied up in the purchase (or continued holding) of the durable good is taken into account. Finally, note that user costs are not like the prices of nondurables or services because the user cost concept involves pricing the durable at *two* points in time rather than at a single point in time. Because the user cost concept involves prices at two points in time, money received or paid out at the first point in time is more valuable than money paid out or received at the second point in time and so *interest rates* creep into the user cost formula.

Also, because the user cost concept involves prices at two points in time, *expected prices* can be involved if the user cost is calculated at the beginning of the period under consideration instead of at the end. So the price statistician has two options for the choice of P_S^1:

- Use the *expected price* of the durable at the end of the period from the perspective of the beginning of the period, or
- Use the *actual market price* of a similar second hand durable at the end of the period (if such a market price exists).

The use of an expected price leads to an *ex ante user cost* whereas the use of an actual market price for the used durable at the end of the period leads to an *ex post user cost*. Which concept should be used in practice? In the present context it is reasonable to argue for the use of the ex ante concept for two reasons:

- The ex ante user cost concept is likely to be closer to a rental price of the durable good (if it exists),[21] which many price statisticians would view as a preferred price for the services of the durable during the period, and
- The ex ante user cost is closer to the purchaser's *expected cost* for using the durable good during the period; the purchaser cannot know exactly what the end of period price will be and hence must form expectations about the end of period price of the durable, which leads to the ex ante user cost as the expected cost for using the services of the durable during the period. Thus, the ex ante user cost is likely to be the charge for the services of the durable that motivates consumer behavior.

The issue of how exactly one forms expectations for the selling price of a used durable will be examined later when the pricing of housing services is discussed.

[21] If a company is in the business of leasing the services of an automobile for a certain period, it has to form *expectations* about the price of its used autos at the end of the leasing period in order to calculate its schedule of rental or leasing prices for its stock of automobiles.

With all of the above complications, it is no wonder that many price statisticians would like to avoid using user costs as a pricing concept. However, as will be seen, the use of user costs may be unavoidable in the context of pricing the services of owned dwellings under certain circumstances.

The user cost formula (4.A1) can be put into a more familiar form using the end of period 0 *economic depreciation rate* δ^0 and the period 0 *asset inflation rate* i^0. Define the end of period 1 *depreciation rate* δ^0 by

$$(1 - \delta^0) \equiv P_S{}^1/P^1, \qquad (4.A4)$$

where $P_S{}^1$ is the price of a used asset at the end of period 0 and P^1 is the price of a new asset at the end of period 0.[22] The *period 0 inflation rate* for the new asset, i^0, is defined by

$$1+i^0 \equiv P^1/P^0. \qquad (4.A5)$$

Eliminating P^1 from equations (4.A4) and (4.A5) leads to the following formula for the *end of period 0 used asset price*:

$$P_S{}^1 = (1 - \delta^0)(1 + i^0)P^0. \qquad (4.A6)$$

Substitution of (4.A6) into (4.A1) yields the following expression for the *period 0 user cost* u^0:

$$u^0 = [(1 + r^0) - (1 - \delta^0)(1 + i^0)]P^0/(1 + r^0). \qquad (4.A7)$$

Note that $r^0 - i^0$ can be interpreted as a period 0 *real interest rate* and that $\delta(1+i^0)$ can be interpreted as an *inflation adjusted depreciation rate*.

In (4.A7), the user cost u^0 is expressed in terms of prices that are discounted to the *beginning* of period 0. However, it is also possible to express the user cost in terms of prices that are "antidiscounted" or "appreciated" to the *end* of period 0.[23] The *end of period 0 user cost* p^0 is defined as

$$p^0 \equiv (1 + r^0)u^0 = [(1 + r^0) - (1 - \delta^0)(1 + i^0)]P^0 = [r^0 - i^0 + \delta^0(1 + i^0)]P^0. \qquad (4.A8)$$

[22] If the durable that was purchased (or held) by the household at the beginning of the period was a used durable, then interpret P^1 as the second hand market price of a used durable that is in the same condition as the initially held durable.

[23] Thus the beginning of the period user cost u^0 discounts all monetary costs and benefits into their dollar equivalent at the beginning of period 0 whereas p^0 accumulates or appreciates all monetary costs and benefits into their dollar equivalent at the end of period 0. This leaves open how flow transactions that take place within the period should be treated. Following the conventions used in financial accounting suggests that flow transactions taking place within the accounting period be regarded as taking place at the end of the accounting period and hence following this convention, end of period user costs should probably be used by the price statistician. For additional material on beginning and end of period user costs, see Diewert (2005; 485).

where the second equation follows using (4.A7). If the real interest rate r^{0*} is defined as the nominal interest rate r^0 less the asset inflation rate i^0 and if the generally small term $\delta^0 i^0$ is neglected, then the end of the period user cost defined by (4.A8) reduces to[24]

$$p^0 = (r^{0*} + \delta^0)P^0. \tag{4.A9}$$

Abstracting from transactions costs, it can be seen that the end of the period user cost defined by (4.A9) is an *approximate rental cost*; the rental cost for the use of a durable good should equal the (real) opportunity cost of the capital tied up, $r^{0*}P^0$, plus the decline in value of a new asset over the period, $\delta^0 P^0$. Formulae (4.A8) and (4.A9) thus cast some light on the economic determinants of rental or leasing prices for consumer durables.

If the simplified user cost formula defined by (4.A9) above is used, then forming a price index for the user cost of a durable good is not very much more difficult than forming a price index for the purchase price of the durable good, P^0. The price statistician needs only to

- Make a reasonable assumption as to what an appropriate monthly or quarterly real interest rate r^{0*} should be;[25]
- Make an assumption as to what a reasonable monthly, quarterly or annual depreciation rate δ^0 should be;[26]
- Collect purchase prices P^0 for the durable and form the user cost.

There are some additional difficulties associated with the user cost approach to measuring the services of a consumer durable. The above discussion deals only with the formation of a user cost for a newly purchased consumer durable. It is necessary to extend the analysis to price the services of used units of the consumer durable as well. In order to price out the services of a used durable good, it is necessary to make assumptions about the form of depreciation of the

[24] If one takes the ratio of the approximate rental price for the durable good, p^0, to its asset value, P^0, the *rent to value ratio* $p^0/P^0 = r^{0*} + \delta^0$ is obtained, which is equal to the sum of the appropriate *real* interest rate r^{0*} plus the appropriate depreciation rate δ^0. Since real rates of interest and depreciation rates are approximately constant over time, the rent to value ratio will also be approximately constant over time and hence a historical rent to value ratio times a current asset price index will generally give an adequate approximation to an imputed rental rate for the consumer durable. In the housing literature, a rent to value ratio is often called a *capitalization rate*; e.g., see Garner and Short (2009; 237) or Crone, Nakamura and Voith (2009; 70).

[25] This is not completely straightforward. It is difficult to determine exactly what the appropriate household nominal opportunity cost of capital should be and even if we come to agreement on this point, there will be difficulties in estimating expected inflation rates. In the end, it may boil down to picking a somewhat arbitrary real interest rate in the 2% to 5% range (for annual rates), depending on the recent experience of the country under consideration.

[26] The geometric model for depreciation requires only a single monthly or quarterly depreciation rate. Other models of depreciation may require the estimation of a sequence of vintage depreciation rates. If the estimated annual geometric depreciation rate is δ_a, then the corresponding monthly geometric depreciation rate δ can be obtained by solving the equation $(1 - \delta)^{12} = 1 - \delta_a$. Similarly, if the estimated annual real interest rate is r_a^*, then the corresponding monthly real interest rate r^* can be obtained by solving the equation $(1 + r^*)^{12} = 1 + r_a^*$.

good; does the service flow given to the consumer remain constant throughout the useful life of the durable good or does it decline as the good ages? If the service flow remains constant, then we have *one hoss shay* or *light bulb depreciation* whereas if the service flow declines at a constant linear or geometric rate, then we have *straight line* or *geometric depreciation*.[27]

How can one tell whether one hoss shay or geometric depreciation is applicable for a certain consumer durable? The two patterns of depreciation (and user valuation) can be distinguished if *cross sectional information on rentals* of the consumer durable by the age of the rented asset is available. If there is one hoss shay depreciation, then the rental rates for the consumer durable at a given point in time should be approximately constant for all ages of the durable good whereas if there is geometric depreciation, the rental rates for the good should decline at a geometric rate according to the age of the used durable good. Thus, the various patterns of depreciation can be distinguished when rental markets for used durables exist. In a similar fashion, if *cross sectional information on the prices of used units* of the consumer durable is available, then alternative patterns of depreciation can be distinguished.[28]

4.A.2 The User Cost of Owner Occupied Housing

An *owner occupied dwelling* is different from a normal consumer durable good because it is an example of a *unique* consumer durable good so that it will be difficult to use information on used asset prices in order to determine the pattern of depreciation, which is information that is required to measure a user cost for an owned dwelling unit. As was mentioned in the introduction to this chapter, a particular dwelling unit in a particular country is unique for a number of reasons:

- The *location* of each dwelling unit is unique and location will affect the price of the unit.
- Over time, the dwelling unit *depreciates*; unless there is one hoss shay depreciation, the utility generated by a particular dwelling for the occupying household will tend to decline over time due to the effects of the aging of the structure.
- On the other hand, the effects of depreciation can be offset by *renovation expenditures*, which increase the utility of the dwelling unit.

[27] For descriptions of how to construct user costs by the age of the asset for each of these depreciation models, see Diewert and Lawrence (2000) or Diewert (2005; 506-521).

[28] In the housing context where each house can be regarded as a unique asset, it is necessary to make some additional assumptions in order to identify the form of depreciation. The extra assumptions are of the following type: it is assumed that all housing units in a certain class of structures have a similar pattern of depreciation. Using this type of assumption, empirical evidence suggests that one hoss shay depreciation is unlikely in the housing market since renters are generally willing to pay a rent premium for a new unit over an older unit of the same type. For empirical evidence of this age premium, see Malpezzi, Ozanne and Thibodeau (1987; 378) and Hoffman and Kurz (2002; 19).

As was noted in the main text, for some purposes, it is important to decompose the price of a property into land and structures components. To model the fact that housing is a composite good, consider a particular newly constructed dwelling unit that is purchased at the beginning of period 0. Suppose that the purchase price is V^0. This value can be regarded as the sum of the cost of producing the structure, $P_S{}^0Q_S{}^0$, where $Q_S{}^0$ is the number of square meters of floor space in the structure and $P_S{}^0$ is the beginning of period 0 price of construction per square meter, and the cost of the land, $P_L{}^0Q_L{}^0$, where $Q_L{}^0$ is the number of square meters of the land that the structure sits on and the associated yard and $P_L{}^0$ is the beginning of period 0 price of the land per square meter.[29] Thus at the beginning of period 0, *the value of the dwelling unit* is V^0 defined as follows:

$$V^0 = P_S{}^0Q_S{}^0 + P_L{}^0Q_L{}^0. \qquad (4.A10)$$

Suppose that the anticipated price of a unit of a new structure at the beginning of period 1 is $P_S{}^{1a}$ and that the anticipated price of a unit of land at the beginning of period 1 is $P_L{}^{1a}$. Define the *period 0 anticipated inflation rates for new structures and land*, $i_S{}^0$ and $i_L{}^0$ respectively, as follows:

$$1 + i_S{}^0 \equiv P_S{}^{1a}/P_S{}^0; \qquad (4.A11)$$

$$1 + i_L{}^0 \equiv P_L{}^{1a}/P_L{}^0. \qquad (4.A12)$$

Let δ^0 be the period 0 depreciation rate for the structure. The anticipated beginning of period 1 value for the structure and the associated land is then equal to

$$V^{1a} = P_S{}^{1a}(1 - \delta^0)Q_S{}^0 + P_L{}^{1a}Q_L{}^0. \qquad (4.A13)$$

So the anticipated value of the dwelling unit at the end of period 1, V^{1a}, equals the anticipated price (per unit of new structure of the same quality) at the end of the period, $P_S{}^{1a}$, times one minus the period 0 depreciation rate, $(1 - \delta^0)$, times the quantity of structure purchased at the beginning of period 0, $Q_S{}^0$,[30] plus the anticipated price of land at the end of period 0, $P_L{}^{1a}$, times the quantity of land that the structure associated with the structure, $Q_L{}^0$.

Now calculate the cost (including the imputed opportunity cost of capital r^0)[31] of buying the dwelling unit at the beginning of period 0 and (hypothetically) selling it at the end of period 0.

---

[29] If the dwelling unit is part of a multiple unit structure, then the land associated with it will be the appropriate share of the total land area. This share could be 1 divided by the number of units on the plot or the floor space of the unit divided by the total floor space of the entire structure. Either share allocation could be justified.

[30] Thus the period 0 depreciation rate δ^0 is an *end of period anticipated cross sectional depreciation rate*; i.e., δ^0 is defined by the equation $(1-\delta^0) = V_S{}^{1a}/P_S{}^{1a}Q_S{}^0$, where $V_S{}^{1a}$ is the anticipated market value of the (depreciated) structure at the end of period 0 and $P_S{}^{1a}Q_S{}^0$ is the anticipated end of period 0 value of a newly constructed structure with floor space area $Q_S{}^0$.

[31] More elaborate discussions on how to choose the appropriate opportunity cost of capital when the owner of a dwelling unit has a mortgage on the unit can be found in Diewert and Nakamura (2009), Diewert, Nakamura and Nakamura (2009) and Garner and Verbrugge (2009b; 176).

The following *end of period 0 user cost or imputed rental cost* R^0 for the dwelling unit is obtained using (A11)-(A13):

$$\begin{aligned} R^0 &\equiv V^0(1 + r^0) - V^{1a} \qquad (4.A14)\\ &= [P_S^0Q_S^0 + P_L^0Q_L^0](1 + r^0) - [P_S^{1a}(1 - \delta^0)Q_S^0 + P_L^{1a}Q_L^0]\\ &= [P_S^0Q_S^0 + P_L^0Q_L^0](1 + r^0) - [P_S^0(1 + i_S^0)(1 - \delta^0)Q_S^0 + P_L^0(1 + i_L^0)Q_L^0]\\ &= p_S^0Q_S^0 + p_L^0Q_L^0, \end{aligned}$$

where separate period 0 *user costs of structures and land*, p_S^0 and p_L^0, are defined as follows:

$$p_S^0 = [(1 + r^0) - (1 + i_S^0)(1 - \delta^0)]P_S^0 = [r^0 - i_S^0 + \delta^0(1 + i_S^0)]P_S^0; \qquad (4.A15)$$

$$p_L^0 = [(1 + r^0) - (1 + i_L^0)]P_L^0 \qquad = [r^0 - i_L^0]P_L^0. \qquad (4.A16)$$

Note that the above algebra indicates some of the most important determinants of market rents for rental properties.[32] The user cost formulae defined by (4.A15) and (4.A16) can be further simplified if the approximations that were made in section 4.4 are made here as well (recall equation (4.A9) above); i.e., assume that the terms $r^0 - i_S^0$ and $r^0 - i_L^0$ can be approximated by a real interest rate r^{0*} and neglect the small term δ^0 times i_S^0 in (4.A15). Then the user costs defined by (4.A15) and (4.A16) simplify to

$$p_S^0 = (r^{0*} + \delta^0)]P_S^0; \qquad (4.A17)$$

$$p_L^0 = r^{0*}P_L^0. \qquad (4.A18)$$

The above exposition has neglected two other sources of period 0 cost associated with owning a dwelling unit:

- Various maintenance and insurance costs that are associated with the ownership of a dwelling unit and
- Property taxes that may be payable by the owner to local or state governments.

Assume that period 0 maintenance and insurance costs, M_S^0, are mainly associated with the structure rather than the land under the structure. Suppose that these costs are paid at the end of period 0. These costs can be converted into a *per unit structure charge* μ_S^0 as follows:

$$\mu_S^0 \equiv M_S^0/P_S^0Q_S^0. \qquad (4.A19)$$

Suppose the property taxes that fall on the structure, T_S^0, and the property taxes that fall on the land under the structure, T_L^0, are paid at the end of period 0. Then the period 0 structure and land property tax rates, τ_S^0 and τ_L^0, can be defined as follows:

$$\tau_S^0 \equiv T_S^0/P_S^0Q_S^0\ ;\ \tau_L^0 \equiv T_L^0/P_L^0Q_L^0. \qquad (4.A20)$$

[32] Looking at (4.A13), it can be seen that the land user cost could be negative if the anticipated rate of land price appreciation, i_L^0, is greater than the beginning of the period opportunity cost of capital, r^0. Possible solutions to this complication will be discussed below.

These additional maintenance and property tax costs need to be added to the imputed rental cost for using the dwelling unit R^0. Thus (4.A14) now becomes:

$$\begin{aligned} R^0 &\equiv V^0(1 + r^0) - V^{1a} + M_S{}^0 + T_S{}^0 + T_L{}^0 \\ &= p_S{}^0Q_S{}^0 + p_L{}^0Q_L{}^0 \end{aligned} \tag{4.A21}$$

where the new separate period 0 *user costs of structures and land*, $p_S{}^0$ and $p_L{}^0$, are defined as follows:

$$p_S{}^0 = [r^0 - i_S{}^0 + \delta^0(1 + i_S{}^0) + \mu_S{}^0 + \tau_S{}^0]P_S{}^0; \tag{4.A22}$$

$$p_L{}^0 = [r^0 - i_L{}^0 + \tau_L{}^0]P_L{}^0. \tag{4.A23}$$

The *imputed rent for a dwelling unit* using the user cost approach to the valuation of housing services is thus made up of six main costs:

- The real opportunity cost of the financial capital tied up in the structure, $(r^0 - i_S{}^0)P_S{}^0Q_S{}^0$;
- The real opportunity cost of the financial capital tied up in the land, $(r^0 - i_L{}^0)P_L{}^0Q_L{}^0$;
- The depreciation cost of the structure, $\delta^0(1 + i_S{}^0)P_S{}^0Q_S{}^0$;
- The maintenance and insurance costs associated with the structure, $\mu_S{}^0P_S{}^0Q_S{}^0$;
- The property taxes associated with the structure, $\tau_S{}^0P_S{}^0Q_S{}^0$, and
- The property taxes associated with the land underneath and surrounding the structure, $\tau_L{}^0P_L{}^0Q_L{}^0$.

The above user cost approach to pricing the services of a dwelling unit in period 0 can be applied to various *housing strata*, e.g., to detached dwellings, row houses or duplexes or town houses and apartment blocks. For the last two types of dwelling units, the land component for each individual dwelling unit needs to be constructed. For example, if there are 20 dwelling units in an apartment block, then the land share of each individual dwelling unit could be set to 1/20th of the total land area that the apartment block occupies.[33] Dwelling units can also be grouped according to their construction type, which could be primarily wood, brick, concrete or "traditional".

If a statistical agency produces national balance sheet estimates, then data on the total value of residential land and residential structures should be available. However, data on the quantity of residential land may not be known. Estimates of the country's total real stock of residential structures can be obtained by deflating the balance sheet estimate of the value of residential housing by the country's corresponding investment price deflator for residential housing.

[33] It is not completely straightforward to allocate the common land shared by the dwelling units into individual shares; i.e., instead of an equal division of the land, we could use the relative floor spaces of each apartment as the allocator. There are also problems associated with the relative height of the individual apartment units; i.e., an apartment on a higher floor will typically rent for more than an apartment on a lower floor.

There are at least two uses for the above user cost approach to pricing the services of housing:

- The user costs can be compared to market rents for dwelling units that are actually rented during the period under consideration and
- The user costs can be used to value the services of owner occupied housing.

As will be seen later in this section, it turns out that user costs do approximate market rents (for lower cost housing in the US at least), provided expectations of future inflation in house prices are formed in a certain way.

As mentioned before, *two main methods for valuing the services of owner occupied housing* have been suggested for National Accounts purposes: the user cost approach just explained and the rental equivalence approach. The rental equivalence approach is straightforward; for owner occupied houses in a certain stratum, we look for similar rented dwelling units and impute the market rental to the corresponding owner occupied house. In many countries, the rental equivalence approach works well, but it does not work well if rental markets are thin or if there are price controls on rents.

If user costs are used to value the services of owner occupied dwelling units in a country, then the maintenance and insurance rate term μ_S^0 in the user cost of structures formula (4.A22) should be dropped from the formula, since maintenance and insurance expenditures for owner occupied houses will generally be captured elsewhere in the household expenditure accounts.

The simplified approach to the user cost of housing explained above in equations (4.A17) and (4.A18) can be even further simplified by assuming that the ratio of the quantity of land to structures is fixed and so the aggregate user cost of housing is equal to $[r^{0*} + \delta + \mu + \tau]P_H^0$, where P_H is a quality adjusted price index that is applicable to the country's entire housing stock (including both structures and the underlying land) for the period under consideration and δ, μ and τ are respectively a depreciation rate, a maintenance and insurance rate and a property tax rate that applies to the composite of structures and land. Under this simplified approach to value the services of owner occupied housing, as was seen in the last paragraph above, the term μ should be dropped from the simplified user cost. The resulting *simplified approach* is applied in Iceland; see Gudnason (2004) and Gudnason and Jónsdóttir (2009)[34] and in some European countries; see the detailed exposition of the method by Katz (2009).[35]

[34] The real interest rate that is used is approximately 4% per year and the combined depreciation rate for land and structures is assumed to equal 1.25% per year. The depreciation rate for structures alone is estimated to be 1.5% per year. Property taxes are accounted for separately in the Icelandic CPI. Housing price information is provided by the State Evaluation Board based on property sales data of both new and old housing. The SEB also estimates the value of the housing stock and land in Iceland, using a hedonic regression model based on property sales data. The value of each household's dwelling is collected in the Household Budget Survey.

[35] Katz (2009) and Garner and Verbrugge (2009b; 176) give further references to the literature on the simplified user cost method.

A variant of this approach is used by the US Bureau of Economic Analysis: Lebow and Rudd (2003; 168) note that the US national accounts imputation for the services of owner occupied housing is obtained by applying *rent to value ratios* for tenant occupied housing to the stock of owner occupied housing with the same characteristics as the rented property.[36] The rent to value ratio can be seen as an estimate of the applicable real interest rate plus the depreciation rate plus a maintenance and insurance rate plus the property tax rate, $r^{0*} + \delta + \mu + \tau$.[37]

How exactly should the real interest rate, r^{0*}, be estimated? One possible method is to just make a reasonable guess:[38]

"The remaining question was what value of the real rate of return is appropriate? Evidence was presented to the task force that suggested that, at least in Western European countries, the appropriate real rate of return for owner-occupied dwellings was lower than that for other durables, perhaps in the 2.5 to 3.0 percent range. It was the consensus of the task force that given the actual situation in the CCs [Candidate Countries from Eastern Europe], real rates of return on both dwellings and land should be assumed to be 2.5 percent." Arnold J. Katz (2009; 46).

A second method is to use mortgage interest rates as estimates for the nominal opportunity cost of financial capital tied up in housing and to use econometric forecasting techniques to estimate predicted house price inflation rates (and then the real interest rate can be set equal to the nominal interest rate less the predicted house price inflation rate). Several variants of this second approach were tried by Verbrugge (2008) and Garner and Verbrugge (2009a) (2009b) using US data. However, as these authors show, this approach was not successful in that the resulting user cost estimates were extremely volatile (and frequently negative) and not at all close to corresponding market rents.

A third approach to the determination of an appropriate real interest rate to be used in a user cost formula for housing services was carried out by Garner and Verbrugge (2009b) using US data. They used applicable mortgage interest rates as estimates for the nominal opportunity cost of financial capital and used current period estimates of consumer price index inflation as their estimate of expected house price appreciation. Much to their surprise, they found that the resulting user costs tracked market rents rather well.[39] The conclusion is that either making a

[36] See also Crone, Nakamura and Voith (2009) and Garner and Short (2009; 237) for a description of this *capitalization method* for determining rental prices for housing units from estimates of the corresponding asset values. It can be seen that this method is actually a method for implementing the rental equivalence approach to valuing the services of owner occupied dwelling units.

[37] If an owned dwelling unit has the value V^0 and a rented dwelling unit with the same characteristics has the rent to value ratio $\gamma = r^{0*} + \delta + \mu + \tau$, then the imputed rent for the owned dwelling unit is set equal to $(\gamma - \mu)V^0 = (r^{0*} + \delta + \tau)V^0$, since insurance and maintenance expenditures on the owned dwelling will be recorded elsewhere in the System of National Accounts.

[38] The Australian Bureau of Statistics assumes a constant real interest rate equal to 4% per year when constructing its estimates of capital services.

[39] Using this approach, Garner and Verbrugge (2009b; 179) also found that there were no negative estimated user costs in their US data set.

reasonable guess for the real interest rate or using CPI inflation as a proxy for expected house price inflation gives rise to reasonable user costs that are likely to be fairly similar to market rents, at least for relatively inexpensive housing units.

It is evident that the main drivers for the user costs of structures and land are price indexes for new dwelling construction, P_S^t, and for residential land, P_L^t. Most statistical agencies have a constant quality price index for new residential structures, because this index is required in the national accounts in order to deflate investment expenditures on residential structures. This index could be used as an approximation to P_S^t.[40]

This completes the overview of the user cost approach to pricing residential housing services. In the following section, another approach to pricing the services of Owner Occupied Housing will be reviewed: the *opportunity cost approach*.

4.A.3 The Opportunity Cost Approach to the Valuation of Owner Occupied Housing Services

Recall the two main methods for valuing the services of owner occupied housing (OOH): the rental equivalence approach and the user cost approach. In the rental equivalence approach, an owner of a dwelling unit who chooses to live in it (or at least not rent it out to someone else) values the services of the dwelling by the market rent which is foregone. This is a very *direct opportunity cost* of using the dwelling. On the other hand, the user cost approach to valuing dwelling services is basically a *financial opportunity cost* of using the services of the dwelling unit during the period under consideration. It has been suggested that the true opportunity cost of using the services of an owned dwelling unit is the *maximum of the rent foregone and the user cost*:

"We conclude this section with the following (controversial) observation: perhaps the 'correct' opportunity cost of housing for an owner occupier is not his or her internal user cost but the *maximum* of the internal user cost and what the property could rent for on the rental market. After all, the concept of opportunity cost is supposed to represent the *maximum sacrifice* that one makes in order to consume or use some object and so the above point would seem to follow." W. Erwin Diewert (2009b; 113).

Diewert and Nakamura (2009) and Diewert, Nakamura and Nakamura (2009) pursued this opportunity cost approach to the valuation of owner occupied housing services in more detail but it can be seen that this approach seems to be a valid one. Moreover, it has the advantage of eliminating the problem with the user cost approach: namely, that the user cost approach can generate *negative* user costs if ex post or forecasted housing inflation rates are used in the user cost formula.

[40] This index may only be an approximation since it covers the construction of rental properties as well as owner occupied dwellings.

In practice, the opportunity cost approach to pricing OOH services may lead to similar results as the rental equivalence approach provided that expected inflation in the user cost formula is set equal to CPI inflation, since Garner and Verbrugge (2009b) show that for most low end rental properties, the rental equivalence and user cost approaches give much the same answer, at least in the US. However, there is evidence that user costs may be considerably higher than the corresponding market rentals for high end properties. Table 4.1 is taken from Heston and Nakamura (2009a; 113) (2009b; 277) and shows average market rent to market value of rental properties in a number of regions; i.e., it shows capitalization ratios as a function of the value of the rental property. Table 4.1 is based on a survey of US federal government employees conducted as part of a Safe Harbor process regarding the Cost of Living Allowance (COLA) program administered by the United States Office of Personnel Management. This program began in 1948 and pays an allowance above the federal salary schedule in three geographic areas (Alaska, the Caribbean and the Pacific) based on prices in these COLA areas relative to the Washington D.C. housing area.[41]

Table 4.1. Estimated Rent to Value Ratios as Percentages (Capitalization Ratios)

	Renter			
	Alaska	**Wash D.C.**	**Carib**	**Hawaii-Pacific**
Value($)	(1)	(2)	(3)	(4)
50,000	13.0	8.9	6.3	6.9
100,000	12.0	8.2	5.8	6.4
200,000	10.2	6.9	4.9	5.4
500,000	6.2	4.3	3.0	3.3

Two facts emerge from the Table 4.1:

- Capitalization ratios differ substantially across regions[42], and
- As one moves from inexpensive properties to more expensive properties the capitalization ratio for the high end properties is about one half the ratio for low end properties for all regions.

The second point listed above also emerges from the much more extensive US data on annual rents for the years 2004-2006 as a function of the corresponding home values found in Figure 1 in Garner and Verbrugge (2009b; 178). For a $100,000 home, the corresponding average

[41] This program is directed at comparing the costs of living for federal employees in the non-continental United States to Washington D.C. area. Housing is one of the most important and most difficult of the comparisons required under this program. The COLA areas include Alaska, Guam, Hawaii, Puerto Rico, and the U.S. Virgin Islands: a very diverse range of climates and housing needs.

[42] The relatively high capitalization ratios for Alaska may be due to the inclusion of heating services in the rent.

annual rent was about \$10,000 while for a \$900,000 home the corresponding average annual rent was about \$30,000. Thus the capitalization ratio fell from about 10% to about 3.3% as the home value increased from \$100,000 to \$900,000.

What factors could explain this dramatic drop in the capitalization ratio as we move from inexpensive properties to more expensive properties? As was indicated in section 4.A.2, the rent to value ratio can be regarded as an estimate of the applicable real interest rate plus the depreciation rate plus the property tax rate, $r^{0*} + \delta + \mu + \tau$, and these rates should not be all that different for properties of differing value. There are at least three possible explanations:

- High value properties may have a much higher proportion of land, hence the depreciation rate δ, regarded as a decline in value of the property due to aging of the structure, will be smaller as the land to structure ratio increases.[43]
- A substantial fraction of a landlord's monitoring, accounting and billing expenses may be in the nature of a fixed cost and hence these costs will drop as a fraction of the rent as the value of the property increases.
- Rentals of high value residential properties are not made on a commercial basis; i.e., they may be made on a temporary basis, with the renters serving as "house sitters" who pay somewhat subsidized rents as compared to the owner's financial opportunity cost.

It seems unlikely that the imperfect determination of the depreciation rate can explain the big decline in capitalization ratios as the value of the property increases; estimates of housing depreciation rates are generally in the 1 to 2% per year range,[44] and these rates are too low to fully explain the declines in the capitalization ratios. Similarly, the costs of maintaining and insuring a rental property that are collected in the term μ are likely to be relatively small and thus are unlikely to fully explain the phenomenon. Thus it may be that the third explanation is an important explanatory factor. If this is indeed the case, then the opportunity cost approach to the valuation of OOH services would give a much higher valuation to OOH services than the rental equivalence approach.[45]

[43] This explanation was suggested by Diewert (2009a; 486) and Garner and Verbrugge (2009b; 182).

[44] Garner and Verbrugge (2009b; 176) and Garner and Short (2009; 244) assume that annual depreciation rates (as fractions of the value of the property including both structures and land) is 1% per year.

[45] Thus the discrepancy between the rental equivalence approach to the valuation of OOH services and the opportunity cost approach may not be very important in the time series context because both measures may move in tandem. But in the context of making international comparisons, this argument will not be applicable due to the fact that the percentage of owner occupied dwelling units differs substantially across countries.

Chapter 5: Stratification or Mix Adjustment Methods

5.1 Simple Mean or Median Indexes

The simplest measures of house price change are based on some measure of central tendency from the distribution of house prices sold in a period, in particular the mean or the median. Since house price distributions are generally positively skewed (predominantly reflecting the heterogeneous nature of housing, the positive skew in income distributions and the zero lower bound on transaction prices), the median is typically used rather than the mean. As no data on housing characteristics are required to calculate the median, a price index that tracks changes in the price of the median house sold from one period to the next can be easily constructed. Another attraction of median indexes is that they are easy to understand.

An important drawback of simple median indexes is that they will provide noisy estimates of price change. The set of houses actually traded in a period, or a sample thereof, is typically small and not necessarily representative of the total stock of housing. Changes in the mix of properties sold will therefore affect the sample median price much more than the median price of the housing stock. For example, think of a city with two regions, A and B, and that region A has more expensive houses than region B. Suppose that the median house sold in 2006 and 2008 comes from region A, while the median house in 2007 comes from region B. It follows that the median index could record a large rise from 2006 to 2007 and then a large fall from 2007 to 2008. Such an index would be a very poor indicator of what is actually happening in the housing market. Thus, a median (or mean) index will be a very inaccurate guide to price change when there is substantial change in the composition of houses sold between periods. If there is a correlation between turning points in house price cycles and compositional change, then a median could be especially misleading in periods when the premium on accuracy is highest.

A perhaps bigger problem than short-term noise is systematic error, or bias. A simple median index will be subject to bias when the quality of the housing stock changes over time. The median index will be upward biased if the average quality improves over the years. Bias can also arise if certain types of houses are sold more frequently than other types of houses and at the same time exhibit different price changes. For example, when higher quality houses sell more frequently and also rise in price faster than lower quality houses, a downward bias may result if the number of sales per type of house does not properly reflect the number of houses in stock. This is sometimes referred to as a sample selection problem. The fact that houses traded are usually a small and not necessarily representative part of the total housing stock can bias other property price index methods as well, including hedonic and repeat sales methods (to be discussed in Chapters 6 and 7).

5.2 Stratification

Post-stratification of a sample is a general technique for reducing sample selection bias. In the case of residential property price indexes, stratification is the simplest tool for controlling for changes in the composition or “quality mix” of the properties sold. The method is therefore also known as mix adjustment. Stratification is also needed if users desire price indexes for different housing market segments.

Stratification is nothing else than separating the total sample of houses into a number of sub-samples or strata. After constructing a measure of the change in the central tendency for each stratum, such as a mean or median price index, the aggregate mix-adjusted RPPI is typically calculated as a weighted average of indexes for each stratum. With M different strata, the mix-adjusted index, as calculated in practice in various countries, can be written in mathematical form as follows:

$$P^{0t} = \sum_{m=1}^{M} w_m^0 P_m^{0t} , \qquad (5.1)$$

where P_m^{0t} is the index for stratum m which compares the mean (median) price in the current or comparison period t with the mean (median) price in an earlier or base period 0, and where w_m^0 denotes the weight of stratum m. The weights are value shares pertaining to the strata. They refer to the base period, which is usually a year (whereas the comparison periods may be months or quarters). For practical reasons, the weights are often kept fixed for several years, but keeping weights fixed for a long time is generally not good practice. Section 5.3 provides more details on aggregation and weighting issues in this context.

Which type of value weights is used, depends on the target index that the RPPI is supposed to estimate. If the purpose is to track the price change of the housing stock then obviously stock-weights – the stock value shares of the strata – should be used. If, on the other hand, the target is a sales or acquisitions RPPI, then sales (expenditure) weights should be applied.[1]

The effectiveness of stratification will depend upon the stratification variables used because a mix-adjusted measure only controls for compositional change across the various groups. For example, if house sales are separated solely according to their location, a mix-adjusted index will control for changes in the mix of property types across the defined locations. But the mix-adjusted measure will not account for any changes in the mix of property types sold that are unrelated to location. Also, a mix-adjusted index does not account for changes in the mix of properties sold within each subgroup, in this case changes in the mix of properties sold within the boundaries of each location.

[1] The house price indexes compiled in the EU as part of a Eurostat pilot study are examples of such acquisitions indexes (see Makaronidis and Hayes, 2006 or Eurostat, 2010).

Very detailed stratification according to housing characteristics such as size of the structure, plot size, type of dwelling, location and amenities will increase homogeneity and thus reduce the quality-mix problem, although some quality mix changes will most likely remain. There is, however, a tradeoff to be considered. Increasing the number of strata reduces the average number of observations per stratum, and a very detailed stratification might raise the standard error of the overall RPPI. Needless to say, a detailed stratification scheme can be constructed only if the strata-defining characteristics are available for all sample data. Another potential practical problem is that it might be difficult to obtain accurate data on the (stock) weights for small subgroups.

When using only physical and locational stratification variables, like those mentioned above, then the stratification method does not control for quality changes of the individual properties. By quality changes we mean the effect of renovations and remodeling done to the properties in combination with depreciation of the structures. This can also be called ‘net depreciation’. Depreciation obviously depends on the age of the structure, although depreciation rates may differ across different types of dwellings or even across different locations. This is why age of the structure was listed in section 4.4 as one of the most important price determining quality attributes. Consequently, stratifying according to age class may help to reduce the problem of quality change.

Introducing age class as another stratification variable will further reduce the average number of observations per stratum and may give rise to unreliable estimates of price changes. Under these circumstances, hedonic regression techniques – which are discussed in Chapter 5 of this Handbook – will generally work better than stratification. As mentioned earlier, some sort of hedonic regression method will also be needed to decompose the overall RPPI into land and structures components if this is required for any of the purposes discussed in Chapter 3. Such a decomposition cannot be provided by stratification methods.

Mix-adjusted RPPIs have been compiled by numerous statistical offices and other government agencies, including the UK Department of the Environment (1982) and the Australian Bureau of Statistics (ABS, 2006). While mix adjustment has received relatively little attention in the academic literature,[2] there is a growing body of work on market segmentation using statistical techniques like cluster analysis and factor analysis; see e.g. Dale-Johnson (1982), Goodman and Thibodeau (2003), and Thibodeau (2003). These techniques could in principle be used to define housing sub-markets, which could subsequently be used as strata for the construction of a mix-adjusted RPPI. The Australian Bureau of Statistics experimented with this approach (ABS, 2005).

[2] However, stratified median house price indexes have been used by several researchers, mostly for comparison purposes; see e.g. Mark and Goldberg (1984), Crone and Voith (1992), Gatzlaff and Ling (1994), and Wang and Zorn (1997).

Prasad and Richards (2006) (2008) proposed a novel stratification method and tested it on an Australian data set. They grouped together suburbs according to the long-term average price level of dwellings in those regions, rather than just clustering smaller geographic regions into larger regions. Their method of stratification was specifically designed to control for what may be the most important form of compositional change, namely changes in the proportion of houses sold in higher- and lower-priced regions in any period.[3] Note that they used median price indexes at the stratum level. McDonald and Smith (2009) followed-up on this study and constructed a similar stratified median house price measure for New Zealand.

5.3 Aggregation and Weighting Issues

First-stage aggregation

Stratification involves a two-stage procedure: price indexes are compiled at the stratum level, which are then aggregated across the various strata. As was mentioned in Section 5.1, median strata indexes have typically been used, in particular because they will often be more stable than the corresponding mean indexes. Yet, we will focus on mean prices rather than medians. Conventional index number theory deals with aggregation issues, in this case aggregation of house price observations within strata. Unlike the median, means are aggregator functions, which link up with index number theory. The question then arises: what kind of mean should be taken?

The *Consumer Price Index Manual* (2004) makes recommendations about how to construct price indexes at the first stage of aggregation if information on quantities is unavailable and then at the second stage of aggregation when both price and value (or quantity) information is available. At the first stage of aggregation, Chapter 20 in the *Consumer Price Index Manual* generally recommends using the unweighted geometric mean or Jevons index to aggregate individual price quotations into an index. However, this general advice is not applicable in the present context.

If the aim is to construct a price index for the sales of residential properties, the appropriate concept of (elementary) price in some time period t for a homogeneous stratum or cell in the stratification scheme is a *unit value*. Because each sale of a residential property comes with its own quantity, which is equal to one, the corresponding quantity for that cell is the simple *sum* of the properties transacted in period t. We can formally describe this as follows. Suppose that in period t there are $N(t,m)$ property sales observed in a particular cell m, with the selling

[3] A general rule is that stratification according to the variable of interest should not be used since that can lead to biased results. The study variable used by Prasad and Richards (2006) (2008) is (long-term) house price *change*, not house price level, so their stratification method could perhaps be defended. However, little is known about the statistical properties of this type of stratification index and it would be advisable to investigate the issue of potential bias before producing such an index.

price (value) of property *n* equal to V_n^t for $n = 1,...,N(t,m)$. Then the appropriate price and quantity for cell *m* in period *t* are:

$$P_m^t \equiv \sum_{n=1}^{N(t,m)} V_n^t / N(t,m) ; \tag{5.2}$$

$$Q_m^t \equiv N(t,m) . \tag{5.3}$$

This narrowly defined unit value concept is actually recommended in the CPI Manual (2004; 356). If the stratification scheme leads to cells that are not sufficiently narrow defined, then of course, some unit value bias may arise, which is equivalent to saying that some quality mix bias may remain.[4]

Second-stage aggregation

The next issue to be resolved is: what index number formula should be used to aggregate the elementary prices and quantities into one overall RPPI? The CPI Manual discusses this choice of formula issue at great length. A number of index number formulae are recommended but a good overall choice appears to be the Fisher ideal index since this index can be justified from several different perspectives.[5] The Fisher index is the geometric mean of the Laspeyres and Paasche indexes.

To illustrate this point, let $P^t \equiv [P_1^t,...,P_M^t]$ and $Q^t \equiv [Q_1^t,...,Q_M^t]$ denote the period *t* vectors of cell prices and quantities. The Laspeyres price index, P_L^{st}, going from (the base) period *s* to (the comparison) period *t* can be defined as follows:

$$P_L^{st}(P^s, P^t, Q^s) \equiv \frac{\sum_{m=1}^{M} P_m^t Q_m^s}{\sum_{m=1}^{M} P_m^s Q_m^s} ; \tag{5.4}$$

Note that equation (5.4) can be rewritten in the form of (5.1) if $s = 0$ with cell price indexes $P_m^{0t} = P_m^t / P_m^0$ and value shares $w_m^0 = P_m^0 Q_m^0 / \sum_{m=1}^{M} P_m^0 Q_m^0$. The Paasche price index going from period *s* to *t*, P_P^{st}, is defined as follows:

$$P_P^{st}(P^s, P^t, Q^t) \equiv \frac{\sum_{m=1}^{M} P_m^t Q_m^t}{\sum_{m=1}^{M} P_m^s Q_m^t} . \tag{5.5}$$

[4] In practice, crude stratification according to region and type of dwelling is often used. The stratification method according to price bands proposed by Prasad and Richards (2008), could be useful to militate against unit value bias. See Balk (1998) (2008; 72-74), Silver (2009a) (2009b) (2010), and Diewert and von der Lippe (2010) for more general discussions of unit value bias.

[5] See CPI Manual (2004; Chapters 15-18) for alternative justifications for the use of the Fisher formula.

The Fisher price index for period *t* relative to period *s*, P_F^{st}, can be defined as the geometric mean of (5.4) and (5.5):

$$P_F^{st}(P^s, P^t, Q^s, Q^t) \equiv [P_L^{st}(P^s, P^t, Q^s) P_P^{st}(P^s, P^t, Q^t)]^{1/2}. \quad (5.6)$$

Recall that all the quantities occurring in these three formulas are numbers of transactions; that is, numbers of observed prices. Thus, for calculating a Laspeyres, Paasche, or Fisher price index one needs the same information.

The Laspeyres, Paasche and Fisher price indexes defined by equations (5.4), (5.5) and 5.6) are *fixed base indexes*. For example, if there are 3 periods of sales data, including the base period 0, then the Fisher formula (5.6) would generate the following index number series for those 3 periods:

$$1; P_F^{01}(P^0, P^1, Q^0, Q^1); P_F^{02}(P^0, P^2, Q^0, Q^2). \quad (5.7)$$

Chaining

An alternative to the fixed base method is the use of chaining. The *chain method* uses the data of the last two periods to calculate a period to period chain link index which is used to update the index level from the previous period. Chaining would, for example, generate the following Fisher index number series for the 3 periods:

$$1; P_F^{01}(P^0, P^1, Q^0, Q^1); P_F^{01}(P^0, P^1, Q^1, Q^1) P_F^{12}(P^1, P^2, Q^1, Q^2). \quad (5.8)$$

The next issue to be discussed is whether RPPIs should be constructed by using fixed base or chain indexes. Both the System of National Accounts and the Consumer Price Index Manual recommend the use of chain indexes provided that the underlying price data have reasonably smooth trends.[6] On the other hand, if there is a great deal of variability in the data, particularly when prices bounce erratically around a trend, the use of fixed base indexes is recommended. Property price changes tend to be fairly smooth,[7] so it is likely that chained indexes will work well in many cases. However, more experimentation with actual data is required in order to give definitive advice on this issue. There may also be seasonal variation in house prices as the example for the Dutch town of "A", presented in section 5.5, suggests. In such cases too, one should be careful with using chain indexes.

Stock RPPIs

---

[6] See SNA (1993; 388) and CPI Manual (2004; 349).

[7] Although prices do not bounce around erratically in the real estate context, quantities do exhibit considerable variability, particularly if there are a large number of cells in the stratification setup with a limited number of observations in each cell. There is also a considerable amount of seasonal variation in quantities; i.e., sales of residential properties fall off dramatically during the winter months of the year.

The above discussion was on the construction of a price index for the *sales* of residential properties when a stratification method is used. But how should a RPPI be constructed for the *stock* of residential properties? Assuming that, for each cell m, the properties sold are random (or 'representative') selections from the stock of dwelling units defined by cell m, the period t unit value prices P_m^t defined by (5.2) can still be used as (estimates of the) cell prices for a stock RPPI. The quantities Q_m^t defined by (5.3) are, however, no longer appropriate; they need to be replaced by (estimates of) the number of dwelling units of the type defined by cell m that are in the reference stock at time t, say Q_m^{t*}, for $m = 1,...,M$. With these *population* quantity weights, the rest of the details of the index construction are the same as was the case for the sales RPPI.

To compile stock weights, it will be necessary to have a periodic census of the housing stock with enough details on the properties so that it can be decomposed into the appropriate cells in the stratification scheme for a base period. If information on new house construction and on demolitions is available in a timely manner, then the census information can be updated and estimates for the housing stock by cell (the Q_m^{t*}) can be made in a timely manner. The stock RPPI can be constructed using a (chained) Fisher index as was the case for the sales RPPI. On the other hand, if timely data on new construction and demolitions is lacking, it will only be possible to construct a fixed base Laspeyres index using quantity data from the last available housing census (in say period 0), $Q^{0*} = [Q_1^{0*},...,Q_M^{0*}]$, until information from a new housing census is made available (in say period T). The Laspeyres stock RPPI thus is

$$P_L^{0t}(P^0, P^t, Q^{0*}) \equiv \frac{\sum_{m=1}^{M} P_m^t Q_m^{0*}}{\sum_{m=1}^{M} P_m^0 Q_m^{0*}}. \qquad t = 0,...,T \qquad (5.9)$$

As mentioned in Chapter 4, for some purposes it is useful to have a stock RPPI for Owner Occupied Housing, i.e. excluding rented homes. The construction of such an index proceeds in the same way as for the construction of a RPPI for the entire housing stock except that the cells in the stratification scheme are now restricted to owned dwellings. This will be possible if the periodic housing census collects information on whether each dwelling unit is owned or rented.

It should be noted that the construction of a stratified (stock or sales) RPPI becomes more complex when some of the cells in the stratification scheme are empty for some periods. In section 5.5, where an empirical example using data on housing sales for the Dutch town of "A" is presented, a matched-model approach will be outlined that can be used in case some cells are empty.

5.4 Main Advantages and Disadvantages

We will summarize the main advantages and disadvantages of the stratified median or mean approach. The main advantages are:

- Depending on the choice of stratification variables, the method adjusts for compositional change of the dwellings.
- The method is reproducible, conditional on an agreed list of stratification variables.
- Price indexes can be constructed for different types and locations of housing.
- The method is relatively easy to explain to users.

The main disadvantages of the stratified median or mean method are:

- The method cannot deal adequately with depreciation of the dwelling units unless age of the structure is a stratification variable.
- The method cannot deal adequately with units that have undergone major repairs or renovations (unless renovations are a stratification variable).
- The method requires information on housing characteristics so that sales transactions can be allocated to the correct strata.
- If the classification scheme is very coarse, compositional changes will affect the indexes, i.e., there may be some unit value bias in the indexes.
- If the classification scheme is very fine, the cell indexes may be subject to a considerable amount of sampling variability due to small sample sizes or some cells may be empty for some periods causing index number difficulties.

An overall evaluation of the stratification method is that it can be satisfactory if:

- an appropriate level of detail is chosen;
- age of the structure is one of the stratification variables, and
- a decomposition of the index into structure and land components is not required.

Stratification can be interpreted as a special case of regression.[8] Chapter 6 discusses this more general technique, known as hedonic regression when applied to price index construction and quality adjustment.

[8] See Diewert (2003a) who showed that stratification techniques or the use of dummy variables can be viewed as a nonparametric regression technique. In the statistics literature, these partitioning or stratification techniques are known as analysis of variance models; see Scheffé (1959).

5.5 An Example Using Dutch Data for the Town of "A"

This chapter will be concluded by a worked example for the construction of a stratified index using data on sales of detached houses for a small town (the population is around 60,000) in the Netherlands, town "A", for 14 quarters, starting in the first quarter of 2005 and ending in the second quarter of 2008. The same data set will be exploited in Chapters 6, 7, 8 and 9 to illustrate the other methods for constructing house price indexes and the numerical differences that can arise in practice.[9]

A dwelling unit has a number of important *price determining characteristics*:

- The land area L of the property;
- The floor space area S of the structure; i.e., the size of the structure that sits on the land underneath and surrounding the structure;
- The age A of the structure; this determines (on average) how much physical deterioration or depreciation the structure has experienced;
- The amount of renovations that have been undertaken for the structure;
- The location of the structure; i.e., its distance from amenities such as shopping centers, schools, restaurants and work place locations;
- The type of structure; i.e., single detached dwelling unit, row house, low rise apartment or high rise apartment or condominium;
- The type of construction used to build the structure;
- Other special price determining characteristics that are different from "average" dwelling units in the same general location such as swimming pools, air conditioning, elaborate landscaping, the height of the structure or views of oceans or rivers.

The variables used in this study can be described as follows:

- V_n^t is the selling price of property n in quarter t in Euros;
- L_n^t is the area of the plot for the sale of property n in quarter t in meters squared;
- S_n^t is the living space area of the structure for the sale of property n in quarter t in meters squared;
- A_n^t is the approximate age (in decades) of the structure on property n in quarter t.

It can be seen that not all of the price determining characteristics listed above were used in the present study. In particular, the last five sets of characteristics of the property were neglected. There is an implicit assumption that quarter to quarter changes in the amount of renovations

[9] This material is drawn from Diewert (2010).

that have been undertaken for the structures, the location of the house, the type of structure, the type of construction and any other price determining characteristics of the properties sold in the quarter did not change enough to be a significant determinant of the average price for the properties sold once changes in land size, structure size and the age of the structures were taken into account.[10]

The determination of the values for the age variable A_n^t needs some explanation. The original data were coded as follows: if the structure was built in 1960-1970, then the observation was assigned the decade indicator variable BP = 5; 1971-1980, BP=6; 1981-1990, BP=7; 1991-2000, BP=8; 2001-2008, BP=9. The age variable in this study was set equal to 9 − BP. For a recently built structure n in quarter t, A_n^t = 0. Thus, the age variable gives the (approximate) age of the structure in decades.

Houses which were older than 50 years at the time of sale were deleted from the data set. Two observations which had unusually low selling prices (36,000 and 40,000 Euros) were deleted as were 28 observations which had land areas greater than 1200 m^2. No other outliers were deleted from the sample. After this cleaning of the data, we were left with 2289 observations over the 14 quarters in the sample, or an average of 163.5 sales of detached dwelling units per quarter. The overall sample mean selling price was 190,130 Euros, whereas the median price was 167,500 Euros. The average plot size was 257.6 m^2 and the average size of the structure (living space area) was 127.2 m^2. The average age of the properties sold was approximately 18.5 years.

The stratification approach to constructing a house price index is conceptually very simple: for each of the important price explaining characteristic, divide up the sales into relatively homogeneous groups. Thus in the present case, sales were classified into 45 groups or cells, consisting of 3 groupings for the land area L, 3 groupings for the structure area S and 5 groups for the age A (in decades) of the structure (3×3×5 = 45 separate cells). Once quarterly sales were classified into the 45 groupings of sales, the sales within each cell in each quarter were summed and then divided by the number of units sold in that cell in order to obtain unit value prices, the cell prices P_m^t. These unit values were then combined with the number of units sold in each cell, the Q_m^t, to form the usual p's and q's that can be inserted into a bilateral index number formula, like the Laspeyres, Paasche and Fisher ideal formulae defined by (5.4)-(5.6)

[10] To support this assumption, it should be noted that the hedonic regression models discussed in later chapters consistently explained 80-90% of the variation in the price data using just the three main explanatory variables: L, S and A. The R^2 between the actual and predicted selling prices ranged from .83 to .89. The fact that it was not necessary to introduce more price determining characteristics for this particular data set can perhaps be explained by the nature of the location of the town of "A" on a flat, featureless plain and the relatively small size of the town; i.e., location was not a big price determining factor since all locations have more or less the same access to amenities.

above,[11] yielding a stratified index of house prices of each of these types. However, since there are only 163 or so observations for each quarter and 45 cells to fill, each cell had only an average of 3 or so observations in each quarter, and some cells were empty for some quarters. This problem will be addressed subsequently.

How should the size limits for the *L* and *S* groupings be chosen? One approach would be to divide the range of *L* and *S* by three and create three equal size cells. However, this approach leads to a large number of observations in the middle cells. In the present study, size limits were therefore chosen such that roughly 50% of the observations would fall into the middle sized categories and roughly 25% would fall into the small and large categories. For the land size variable *L*, the cutoff points chosen were 160 m^2 and 300 m^2, while for the structure size variable *S*, the cutoff points chosen were 110 m^2 and 140 m^2. Thus if $L < 160$ m^2, then the observation fell into the small land size cell; if 160 $m^2 \le L < 300$ m^2, then the observation fell into the medium land size cell and if 300 $m^2 \le L$, then the observation fell into the large land size cell. The resulting sample probabilities for falling into these three *L* cells over the 14 quarters were .24, .51 and .25 respectively. Similarly, if $S < 110$ m^2, the observation fell into the small structure size cell; if 110 $m^2 \le S < 140$ m^2, then the observation fell into the medium structure size cell and if 140 $m^2 \le S$, then the observation fell into the large structure size cell. The resulting sample probabilities for falling into these three *S* cells over the 14 quarters were .21, .52 and .27 respectively.

As mentioned earlier, the data that were used did not have an exact age for the structure; only the decade when the structure was built was recorded. So there was no possibility of choosing exact cutoff points for the age of the structure. $A = 0$ corresponds to houses that were built during the years 2001-2008; $A = 1$ for houses built in 1991-2000; $A = 2$ for houses built in 1981-1990, $A = 3$ for houses built in 1971-1980; and $A = 4$ for houses built in 1961-1970. The resulting sample probabilities for falling into these five cells over the 14 quarters were .15, .32, .21, .20 and .13 respectively. See Table 1 for the sample joint probabilities of a house sale belonging to each of the 45 cells.

There are several points of interest to note about Table 5.1:

- There were no observations for houses built during the 1960s ($A = 4$) which had a small lot (*L* = small) and a large structure (*S* = large), so this cell is entirely empty;

[11] The international manuals on price measurement recommend this unit value approach to the construction of price indexes at the first stage of aggregation; see CPI Manual (2004), PPI Manual (2004), and XMPI Manual (2009). However, the unit value aggregation should take place over homogeneous items and this assumption may not be fulfilled in the present context, since there is a fair amount of variability in *L*, *S* and *A* within each cell. But since there are only a small number of observations in each cell for the data set under consideration, it would be difficult to introduce more cells to improve homogeneity since this would lead to an increased number of empty cells and a lack of matching for the cells.

- There are many cells which are almost empty; in particular the probability of a sale of a large plot with a small house is very low as is the probability of a sale of a small plot with a large house;[12]
- The "most representative model" sold over the sample period corresponds to a medium sized lot, a medium sized structure and a house that was built in the 1990s ($A = 1$). The sample probability of a house sale falling into this highest probability cell is 0.09262.

Table 5.1. Sample Probability of a Sale in Each Cell

L	L	$A = 0$	$A = 1$	$A = 2$	$A = 3$	$A = 4$
small	small	0.00437	0.02665	0.01660	0.02053	0.02097
medium	small	0.00349	0.02840	0.01966	0.01092	0.03888
large	small	0.00087	0.00175	0.00044	0.00218	0.00612
small	medium	0.01223	0.05242	0.04281	0.02053	0.00699
medium	medium	0.03277	0.09262	0.08869	0.07907	0.02141
large	medium	0.00786	0.02315	0.01005	0.01442	0.01398
small	large	0.00306	0.00218	0.00175	0.00568	0.00000
medium	large	0.03145	0.03495	0.00786	0.02097	0.00306
large	large	0.04893	0.05461	0.02315	0.02490	0.01660

The average selling price of the representative house, falling into the medium L, medium S and $A = 1$ category, is graphed in Figure 1 along with the overall sample mean and median price in each quarter. These average prices have been converted into indexes which start at 1 for quarter 1, which is the first quarter of 2005. It should be noted that these three house price indexes are rather variable.

Some additional indexes are plotted in Figure 1, including a fixed base matched model Fisher ideal index and a chained matched model Fisher ideal price index. It is necessary to explain what a "matched model" index in this context means. If at least one house was sold in each quarter for each of the 45 cells, then the ordinary Laspeyres, Paasche and Fisher price indexes comparing the prices of quarter t to those of quarter s would be defined by equations (5.4)-(5.6) respectively, where M = 45. This algebra is applicable to the situation where there are transactions in all cells for the two quarters being compared. But for the present data set, on average only about 30 out of the 45 categories can be matched across any two quarter, and the formulae (5.4)-(5.6) need to be modified in order to deal with this *lack of matching problem*. Thus, when considering how to form an index number comparison between quarters s and t,

[12] Thus lot size and structure size are positively correlated with a correlation coefficient of .6459. Both L and S are fairly highly correlated with the selling price variable P: the correlation between P and L is .8234 and between P and S is .8100. These high correlations lead to multicollinearity problems in the hedonic regression models to be considered later.

define the set of cells m that have at least one transaction in each of quarters s and t as the set $S(s,t)$. Then the *matched model counterparts*, P_{ML}^{st}, P_{MP}^{st} and P_{MF}^{st}, to the regular Laspeyres, Paasche and Fisher indexes between quarters s and t given by (5.4), (5.5) and (5.6) are defined as follows:[13]

$$P_{ML}^{st} \equiv \frac{\sum_{m\in S(s,t)} P_m^t Q_m^s}{\sum_{m\in S(s,t)} P_m^s Q_m^s}; \tag{5.10}$$

$$P_{MP}^{st} \equiv \frac{\sum_{m\in S(s,t)} P_m^t Q_m^t}{\sum_{m\in S(s,t)} P_m^s Q_m^t}; \tag{5.11}$$

$$P_{MF}^{st} \equiv [P_{ML}^{st} P_{MP}^{st}]^{1/2}. \tag{5.12}$$

In Figure 5.1, the Fixed Base Fisher index is the matched model Fisher price index defined by (5.12), where the base period s is kept fixed at quarter 1; i.e., the indexes $P_{MF}^{1,1}$, $P_{MF}^{1,2}$,…, $P_{MF}^{1,14}$ are calculated and labeled as the Fixed Base Fisher Index, P_{FFB}. The index that is labeled the matched model Chained Fisher Index, P_{FCH}, is the price index $P_{MF}^{1,1}$, $P_{MF}^{1,1}\,P_{MF}^{1,2}$, $P_{MF}^{1,1}\,P_{MF}^{1,2}\,P_{MF}^{2,3}$, …, $P_{MF}^{1,1}\,P_{MF}^{1,2}\ldots\,P_{MF}^{13,14}\,P_{MF}^{1,14}$. Notice that the Fixed Base and Chained (matched model) Fisher indexes are quite close to each other and are much smoother than the corresponding Mean, Median and Representative Model indexes.[14] The data for these 5 series plotted in Figure 5.1 are listed in Table 5.2.

The matched model Fisher indexes must be regarded as being more accurate than the other indexes which use only a limited amount of the available price and quantity information. As the trend of the Fisher indexes is fairly smooth, the chained Fisher index should be preferred over the fixed base Fisher index, following the advice given in Hill (1988) (1993) and in the CPI Manual (2004). Recall also that there is no need to use Laspeyres or Paasche indexes in this situation since data on sales of houses contains both value and quantity information. Under these conditions, Fisher indexes are preferred over the Laspeyres and Paasche indexes

[13] A justification for this approach to dealing with a lack of matching in the context of bilateral index number theory can be found in the discussion by Diewert (1980; 498-501) on the related problem of dealing with new and disappearing goods. Other approaches are also possible. For approaches based on maximum matching over all pairs of periods; see Ivancic, Diewert and Fox (2011) and de Haan and van der Grient (2011) for approaches based on imputation methods; see Alterman, Diewert and Feenstra (1999). A useful imputation approach could be to estimate imputed prices for the empty cells using hedonic regressions. The discussion is left until various hedonic regression methods have been discussed.

[14] The means (and standard deviations) of the 5 series mentioned thus far are as follows: P_{FCH} = 1.0737 (0.0375), P_{FFB} = 1.0737 (0.0370), P_{Mean} = 1.0785 (0.0454), P_{Median} = 1.0785 (0.0510), and $P_{Represent}$ = 1.0586 (0.0366). Thus the representative model price index has a smaller variance than the two matched model Fisher indexes but it has a substantial bias relative to the two matched model Fisher indexes: the representative model price index is well below the Fisher indexes for most of the sample period.

(which do not use all of the available price and quantity information for the two periods being compared).

Figure 5.1. Matched Model Fisher Chained and Fixed Base Price Indexes, Mean, Median and Representative Model Price Indexes

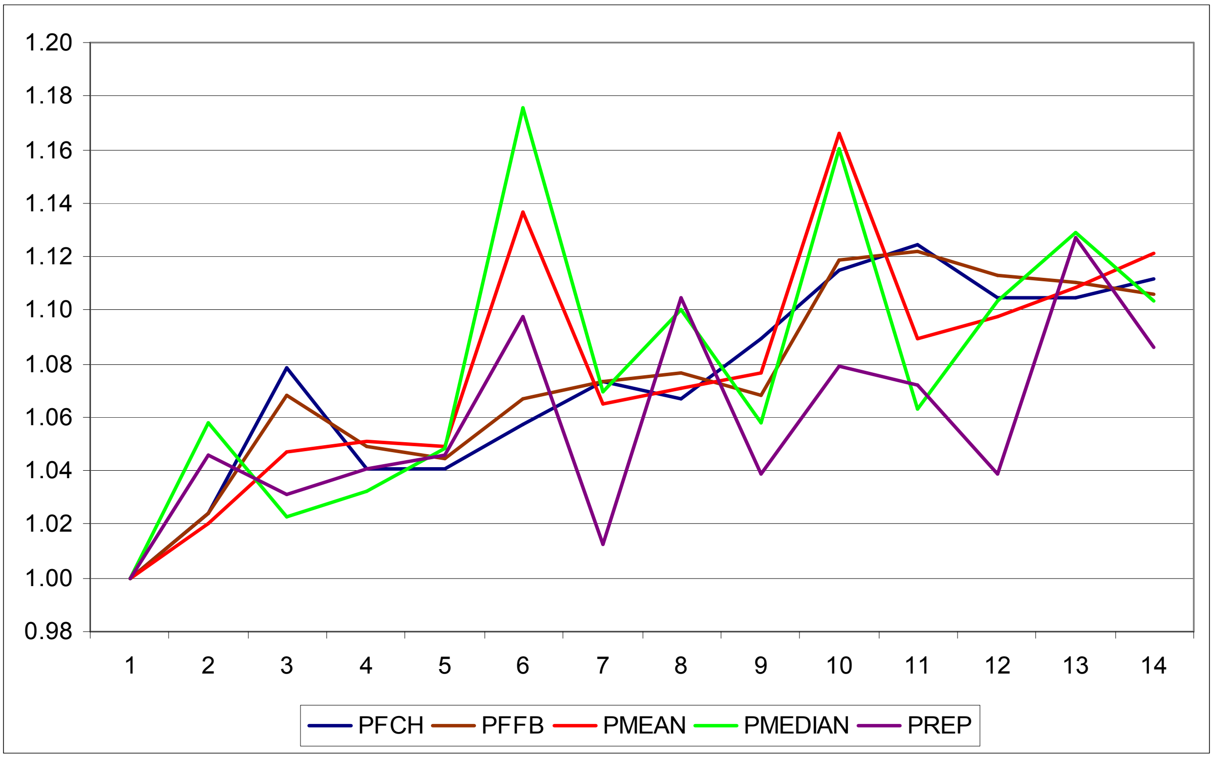


Table 5.2. Matched Model Fisher Chained and Fixed Base Price Indexes, Mean, Median and Representative Model Price Indexes

Quarter	P_{FCH}	P_{FFB}	P_{Mean}	P_{Median}	$P_{Represent}$
1	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.02396	1.02396	1.02003	1.05806	1.04556
3	1.07840	1.06815	1.04693	1.02258	1.03119
4	1.04081	1.04899	1.05067	1.03242	1.04083
5	1.04083	1.04444	1.04878	1.04839	1.04564
6	1.05754	1.06676	1.13679	1.17581	1.09792
7	1.07340	1.07310	1.06490	1.06935	1.01259
8	1.06706	1.07684	1.07056	1.10000	1.10481
9	1.08950	1.06828	1.07685	1.05806	1.03887
10	1.11476	1.11891	1.16612	1.16048	1.07922
11	1.12471	1.12196	1.08952	1.06290	1.07217
12	1.10483	1.11321	1.09792	1.10323	1.03870
13	1.10450	1.11074	1.10824	1.12903	1.12684
14	1.11189	1.10577	1.12160	1.10323	1.08587

Since there is a considerable amount of heterogeneity in each cell of the stratification scheme, there is the strong possibility of some unit value bias in the matched model Fisher indexes. However, if a finer stratification were used, the amount of matching would drop dramatically. Already, with the present stratification, only about 2/3 of the cells could be matched across any two quarters. There is a trade-off between having too few cells with the possibility of unit value bias and having a more detailed stratification scheme but with a much smaller degree of matching of the data within cells across the two time periods being compared.

Looking at Table 5.2 and Figure 5.1, it can be seen that the chained Fisher index shows a drop in house prices during the fourth quarters of 2005, 2006 and 2007. There is a possibility that house prices drop for seasonal reasons in the fourth quarter of a year. In order to deal with this possibility, in section 5.6 a rolling year matched model Fisher index will be constructed.

5.6 The Treatment of Seasonality for the Dutch Example

Assuming that each commodity in each season of the year is a separate "annual" commodity is the simplest and theoretically most satisfactory method for dealing with seasonal goods when the goal is to construct *annual* price and quantity indexes. This idea can be traced back to Mudgett in the consumer price context and to Stone in the producer price context:

"The basic index is a yearly index and as a price or quantity index is of the same sort as those about which books and pamphlets have been written in quantity over the years." Bruce D. Mudgett (1955; 97).

"The existence of a regular seasonal pattern in prices which more or less repeats itself year after year suggests very strongly that the varieties of a commodity available at different seasons cannot be transformed into one another without cost and that, accordingly, in all cases where seasonal variations in price are significant, the varieties available at different times of the year should be treated, in principle, as separate commodities." Richard Stone (1956; 74-75).

Diewert (1983) generalized the Mudgett-Stone annual framework to allow for *rolling year comparisons* for 12 consecutive months of data with a base year of 12 months of data or for comparisons of 4 consecutive quarters of data with a base year of 4 consecutive quarters of data; i.e., the basic idea is to compare the current rolling year of price and quantity data to the corresponding data of a base year where the data pertaining to each season is compared.[15] In the present context, we have in principle,[16] price and quantity data for 45 classes of housing

[15] For additional theory and examples of this rolling year approach, see the chapters on seasonality in the CPI Manual (2004) and the PPI Manual (2004), Diewert (1998), and Balk (2008; 151-169). To justify the rolling year indexes from the viewpoint of the economic approach to index number theory, some restrictions on preferences are required; details can be found in Diewert (1999; 56-61). It should be noted that weather and the lack of fixity of Easter can cause "seasons" to vary and a breakdown in the approach; see Diewert, Finkel and Artsev (2009). However, with quarterly data, these limitations of the rolling year index are less important.

[16] In practice, as we have seen in the previous section, many of the cells are empty in each period.

commodities in each quarter. If the sale of a house in each season is treated as a separate good, then there are 180 annual commodities.

For the first index number value, the four quarters of price and quantity data on sales of detached dwellings in the town of "A" (180 series) are compared with the same data using the Fisher ideal formula. Naturally, the resulting index is equal to 1. For the next index number value, the data for the first quarter of 2005 are dropped and the data pertaining to the first quarter of 2006 are appended to the data for quarters 2-4 of 2005. The resulting Fisher index is the second entry in the Rolling Year (RY) Matched Model series that is illustrated in Figure 5.2. However, as was the case with the chained and fixed base Fisher indexes that appeared in Figure 5.1, not all cells could be matched using the rolling year methodology; i.e., some cells were empty in the first quarter of 2006 which corresponded to cells in the first quarter of 2005 which were not empty and vice versa. So when constructing the rolling year index P_{RY} plotted in Figure 5.2, the comparison between the rolling year and the data pertaining to 2005 was restricted to the set of cells which were non empty in both years; i.e., the Fisher rolling year indexes plotted in Figure 5.2 are matched model indexes. Unmatched models are omitted from the index number comparison.[17]

The results are shown in Figure 5.2. Note that there is a definite downturn at the end of the sample period but that the downturns which showed up in Figure 5.1 for quarters 4 and 8 can be interpreted as seasonal downturns; i.e., the rolling year indexes in Figure 5.2 did not turn down until the end of the sample period. Note further that the index value for observation 5 compares the data for calendar year 2006 to the corresponding data for calendar year 2005 and the index value for observation 9 compares the data for calendar year 2007 to the corresponding data for calendar year 2005; i.e., these index values correspond to Mudgett-Stone annual indexes.

It is a fairly labour intensive job to construct the rolling year matched model Fisher indexes because the cells that are matched over any two periods vary with the periods. A short-cut method (which is less accurate) for seasonally adjusting a series, such as the matched model chained Fisher index P_{FCH} and the fixed base Fisher index P_{FFB} listed in Table 5.2 in section 5.5, is to simply take a 4 quarter *moving average* of these series. The resulting rolling year series, P_{FCHMA} and P_{FFBMA}, can be compared with the rolling year Mudgett-Stone-Diewert series P_{RY}; see Figure 5.2. The data that corresponds to Figure 5.2 are listed in Table 3.

[17] There are 11 rolling year comparisons that can be made with the data for 14 quarters that are available. The numbers of unmatched or empty cells for rolling years 2, 3, ..., 11 are as follows: 50, 52, 55, 59, 60, 61, 65, 65, 66, 67. The relatively low number of unmatched or empty cells for rolling years 2, 3 and 4 is due to the fact that for rolling year 2, ¾ of the data are matched, for rolling year 3, ½ of the data are matched and for rolling year 4, ¼ of the data are matched.

Figure 5.2. Rolling Year Fixed Base Fisher, Fisher Chained Moving Average and Fisher Fixed Base Moving Average Price Indexes

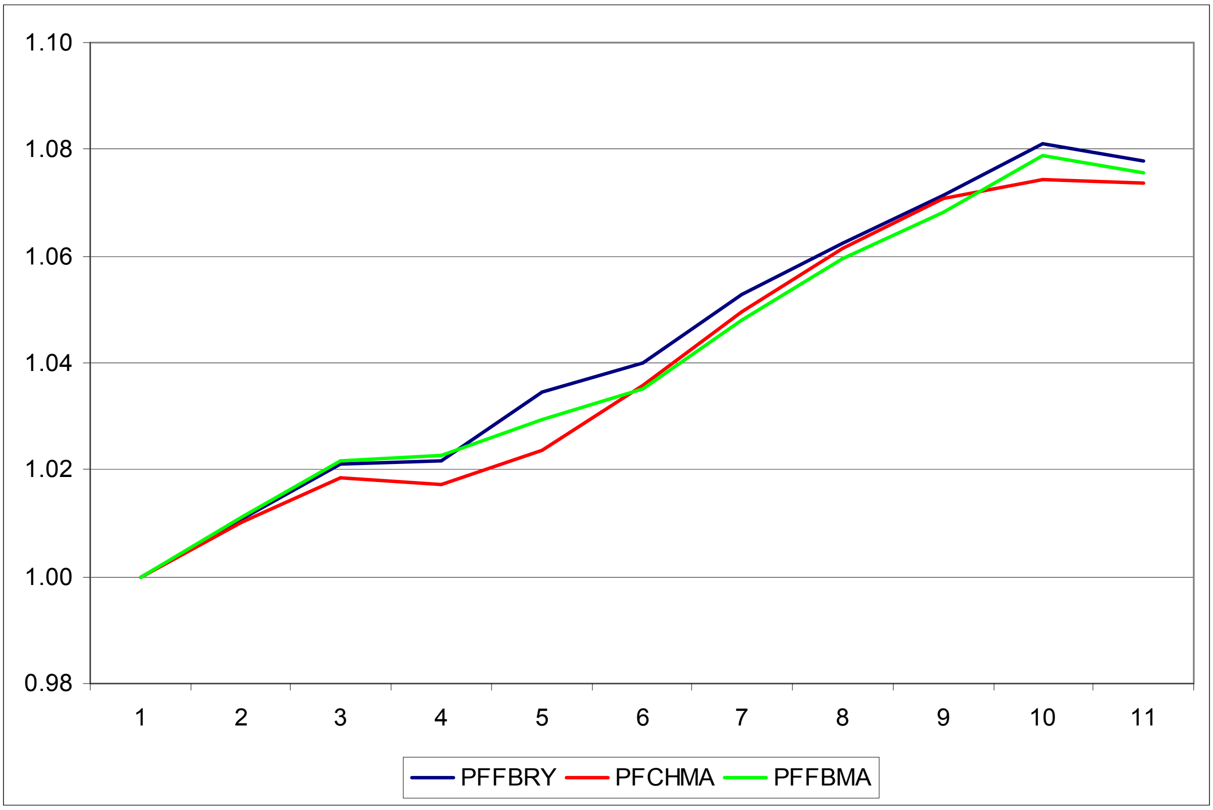


Table 5.3. Rolling Year Fixed Base Fisher, Fisher Chained Moving Average and Fisher Fixed Base Moving Average Price Indexes

Rolling Year	P_{FFBRY}	P_{FCHMA}	P_{FFBMA}
1	1.00000	1.00000	1.00000
2	1.01078	1.01021	1.01111
3	1.02111	1.01841	1.02156
4	1.02185	1.01725	1.02272
5	1.03453	1.02355	1.02936
6	1.04008	1.03572	1.03532
7	1.05287	1.04969	1.04805
8	1.06245	1.06159	1.05948
9	1.07135	1.07066	1.06815
10	1.08092	1.07441	1.07877
11	1.07774	1.07371	1.07556

It can be seen that a moving average of the chained and fixed base Fisher quarter to quarter indexes, P_{FCH} and P_{FFB}, listed in Table 5.2, approximates the theoretically preferred rolling year fixed base Fisher index P_{FFBRY} fairly well. There are differences of up to 1% between the preferred rolling year index and the moving average index, however. Recall that the fixed base Fisher index constructed in section 5.5 compared the data of quarters 1 to 14 with the corresponding data of quarter 1. Thus the observations for, say, quarters 2 and 1, 3 and 1, and 4 and 1 are not as likely to be as comparable as the rolling year indexes where data in any one quarter is always lined up with the data in the corresponding quarter of the base year. A similar argument applies to the moving average index P_{FCHMA}; the comparisons that go into the links in this index are from quarter to quarter and they are unlikely to be as accurate as comparisons across the years for the same quarter.[18]

[18] The stronger is the seasonality, the stronger will be this argument in favour of the accuracy of the rolling year index. The strength of this argument can be seen if all house price sales for each cell turn out to be *strongly seasonal*; i.e., the sales for any given cell occur in only one quarter in each year. Quarter to quarter comparisons are obviously impossible in this situation but rolling year indexes will be perfectly well defined.

Chapter 6: Hedonic Regression Methods

6.1 Hedonic Modeling and Estimation

The hedonic method recognizes that heterogeneous goods can be described by their attributes or characteristics. That is, a good is essentially a bundle of (performance) characteristics.[1] In the housing context, this bundle may contain attributes of both the structure and the location of the properties. There is no market for characteristics, since they cannot be sold separately, so the prices of the characteristics are not independently observed. The demand and supply for the properties implicitly determine the characteristics' marginal contributions to the prices of the properties. Regression techniques can be used to estimate those marginal contributions or shadow prices. One purpose of the hedonic method might be to obtain estimates of the willingness to pay for, or marginal cost of producing, the characteristics. Here we focus on the second main purpose, the construction of quality-adjusted price indexes.

Hedonic Modeling

The starting point is the assumption that the price p_n^t of property *n* in period *t* is a function of a fixed number, say *K*, characteristics measured by "quantities" z_{nk}^t. With *T*+1 time periods, going from the base period 0 to period *T*, we have

$$p_n^t = f(z_{n1}^t,...,z_{nK}^t,\varepsilon_n^t); \qquad (t = 0,...,T), \qquad (6.1)$$

where ε_n^t is a random error term (white noise). In order to be able to estimate the marginal contributions of the characteristics using standard regression techniques, equation (6.1) has to be specified as a parametric model. The two best-known hedonic specifications are the fully linear model

$$p_n^t = \beta_0^t + \sum_{k=1}^{K} \beta_k^t z_{nk}^t + \varepsilon_n^t, \qquad (6.2)$$

and the logarithmic-linear model

$$\ln p_n^t = \beta_0^t + \sum_{k=1}^{K} \beta_k^t z_{nk}^t + \varepsilon_n^t, \qquad (6.3)$$

where β_0^t and β_k^t are the intercept term and the characteristics parameters to be estimated. In both specifications the characteristics may be transformations, like logarithms, of continuous

[1] The hedonic regression approach dates back at least to Court (1939) and Griliches (1961). Lancaster (1966) and Rosen (1974) laid down the conceptual foundations of the approach. Colwell and Dilmore (1999) argue that the first published hedonic study was a 1922 University of Minnesota master's thesis on agricultural land values.

variables. In practice, many explanatory variables will be categorical rather than continuous and represented by a set of dummy variables which take the value of 1 if a property belongs to the category in question and the value of 0 otherwise.

For products such as high-tech goods, the log-linear model (6.3) is usually preferred, among other things because it most likely reduces the problem of heteroskedasticity (non-constant variance of the errors) as prices tend to be log-normally distributed (Diewert, 2003b). In the housing context, on the other hand, the linear model has much to recommend. In Chapter 4, the size of the structure and the size of the land it is built on were mentioned as two important price determining variables. Since the value of a property is generally equal to the *sum* of the price the structure and the price of land, it can be argued that land and structures should be included in the model in a linear fashion, provided that the data are available. Chapter 9 will discuss this issue in more detail, including a decomposition of the hedonic price index into land and structures components. Unfortunately, not all data sets will contain data on land and structure size. Information on land size in particular may be lacking. When land (or structure) size is not included as an explanatory variable, many empirical studies have found log-linear models to perform reasonably well.

The characteristics parameters β_k^t in (6.2) and (6.3) are allowed to change over time. This is in line with the idea that housing market conditions determine the marginal contributions of the characteristics: when demand and supply conditions change, there is no a priori reason to expect that those contributions are constant (Pakes, 2003). Yet, it seems likely that market conditions change gradually. So we could make the simplifying assumption, perhaps only for the short term, that the characteristics parameters (but not the intercept term) are constant across time. In the log-linear case this would give rise to the following constrained version of (6.3):

$$\ln p_n^t = \beta_0^t + \sum_{k=1}^{K} \beta_k z_{nk}^t + \varepsilon_n^t . \tag{6.4}$$

As will be seen below, the time dependent intercept terms (the β_0^t) can be converted into a constant quality price index.

Suppose we have data on selling prices and characteristics for the samples $S(0), S(1),..., S(T)$ of properties sold in periods $t = 0,...,T$ with sizes $N(0), N(1),..., N(T)$. Under the classic error assumptions, in particular a zero mean and constant variance, the parameters of the hedonic models (6.2) and (6.3) can be estimated by Ordinary Least Squares (OLS) regression on the sample data of each time period separately. The constrained version (6.4) can be estimated on the pooled data pertaining to all time periods, provided that dummy variables are included which indicate the time periods (leaving out one dummy to prevent perfect collinearity). The estimating equation for the constrained log-linear model (6.4), which is generally referred to as the *time dummy variable hedonic model*, thus becomes

$$\ln p_n^t = \beta_0 + \sum_{\tau=1}^{T} \delta^\tau D_n^\tau + \sum_{k=1}^{K} \beta_k z_{nk}^t + \varepsilon_n^t, \qquad (6.5)$$

where the time dummy variable D_n^τ has the value 1 if the observation comes from period τ and 0 otherwise; the time dummy for the base period 0 is left out. The reason for choosing a log-linear specification will become clear in section 6.2. Although unusual, it is also possible to specify a time dummy model with the untransformed price as the dependent variable. This specification will be considered in section 6.7.

Some Practical Issues

An important issue is the choice of the set of explanatory variables included in the hedonic equation. If some relevant variables – characteristics that can be expected to affect the price of a property (listed in Chapter 4) – are excluded, then the estimated parameters of the included characteristics will suffer from omitted variables bias. The bias carries over to the predicted prices computed from the regression coefficients and to the hedonic indexes. Each property can be viewed as a unique good, for a large part due to its location. But detailed information on location and neighbourhood can be hard to obtain (Case, Pollakowski, and Wachter, 1991). Other characteristics may be unavailable also and some could be difficult to measure directly. So it is fair to say that in practice some omitted variables bias will always be present when estimating a hedonic model for housing.[2] The sign and magnitude of the bias, and its impact on the price index, are difficult to predict. The magnitude depends among other things on the correlation between the omitted and included variables.

The importance of location has led researchers to make use of 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 appropriate (though rather specialized) econometric methods to allow for spatial dependence in the estimated equation. Explicitly accounting for spatial dependence can ameliorate the omitted locational variables problem. Spatial dependence can be captured either in the regressors or the error term. The first approach, i.e., including location as an explanatory variable using geospatial data, is the most straightforward one. This can be done parametrically or nonparametrically, for example by making use of splines, as demonstrated by Hill, Melser and Reid (2010). For an elaborate discussion and a review of the literature on spatial dependence, the use of geospatial data and also on nonparametric estimation, we refer the reader to Hill (2011).[3]

---

[2] A related point is that the characteristics of each house in the sample should be available in real time. House characteristics can change over time (which is actually the reason why they are given a superscript for time *t* in the hedonic models above). Keeping the characteristics fixed implies that the hedonic price index would not be adjusted for such quality changes.

[3] Colwell (1998) proposed a nonparametric spatial interpolation method which seems well adapted to model land prices as a function of the property's geographical two-dimensional coordinates.

Multicollinearity is a well-known problem in hedonic regressions. A high correlation between some of the included variables increases the standard errors of the regression coefficients; the coefficients become unstable. Again, it is difficult to say a priori how this will affect hedonic indexes. For some purposes, multicollinearity may not be too problematic. For example, if we are not so much interested in the values of the parameters but merely in the predicted prices to be used in the estimation of the overall quality-adjusted house price index, then the problem of multicollinearity should not be exaggerated. In this case it is better to include a relevant variable, even if this would cause multicollinearity, than leaving it out as the latter gives rise to omitted variables bias. But when the parameter values are of interest as such, for example when we are trying to decompose the property prices into land and structures components, then multicollinearity does pose problems. In Chapter 9 it will be shown that this is indeed a problem.

As with other methods, some data cleaning might be necessary. Obvious entry errors should be deleted. Yet, a cautious approach is called for. Deleting outliers from a regression with the aim of producing more stable coefficients (hence, more stable price indexes) is often arbitrary and could lead to biased estimates. The use of hedonics requires data on all characteristics included in the model. Unfortunately, partial non-response is present in many data sets. That is, the information on one or more characteristics may be missing for a part of the sample. Procedures have been developed to impute the missing data, but again it is important to avoid arbitrary choices that can have an impact on the results.

In sections 6.2 and 6.3 the two main hedonic approaches, the time dummy approach and the imputations approach, to constructing quality-adjusted house price indexes will be discussed. Without denying potential econometric problems, our focus will be on the use of least squares regression to estimate the models.

6.2 Time Dummy Variable Method

The time dummy variable approach to constructing a hedonic house price index has been used frequently in academic studies but not so much by statistical agencies.[4] One advantage of this approach is its simplicity; the price index follows immediately from the estimated pooled time dummy regression equation (6.5). Running one overall regression on the pooled data of the samples $S(0), S(1),..., S(T)$ relating to periods $t = 0,...,T$ (with sizes $n(0), n(1),..., n(T)$) yields coefficients $\hat{\beta}^0$, $\hat{\delta}^t$ $(t = 1,...,T)$ and $\hat{\beta}_k$ $(k = 1,..., K)$. The time dummy parameter shifts the

[4] This method was originally developed by Court (1939; 109-111) as his hedonic suggestion number two. The terminology adopted by us is not uniformly employed in the real estate literature. For example, Crone and Voith (1992) refer to the time dummy method as the "constrained hedonic" method, Gatzlaff and Ling (1994) call it the "explicit time-variable" method, while Knight, Dombrow and Sirmans (1995) name it the "varying parameter" method. Other terms also appear in the literature so that statements about the relative merits of different hedonic methods require careful interpretation.

hedonic surface upwards or downwards and measures the effect of "time" on the logarithm of price. Exponentiating the time dummy coefficients thus controls for changes in the quantities of the characteristics and provides a measure of quality-adjusted house price change between the base period 0 and each comparison period *t*. In other words, the time dummy index going from period 0 to period *t* is given by[5]

$$P_{TD}^{0t} = \exp(\hat{\delta}^t) . \tag{6.6}$$

Pooling cross-section data preserves degrees of freedom. The regression coefficients $\hat{\beta}_k$ will therefore generally have lower standard errors than the coefficients $\hat{\beta}_k^t$ that would be obtained by estimating equation (6.19) on the data of the samples $S(0), S(1), ..., S(T)$ separately. While the increased efficiency can be seen as an advantage, it comes at an expense: the assumption of fixed characteristics parameters is a disadvantage of the time dummy method.

When using OLS, the time dummy hedonic index can be written as (see e.g. Diewert, Heravi and Silver, 2009; de Haan, 2010a)

$$P_{TD}^{0t} = \frac{\prod_{n \in S(t)} (p_n^t)^{1/N(t)}}{\prod_{n \in S(0)} (p_n^0)^{1/N(0)}} \exp\left[\sum_{k=1}^{K} \hat{\beta}_k (\bar{z}_k^0 - \bar{z}_k^t)\right], \tag{6.7}$$

where $\bar{z}_k^s = \sum_{n \in S(s)} z_{nk}^s / N(s)$ is the sample mean of characteristic *k* in period *s* $(s = 0, t)$. Equation (6.7) tells us that the time dummy index is essentially the product of two factors. The first factor is the ratio of the geometric mean prices in the periods *t* and 0. The second factor, $\exp[\sum_{k=1}^{K} \hat{\beta}_k (\bar{z}_k^0 - \bar{z}_k^t)]$, adjusts this ratio of raw sample means for differences in the average characteristics $\bar{z}_k^0$ and $\bar{z}_k^t$; it serves as a quality-adjustment factor which accounts for both changes in the quality mix and quality changes of the individual properties (provided that all relevant quality-determining attributes are included in the hedonic model). Notice that the time dummy price index simplifies to the ratio of geometric mean prices if $\bar{z}_k^t = \bar{z}_k^0$, i.e. if the average characteristics in period *t* and period 0 happen to be equal.

Suppose for simplicity that the housing stock is constant, in the sense that there are no houses entering or exiting, and that the quality of the individual properties does not change. Suppose further that $S(0)$ and $S(t)$ are random or "representative" selections from the housing stock. In that case the time dummy method implicitly aims at a ratio of geometric mean prices for the total stock, which is equal to the geometric mean of the individual price ratios.[6] Although it is true that the target of measurement may be different for different purposes, it is difficult

[5] The expected value of the exponential of the time dummy coefficient is not exactly equal to the exponential of the time dummy parameter. The associated bias is often referred to as small sample bias: it diminishes when the sample size grows. Unless the sample size is extraordinary small, the bias will be small compared to the standard error and can usually be neglected in practice.

[6] In index number theory such an index is referred to as a Jevons index.

to see what purposes a geometric stock RPPI would meet. Arithmetic target RPPIs, such as an index that tracks the value of the fixed housing stock over time, seem to be more appropriate (see also Chapters 5 and 9).

The samples of houses traded, $S(0)$ and $S(t)$, may not be representative for the total housing stock (or for the total population of houses sold). A solution could be to weight the samples in order to make them representative. Running an OLS regression on the (pooled) weighted data set is equivalent to running a Weighted Least Squares (WLS) regression on the original data set. Under the assumption of a constant variance of the errors, econometric textbooks do not suggest the use of WLS since this will introduce heteroskedasticity. Note that a WLS time dummy method will still generate a geometric index, in this case a weighted one.

A better option than using WLS regressions could be to stratify the samples, run separate OLS regressions on the data of the different strata, and then explicitly weight the stratum-specific hedonic indexes using stock (or sales) weights to construct an overall RPPI with an arithmetic structure at the upper level of aggregation. This stratified hedonic approach has several other advantages as well, as will be explained in section 6.4.

A problem with the time dummy method is the revision that goes with it. If the time series is extended to $T+1$ and new sample data is added, the characteristics coefficients will change. Consequently, the newly computed price index numbers for the periods $t=1,...,T$ will differ from those previously computed.[7] When additional data become available, the efficiency due to the pooling of data increases and better estimates can be made. This can actually be seen as a strength rather than a weakness of the method. On the other hand, statistical agencies and their users will most likely be reluctant to accept continuous revisions of previously published figures.

The multiperiod time dummy method therefore appears to be of limited use for the production of official house price indexes although there are ways to deal with the problem of revisions. One way would be to estimate time dummy indexes for adjacent periods *t*-1 and *t* and then multiply them to obtain a time series which is free of revisions. This high-frequency chaining has the additional advantage of relaxing the assumption of fixed parameters. It is, however, not entirely without problems. Drift in the index can occur when the data exhibit systematic fluctuations such as seasonal fluctuations.[8]

[7] In the words of Hill (2004), the time dummy approach violates time fixity.

[8] An alternative approach would be the use of a moving window. For example, suppose we initially estimated a time dummy index on the data of twelve months. Next, we delete the data of the first month and add the data of the thirteenth month and estimate a time dummy index on this data set, and so on. By multiplying (chaining) the last month-to-month changes a non-revised time series is obtained. For an application, see Shimizu, Nishimura and Watanabe (2010). In section 9.7 an example is given for the town of "A" where drift does not seem to be a problem; the moving window method gives much the same results as the multiperiod time dummy regression.

6.3 Characteristics Prices and Imputation Methods

In the second main approach to compiling a hedonic price index, separate regressions are run for all time periods and the index is constructed by making use of the predicted prices based on the regression coefficients. Because the implicit characteristics prices are allowed to vary over time, this method is more flexible than the time dummy variable method. Two variants can be distinguished: the *characteristics prices approach* and the *imputations approach*. It will be shown that, under certain circumstances, both approaches are equivalent. We will first discuss the characteristics prices approach.[9]

6.3.1 Characteristics Prices Approach

To illustrate this approach, suppose as before that sample data are available on prices and relevant characteristics of houses sold in the base period 0 and each comparison period *t*. We will first assume that the linear hedonic model (6.2) holds true and is estimated on the data of period 0 and period *t* separately. This yields regression coefficients $\hat{\beta}_0^s$ and $\hat{\beta}_k^s$ $(k = 1,...,K)$ for $s = 0,t$. The predicted prices for each individual property are $\hat{p}_n^0 = \hat{\beta}_0^0 + \sum_{k=1}^{K} \hat{\beta}_k^0 z_{nk}^0$ and $\hat{p}_n^t = \hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t z_{nk}^t$. It is also possible to compute predicted period 0 and period *t* prices for a "standardized" property with fixed (quantities of) characteristics z_k^*. The resulting estimated price relative is

$$\frac{\hat{p}^t}{\hat{p}^0} = \frac{\hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t z_k^*}{\hat{\beta}_0^0 + \sum_{k=1}^{K} \hat{\beta}_k^0 z_k^*}. \tag{6.8}$$

Expression (6.8) is a quality-adjusted price index because the characteristics are kept fixed. But different values of z_k^* will give rise to different index numbers. So what would be the preferred choice?

Suppose that we were aiming at a sales-based RPPI. There are two natural choices for z_k^* in (6.8): the sample average characteristics of the base period, $\bar{z}_k^0$, and the sample averages of the comparison period t $(t = 1,...,T)$, $\bar{z}_k^t$. The usual solution in index number theory is to treat the resulting price indexes – which are equally valid – in a symmetric manner by taking the geometric mean. Setting $z_k^* = \bar{z}_k^0$ in (6.8) generates a Laspeyres-type characteristics prices (CP) index:

[9] Again, the terminology differs between authors. For example, Crone and Voith (1992) and Knight, Dombrow and Sirmans (1995) refer to this approach as the "hedonic method" (as opposed to the "constrained hedonic" or "varying parameter" method, what we have called the time dummy variable approach), while Gatzlaff and Ling (1994) refer to it as the "strictly cross-sectional" method.

$$P_{CPL}^{0t} = \frac{\hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t \bar{z}_k^0}{\hat{\beta}_0^0 + \sum_{k=1}^{K} \hat{\beta}_k^0 \bar{z}_k^0}. \tag{6.9}$$

Setting $z_k^* = \bar{z}_k^t$ in (6.8) yields a Paasche-type index:

$$P_{CPP}^{0t} = \frac{\hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t \bar{z}_k^t}{\hat{\beta}_0^0 + \sum_{k=1}^{K} \hat{\beta}_k^0 \bar{z}_k^t}. \tag{6.10}$$

By taking the geometric mean of (6.9) and (6.10) the Fisher-type characteristics prices index is obtained:

$$P_{CPF}^{0t} = \left[P_{CPL}^{0t} P_{CPP}^{0t}\right]^{1/2}. \tag{6.11}$$

The characteristics prices method can also applied in combination with the log-linear model given by (6.3). Running separate regressions of this model on the sample data for periods 0 and t gives rise to predicted prices (after exponentiating) $\hat{p}_n^0 = \exp(\hat{\beta}_0^0)\exp[\sum_{k=1}^{K} \hat{\beta}_k^0 z_{nk}^0]$ and $\hat{p}_n^t = \exp(\hat{\beta}_0^t)\exp[\sum_{k=1}^{K} \hat{\beta}_k^t z_{nk}^t]$. Similar to what was done in (6.8) for the linear model, prices can be predicted for a standardized house. Using the sample averages of the characteristics in the base period to define the standardized house, the geometric counterpart to the Laspeyres-type characteristics prices index (6.9) is found:

$$P_{CPGL}^{0t} = \frac{\exp(\hat{\beta}_0^t)\exp\left[\sum_{k=1}^{K} \hat{\beta}_k^t \bar{z}_k^0\right]}{\exp(\hat{\beta}_0^0)\exp\left[\sum_{k=1}^{K} \hat{\beta}_k^0 \bar{z}_k^0\right]} = \exp(\hat{\beta}_0^t - \hat{\beta}_0^0)\exp\left[\sum_{k=1}^{K} (\hat{\beta}_k^t - \hat{\beta}_k^0)\bar{z}_k^0\right]. \tag{6.12}$$

The geometric counterpart to the Paasche-type hedonic index (6.10) is obtained by using the sample averages of the characteristics in the comparison period:

$$P_{CPGP}^{0t} = \frac{\exp(\hat{\beta}_0^t)\exp\left[\sum_{k=1}^{K} \hat{\beta}_k^t \bar{z}_k^t\right]}{\exp(\hat{\beta}_0^0)\exp\left[\sum_{k=1}^{K} \hat{\beta}_k^0 \bar{z}_k^t\right]} = \exp(\hat{\beta}_0^t - \hat{\beta}_0^0)\exp\left[\sum_{k=1}^{K} (\hat{\beta}_k^t - \hat{\beta}_k^0)\bar{z}_k^t\right]. \tag{6.13}$$

Taking the geometric mean of (6.12) and (6.13) yields

$$P_{CPGF}^{0t} = \left[P_{HGL}^{0t} P_{HGP}^{0t}\right]^{1/2} = \exp(\hat{\beta}_0^t - \hat{\beta}_0^0)\exp\left[\sum_{k=1}^{K} (\hat{\beta}_k^t - \hat{\beta}_k^0)\bar{z}_k^{0t}\right], \tag{6.14}$$

where $\bar{z}_k^{0t} = (\bar{z}_k^0 + \bar{z}_k^t)/2$ in (6.14) denotes the mean of the average characteristics in the base and comparison period.

If the target index is a stock-based rather than a sales-based RPPI, the two natural choices for the characteristics z_k^* in equation (6.8) would be the average *stock* characteristics of the base period and those of the comparison period. The first choice would produce a Laspeyres-type stock RPPI, the second choice a Paasche-type stock RPPI. Both indexes measure the quality-adjusted value change of the housing stock, but the results will usually differ. Not only does the average quality of the housing stock change over time, the Laspeyres-type index ignores new properties that entered the housing market whereas the Paasche-type index does not take into account disappearing properties.

Of course the assumption of known stock averages for all property characteristics included in the hedonic model is unrealistic. In most situations we have to rely on estimates, i.e. on the sample averages $\bar{z}_k^0$ and $\bar{z}_k^t$ which are based on the same characteristics data that is used to estimate the hedonic equations. This leads to formulae (6.9) and (6.10), or the geometric mean (6.11), which describe sales-based RPPIs. Once again we are reminded that sales RPPIs can be seen as estimators of stock RPPIs, provided that the samples are representative for the total stock. The latter is rather doubtful, however, and the usual approach is to stratify the samples and weight the results using stock weights; see further Section 6.4.

6.3.2 Hedonic Imputation Approach

The question arises how the characteristics prices method described in section 6.3.1 relates to the standard (matched-model) methodology to construct price indexes. From an index number point of view we can look at the issue in the following way. The period t prices of properties sold in period 0 cannot be observed and are "missing" because those properties, or at least the greater part, will not be resold in period t. Similarly, the period 0 prices of the properties sold in period t are unobservable. To apply standard index number formulae these "missing prices" must be imputed.[10] Hedonic imputation indexes do this by using predicted prices, evaluated at fixed characteristics, based on the hedonic regressions for all time periods.

Arithmetic Imputation Indexes

The Laspeyres hedonic imputation index imputes period t prices for the properties belonging to the base period sample $S(0)$, evaluated at base period characteristics to control for quality

[10] As noted earlier, the hedonic theory dates back at least to Court (1939; 108). Imputation was his hedonic suggestion number one. His suggestion was followed up by Griliches (1971a; 59-60) (1971b; 6) and Triplett and McDonald (1977; 144). More recent contributions to the hedonic imputations literature include Diewert (2003b), de Haan (2004) (2009) (2010a), Triplett (2004) and Diewert, Heravi and Silver (2009). In a housing context the hedonic imputation method is discussed in detail by Hill and Melser (2008) and Hill (2011).

changes. Thus, using the linear model (6.1), the imputed prices are $\hat{p}_n^t(0) = \hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t z_{nk}^0$ so that the hedonic imputation Laspeyres index becomes

$$P_{HIL}^{0t} = \frac{\sum_{n \in S(0)} 1 \hat{p}_n^t(0)}{\sum_{n \in S(0)} 1 p_n^0} = \frac{\sum_{n \in S(0)} \left[\hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t z_{nk}^0 \right]}{\sum_{n \in S(0)} p_n^0} = \frac{\hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t \bar{z}_k^0}{\sum_{n \in S(0)} p_n^0 / N(0)}. \tag{6.15}$$

Notice that the quantity associated with each price is 1; basically, every house is unique and cannot be matched except through the use of a model.

The hedonic imputation Laspeyres index (6.15) is an example of a *single imputation* index in which the observed prices are left unchanged. It can be argued that it would be better to use a *double imputation* approach, where the observed prices are replaced by the predicted values. This is because biases in the period 0 and period t estimates due to omitted variables likely offset each other, at least to some degree (see e.g. Hill, 2011). Using $\hat{p}_n^0 = \hat{\beta}_0^0 + \sum_{k=1}^{K} \hat{\beta}_k^0 z_{nk}^0$, the hedonic double imputation (DI) Laspeyres price index is

$$P_{HDIL}^{0t} = \frac{\sum_{n \in S(0)} 1 \hat{p}_n^t(0)}{\sum_{n \in S(0)} 1 \hat{p}_n^0} = \frac{\sum_{n \in S(0)} \left[\hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t z_{nk}^0 \right]}{\sum_{n \in S(0)} \left[\hat{\beta}_0^0 + \sum_{k=1}^{K} \hat{\beta}_k^0 z_{nk}^0 \right]} = \frac{\hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t \bar{z}_k^0}{\hat{\beta}_0^0 + \sum_{k=1}^{K} \hat{\beta}_k^0 \bar{z}_k^0} = P_{CPL}^{0t}. \tag{6.16}$$

A comparison with equation (6.12) shows that, using the linear model, the double imputation index equals the Laspeyres-type characteristics prices index. This result does not depend on the estimation method. If we would use OLS regression to estimate the linear model, then the single imputation index would be equal to the double imputation index and also coincide with the characteristics prices index as in this case $\sum_{n \in S(0)} p_n^0 = \sum_{n \in S(0)} \hat{p}_n^0$, due to the fact that the hedonic model includes an intercept term so that the OLS regression residuals sum to zero.

The hedonic single imputation Paasche index imputes base period prices for the properties belonging to the period t sample $S(t)$, evaluated at period t characteristics. Using again the linear model (6.1), these imputed prices are given by $\hat{p}_n^0(t) = \hat{\beta}_0^0 + \sum_{k=1}^{K} \hat{\beta}_k^0 z_{nk}^t$. To save space we will only show the double imputation variant. Here, the observed (period t) prices are replaced by their model-based predictions $\hat{p}_n^t = \hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t z_{nk}^t$. Thus, the hedonic double imputation Paasche price index is

$$P_{HDIP}^{0t} = \frac{\sum_{n \in S(t)} 1 \hat{p}_n^t}{\sum_{n \in S(t)} 1 \hat{p}_n^0(t)} = \frac{\sum_{n \in S(t)} \left[\hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t z_{nk}^t \right]}{\sum_{n \in S(t)} \left[\hat{\beta}_0^0 + \sum_{k=1}^{K} \hat{\beta}_k^0 z_{nk}^t \right]} = \frac{\hat{\beta}_0^t + \sum_{k=1}^{K} \hat{\beta}_k^t \bar{z}_k^t}{\hat{\beta}_0^0 + \sum_{k=1}^{K} \hat{\beta}_k^0 \bar{z}_k^t} = P_{CPP}^{0t}, \tag{6.17}$$

which coincides with the Paasche-type characteristics prices index. If OLS regression is used, then (6.17) is equal to the single imputation Paasche index because in this particular case the numerator equals $\sum_{n \in S(t)} p_n^t$. It will then be unnecessary to estimate the hedonic equations for the comparison periods $t = 1,...,T$; estimating the base period hedonic equation to obtain the base period imputed values will suffice.

The hedonic double imputation Fisher index is found by taking the geometric mean of (6.16) and (6.17):

$$P_{HDIF}^{0t} = \left[P_{HDIL}^{0t} P_{HDIP}^{0t}\right]^{1/2}. \tag{6.18}$$

The above imputation indexes can be given two interpretations. They can be viewed either as estimators of the quality-adjusted value change of the entire housing stock, i.e., as stock-based RPPIs, or as estimators of quality-adjusted sales-based RPPIs. Under the first interpretation, to produce approximately unbiased results, each sample should be a random or representative selection from the housing stock. Sample selection bias problems could be less severe under the second interpretation, although this depends on the sampling design.[11]

Geometric Imputation Indexes

The imputation approach can also be applied to geometric price index number formulae. Let us start with what might be called the geometric counterpart to the imputation Laspeyres price index (6.15). For reasons of "consistency" the imputations will now be computed using the log-linear hedonic model (6.3) instead of the linear model. The imputed period *t* prices for the properties belonging to the base period sample $S(0)$, evaluated at base period characteristics, are $\hat{p}_n^t(0) = \exp(\hat{\beta}_0^t)\exp[\sum_{k=1}^{K} \hat{\beta}_k^t z_{nk}^0]$. The double imputation unweighted geometric index, in which the base period prices are replaced by predicted values $\hat{p}_n^0 = \exp(\hat{\beta}_0^0)\exp[\sum_{k=1}^{K} \hat{\beta}_k^0 z_{nk}^0]$, is

$$P_{HDIGL}^{0t} = \frac{\prod_{n \in S(0)} (\hat{p}_n^t(0))^{1/N(0)}}{\prod_{n \in S(0)} (\hat{p}_n^0)^{1/N(0)}} = \exp(\hat{\beta}_0^t - \hat{\beta}_0^0)\exp\left[\sum_{k=1}^{K} (\hat{\beta}_k^t - \hat{\beta}_k^0)\bar{z}_k^0\right] = P_{CPGL}^{0t}. \tag{6.19}$$

Similarly, the geometric counterpart to the imputation Paasche price index (6.16) is obtained by imputing period 0 prices for the properties belonging to the period *t* sample $S(t)$, which are given by $\hat{p}_n^0(t) = \exp(\hat{\beta}_0^0)\exp[\sum_{k=1}^{K} \hat{\beta}_k^0 z_{nk}^t]$, and replacing the observed period *t* prices by the predictions $\hat{p}_n^t = \exp(\hat{\beta}_0^t)\exp[\sum_{k=1}^{K} \hat{\beta}_k^t z_{nk}^t]$. So we have

[11] If all property transactions are observed, there is no sampling involved from a sales point of view, and sample selection bias is not an issue. In many countries the Land Registry records all transactions, at least for resold houses. However, such data sets usually have limited information on characteristics; see e.g. Lim and Pavlou (2007) or Academetrics (2009).

$$P_{HDIGP}^{0t} = \frac{\prod_{n \in S(t)} (\hat{p}_n^t)^{1/N(t)}}{\prod_{n \in S(t)} (\hat{p}_n^0(t))^{1/N(t)}} = \exp(\hat{\beta}_0^t - \hat{\beta}_0^0) \exp\left[\sum_{k=1}^{K} (\hat{\beta}_k^t - \hat{\beta}_k^0) \bar{z}_k^t\right] = P_{CPGP}^{0t}. \tag{6.20}$$

When OLS is used to estimate the log-linear regression equations, the denominator of (6.19) and the numerator of (6.20) will equal the geometric sample means of the prices in period 0 and period *t*, respectively, and the double imputation indexes coincide with single imputation indexes.

Taking the geometric mean of (6.19) and (6.20) yields

$$P_{HDIGF}^{0t} = \left[P_{HDIGL}^{0t} P_{HDIGP}^{0t}\right]^{1/2} = \exp(\hat{\beta}_0^t - \hat{\beta}_0^0) \exp\left[\sum_{k=1}^{K} (\hat{\beta}_k^t - \hat{\beta}_k^0) \bar{z}_k^{0t}\right] = P_{CPGF}^{0t}, \tag{6.21}$$

where $\bar{z}_k^{0t} = (\bar{z}_k^0 + \bar{z}_k^t)/2$ denotes the mean of the average characteristics in periods 0 and *t*, as before.

The symmetric imputation index equation (6.21) can be rewritten in a way that is surprisingly similar to equation (6.7) for the time dummy index when OLS is used to estimate the hedonic equations (see Diewert, Heravi and Silver, 2009, and de Haan, 2010a):

$$P_{HDIGF}^{0t} = \frac{\prod_{n \in S(t)} (p_n^t)^{1/N(t)}}{\prod_{n \in S(0)} (p_n^0)^{1/N(0)}} \exp\left[\sum_{k=1}^{K} \hat{\beta}_k^{0t} (\bar{z}_k^0 - \bar{z}_k^t)\right], \tag{6.22}$$

where $\hat{\beta}_k^{0t} = (\hat{\beta}_k^0 + \hat{\beta}_k^t)/2$ denotes the average value of the *k*-th coefficient in periods 0 and *t*. Equation (6.22) adjusts the ratio of observed geometric mean prices for any differences in the average sample characteristics. Triplett (2006) refers to this as "hedonic quality adjustment". A comparison with equation (6.7) shows that if the sample averages of all characteristics stay the same $(\bar{z}_k^0 = \bar{z}_k^t)$, then the symmetric hedonic imputation index and the time dummy index coincide and equal the ratio of observed geometric mean prices, but this will obviously, rarely happen. Both types of hedonic indexes also coincide if, for each characteristic, the average coefficient $\hat{\beta}_k^{0t}$ from the two separate regressions would be equal to the coefficient $\hat{\beta}_k$ from the time dummy regression. This is rare as well, but it suggests that both approaches generate similar results if the characteristics parameters are approximately constant over time.

If the characteristics parameters can be assumed constant over time, the average coefficients $\hat{\beta}_k^{0t}$ in equation (6.22) can be replaced by the base period coefficients $\hat{\beta}_k^0$. In that case there would be no need to run a regression in each time period, and we would in fact be using the non-symmetric imputation index given by equation (6.13).[12] Data permitting, the base period

[12] In Europe this type of hedonic quality adjustment is called "hedonic re-pricing", especially in case the sample size is fixed (Destatis, 2009).

regression could be run on a bigger data set to increase the stability of the coefficients. It is advisable to regularly check if the coefficients have significantly changed and to update them when necessary.

As mentioned earlier, geometric are less suitable as estimators of quality-adjusted RPPIs. This is not to say they should never be used. In conjunction with stratification, the use of (6.21) could produce satisfactory results since this would combine quality adjustment (using a log-linear hedonic regression model) and a symmetric index number formula within strata with mix adjustment across strata. The stratified hedonic approach will be discussed in section 6.4 below.

6.4 Stratified Hedonic Indexes

Chapter 5 dealt with stratification or mix adjustment. Stratification is a simple and powerful tool to adjust for changes in the quality mix of the properties sold. However, some quality mix changes within the strata are likely to remain, as essentially every property is a unique good, and some unit value bias could therefore occur. A more detailed stratification scheme may be infeasible, especially when the number of observations is relatively small. Provided that the necessary data on characteristics are available, it could be worthwhile to work with a less fine stratification scheme and use hedonic regression at the stratum level to adjust for quality mix changes. This two-stage approach combines hedonics at the lower (stratum) level and explicit weighting at the upper level to form an overall RPPI.

Two advantages of stratification have been mentioned earlier. First, stratification enables the statistical agency to publish different RPPIs for different market segments. Users will benefit from this because it is well known that different types of houses, different regions, etc. can exhibit quite different price trends. Second, stratification can be helpful for reducing sample selection bias, including bias due to non-response, in particular for a stock-based RPPI.

When using hedonic regression techniques to adjust for quality (mix) changes, stratification is highly recommended. It is very unlikely that a single hedonic model holds true for all market segments, hence separate regressions should be run for different types of properties, different locations, etc. There are in fact two issues involved. Perhaps the biggest issue is that different sets of property characteristics will be needed for different market segments. For example, the characteristics that are relevant for detached dwelling units differ from those that are relevant for high rise apartments, if only because the floor of the apartment seems an important price determining variable. The second, though probably less important, issue is that the parameter values for the same characteristics can differ across housing market segments. Statistical tests for differences in parameter values between sub-samples can be found in any econometrics textbook.

The stratified hedonic approach can be illustrated most easily with reference to the imputation method explained in section 6.3, especially in combination with the Laspeyres index formula. Recall the third expression on the right-hand side of the hedonic single imputation Laspeyres price index (6.15), where the period *t* prices for the houses in the base period sample $S(0)$ are "missing" and imputed (using the estimated hedonic regression model for period *t*) by $\hat{p}_n^t(0)$. Suppose, as in Chapter 5, that the total sample is (post) stratified into *M* sub-samples $S_m(0)$. Equation (6.15) can then be rewritten as

$$P_{HIL}^{0t} = \frac{\sum_{n \in S(0)} \hat{p}_n^t(0)}{\sum_{n \in S(0)} p_n^0} = \frac{\sum_{m=1}^{M} \sum_{n \in S_m(0)} \hat{p}_n^t(0)}{\sum_{m=1}^{M} \sum_{n \in S_m(0)} p_n^0} = \frac{\sum_{m=1}^{M} \sum_{n \in S_m(0)} p_n^0 \left[\sum_{n \in S_m(0)} \hat{p}_n^t(0) \Big/ \sum_{n \in S_m(0)} p_n^0 \right]}{\sum_{m=1}^{M} \sum_{n \in S_m(0)} p_n^0} = \sum_{m=1}^{M} s_m^0 P_{HIL,m}^{0t} , \quad (6.23)$$

where $P_{HIL,m}^{0t} = \sum_{n \in S_m(0)} \hat{p}_n^t(0) / \sum_{n \in S_m(0)} p_n^0$ denotes the hedonic (single) imputation Laspeyres price index between the base period and period *t* for cell *m*; $s_m^0 = \sum_{n \in S_m(0)} p_n^0 / \sum_{n \in S(0)} p_n^0$ is the corresponding sales value share, which serves as the weight for $P_{HIL,m}^{0t}$. Note that the last expression of (6.23) has a similar structure as the mix-adjusted index given by equation (5.1), but in the present case the cell indexes are hedonic imputation indexes rather than unit value indexes.

Equation (6.23) shows that if, as in section 6.3, the imputed prices $\hat{p}_n^t(0)$ for all houses in the sample $S(0)$ are based on one overall hedonic regression, the aggregate hedonic imputation Laspeyres index can be written in the form of a stratified index. But this is just another way of writing things, not what is meant by a stratified hedonic approach. Also, as argued above, the use of a common model is very unrealistic. So instead of running one big hedonic regression, separate regressions should be performed on the data of the sub-samples in each time period to obtain imputed (period *t*) prices and imputation cell indexes. That would lead to a stratified Laspeyres-type hedonic imputation index.

It would be preferable to estimate a stratified Fisher hedonic index rather than a Laspeyres one. This is perfectly feasible for a sales RPPI but may not be feasible for a stock RPPI, as was already mentioned in Chapter 4, since up-to-date census data on the number of properties is often lacking.

6.5 Main Advantages and Disadvantages

This section summarizes the advantages and disadvantages of hedonic regression methods to construct an RPPI. The main advantages are:

- If the list of available property characteristics is sufficiently detailed, hedonic methods can in principle adjust for both sample mix changes and quality changes of the individual properties.

- Price indexes can be constructed for different types of dwellings and locations through a proper stratification of the sample. Stratification has a number of other advantages as well.
- The hedonic method is probably the most efficient method for making use of the available data.
- The imputation variant of the hedonic regression method is analogous to the matched model methodology that is widely used in order to construct price indexes.

The main disadvantages of hedonic regression are:

- It may be difficult to control sufficiently for location if property prices and price trends differ across detailed regions. However, a stratified approach to hedonic regressions will help overcome this problem to some extent.
- The method is data intensive since it requires data on all relevant property characteristics, so it is relatively expensive to implement.[13]
- While the method is essentially reproducible, different choices can be made regarding the set of characteristics included in the model, the functional form, possible transformations of the dependent variable[14], the stochastic specification, etc., which could lead to varying estimates of overall price change. Thus, a lot of metadata may be required.
- The general idea of the hedonic method is easily understood but some of the technicalities may not be easy to explain to users.

The overall evaluation of the hedonic regression method is that it is probably the best method that could be used in order to construct constant quality RPPIs for various types of property.[15] We are in favor of the (double) imputation variant because this is the most flexible hedonic approach and because this approach is analogous to the standard matched-model methodology to construct price indexes.

In the following three sections, the various hedonic regression methods discussed above will be illustrated using the data for the town of "A" that was described at the end of Chapter 5. Sections 6.6 and 6.7 show the results of time dummy hedonic regressions, using the log of the selling price as the dependent variable and using the untransformed selling price, respectively. Section 6.8 illustrates the hedonic imputation method. All of the resulting price indexes are

[13] However, as will be seen from the Dutch example in Sections 6.6-6.8, just having information on location, type of property, its age, its floor space area and the plot area may explain most of the variation in the selling price.

[14] For example, the dependent variable could be the sales price of the property or its logarithm or the sales price divided by the area of the structure and so on.

[15] This evaluation agrees with that of Hoffmann and Lorenz (2006; 15): "As far as quality adjustment is concerned, the future will certainly belong to hedonic methods." Gouriéroux and Laferrère (2009) have shown that it is possible to construct an official nationwide credible hedonic regression model for real estate properties.

for the *sales* of detached houses; some results using the data for the town of "A" for indexes of the *stock* of houses will be postponed until Chapter 9.

6.6 Time Dummy Models Using the Logarithm of Price as the Dependent Variable

The Log Linear Time Dummy Model

Recall the description of the data for the Dutch town of "A" on sales of detached houses. In quarter t, there were $N(t)$ sales of detached houses in "A" where p_n^t is the selling price of house n sold during quarter t. There is information on three characteristics of house n sold in period t: L_n^t is the area of the plot in square meters (m^2); S_n^t is the floor space area of the structure in m^2 and A_n^t is the age in decades of house n in period t. Using these variables, the standard *log linear time dummy hedonic regression model* is defined by the following system of regression equations:[16]

$$\ln p_n^t = \alpha + \beta L_n^t + \gamma S_n^t + \delta A_n^t + \tau^t + \varepsilon_n^t, \qquad t = 1,\ldots,14;\ n = 1,\ldots,N(t);\ \tau^1 \equiv 0; \qquad (6.24)$$

where τ^t is a parameter which shifts the hedonic surface in quarter t upwards or downwards as compared to the surface in quarter 1.[17]

It is easy to construct a price index using model (6.24). Exponentiating both sides of (6.24) and neglecting the error term yields $p_n^t = \exp(\alpha)[\exp(L_n^t)]^\beta[\exp(S_n^t)]^\gamma[\exp(A_n^t)]^\delta \exp(\tau^t)$. If a property with the *same characteristics* could be observed in the base period 1 and in some comparison period $t(>1)$, then the corresponding price relative (again neglecting error terms) would simply be equal to $\exp(\tau^t)$. For two consecutive periods t and t+1, the price relative (again neglecting error terms) would equal $\exp(\tau^{t+1})/\exp(\tau^t)$ and this can serve as the chain link in a price index. Figure 6.1 shows the resulting index, labeled as P_{H1} (hedonic index no. 1), and Table 6.1 lists the index numbers. The R^2 for this model was .8420, which is quite satisfactory for a hedonic regression model with only three explanatory variables.[18] For later comparison purposes, note that the log likelihood was 1407.6.

---

[16] As was mentioned in section 6.1, the *estimating equation* for the pooled data set will include time dummy variables to indicate the quarters. For all the models estimated for the town of "A", it is assumed that the error terms ε_n^t are independently distributed normal variables with mean 0 and constant variance. Maximum likelihood estimation is used in order to estimate the unknown parameters in each regression model. The nonlinear option in Shazam was used for the actual estimation.

[17] The 15 parameters α, $\tau^1,\ldots,\tau^{14}$ correspond to variables that are exactly collinear in the regression (6.24) and thus the restriction $\tau^1 = 0$ is imposed in order to identify the remaining parameters.

[18] Later in this chapter and in chapter 9, some hedonic regressions will be run that use *prices* p_n^t as the dependent variables rather than the *logs* of the prices. To facilitate comparisons of goodness of fit across models, we will transform the predicted values for the log price models into predicted price levels by exponentiating the predicted prices and then calculating the correlation coefficient between these predicted price levels and the actual prices. Squaring this correlation coefficient gives us *a levels type measure of goodness of fit* for the log price models which is denoted by R^{*2}. For this particular model, R^{*2} .= .8061.

A problem with this model is that the underlying price formation model seems implausible: *S* and *L* interact multiplicatively in order to determine the overall house price whereas it seems most likely that lot size *L* and house size *S* interact in an approximately additive fashion to determine the overall house price.

Another problem with the regression model (6.24) is that age is entered in an additive fashion. The problem is that we would expect age to interact directly with the structures variable *S* as a (net) depreciation variable and not interact directly with the land variable *L*, because land does not depreciate. In the following model, this direct interaction of age with structures will be made.

The Log Linear Time Dummy Model with Quality Adjustment of Structures for Age

If age *A* interacts with the quantity of structures *S* in a multiplicative manner, an appropriate explanatory variable for the selling price of a house would be $\gamma(1-\delta)^{A}S$ (i.e., geometric depreciation where δ is the decade geometric depreciation rate) or $\gamma(1-\delta A)S$ (straight line depreciation where δ is the decade straight line depreciation rate) instead of the additive specification $\gamma S+\delta A$. In what follows, the straight line variant of this class of models will be estimated[19]. Thus, the *log linear time dummy hedonic regression model with quality adjusted structures* becomes

$$\ln p_n^t=\alpha+\beta L_n^t+\gamma(1-\delta A_n^t)S_n^t+\tau^t+\varepsilon_n^t, \qquad t=1,...,14;\ n=1,...,N(t);\ \tau^1\equiv 0; \qquad (6.25)$$

Regression model (6.25) was run using the 14 quarters of sales data for the town of "A". Note that a single common straight line depreciation rate δ is estimated. The estimated decade (net) depreciation rate[20] was $\hat{\delta}=11.94\%$ (or around 1.2% per year), which is very reasonable. As was the case with model (6.24), if a house with the *same characteristics* in two consecutive periods *t* and *t*+1 could be observed, the corresponding price relative (neglecting error terms) $\exp(\tau^{t+1})/\exp(\tau^t)$ can serve as the chain link in a price index; see Figure 6.1 and Table 6.1 for the resulting index, labeled P_{H2}. The R^2 for this model was .8345, a bit lower than the previous model and the log likelihood was 1354.9, which is quite a drop from the previous log likelihood of 1407.6.[21]

It appears that the imposition of more theory – with respect to the treatment of the age of the house – has led to a drop in the empirical fit of the model. However, it is likely that this model

[19] This regression is essentially linear in the unknown parameters and hence it is very easy to estimate.

[20] It is a net depreciation rate because we have no information on renovation expenditures, i.e., δ is equal to gross wear and tear depreciation of the house less average expenditures on renovations and repairs.

[21] The levels type R^2 for this model was $R^{*2}=.7647$, which again is quite a drop from the corresponding levels R^2 for the previous log price model.

and the previous one are mis-specified[22]: they both multiply together land area times structure area to determine the price of the house while it is likely that an additive interaction between L and S is more appropriate than a multiplicative one.

Note that, given the depreciation rate δ, *quality adjusted structures* (adjusted for the aging of the structure) for each house n in each quarter t can be defined as follows:

$$S_n^{t*} \equiv (1-\delta A_n^t)S_n^t; \qquad t = 1,...,14;\ n = 1,...,N(t). \qquad (6.26)$$

The Log Log Time Dummy Model with Quality Adjustment of Structures for Age

In the remainder of this section, quality adjusted (for age) structures, $(1-\delta A)S$, will be used as an explanatory variable, rather than the unadjusted structures area, S. The log log model is similar to the previous log linear model, except that now, instead of using L and $(1-\delta A)S$ as explanatory variables in the regression model, the logarithms of the land and quality adjusted structures areas are used as independent variables. Thus the l*og log time dummy hedonic regression model with quality adjusted structures* is the following:[23]

$$\ln p_n^t = \alpha + \beta \ln L_n^t + \gamma \ln[(1-\delta A_n^t)S_n^t] + \tau^t + \varepsilon_n^t, \qquad t = 1,...,14;\ n = 1,...,N(t);\ \tau^1 \equiv 0; \qquad (6.27)$$

Using the data for the Dutch town of "A", the estimated decade (net) depreciation rate was $\hat{\delta} = 0.1050$ (standard error 0.00374). Note that if both sides of (6.27) were exponentiated and the error terms were neglected, the house price p_n^t would equal $\exp(\alpha)[L_n^t]^\beta [S_n^{t*}]^\gamma \exp(\tau^t)$, where S_n^{t*} denotes quality adjusted structures as defined by (6.26). So if we could observe a house with the *same characteristics* in two consecutive periods t and t+1, the corresponding price relative (neglecting error terms) would be equal to $\exp(\tau^{t+1})/\exp(\tau^t)$ and this again can serve as the chain link in a price index; see Figure 6.1 and Table 6.1 for the resulting index, labeled P_{H3}. The R^2 for this model was .8599 (the levels measure of fit was R^{*2}= .8880), which is an increase over models (6.25) and (6.26); the log likelihood was 1545.4, a big increase over the log likelihoods for the other two models (1407.6 and 1354.9).

The house price series generated by the three log-linear time dummy regressions described in this section, P_{H1}, P_{H2} and P_{H3}, are plotted in Figure 6.1 along with the chained stratified

[22] If the variation in the independent variables is relatively small, the difference in indexes generated by the various hedonic regression models considered in this section and the following two sections is likely to be small since virtually all of the models considered can offer roughly a linear approximation to the "truth". But when the variation in the independent variables is large, as it is in the present housing context, the choice of functional form can have a substantial effect. Thus a priori reasoning should be applied to both the choice of independent variables in the regression as well as to the choice of functional form. For additional discussion on functional form issues, see Diewert (2003a).

[23] This hedonic regression model turns out to be a variant of McMillen's (2003) *consumer oriented approach to hedonic housing models*. His theoretical framework draws on the earlier work of Muth (1971) and is outlined in Diewert, de Haan and Hendriks (2010). See also McDonald (1981).

sample mean Fisher index described in section 5.5, P_{FCH}. These four house price series are listed in Table 6.1. All four indexes capture the same trend but there can be differences of over 2 percent between them in some quarters. Notice that all of the indexes move in the same direction from quarter to quarter with decreases in quarters 4, 8, 12 and 13 except that P_{H3} – the index that corresponds to the log log time dummy model – increases in quarter 12.

Figure 6.1. Log-Linear Time Dummy Price Indexes and the Chained Stratified Sample Mean Fisher Price Index

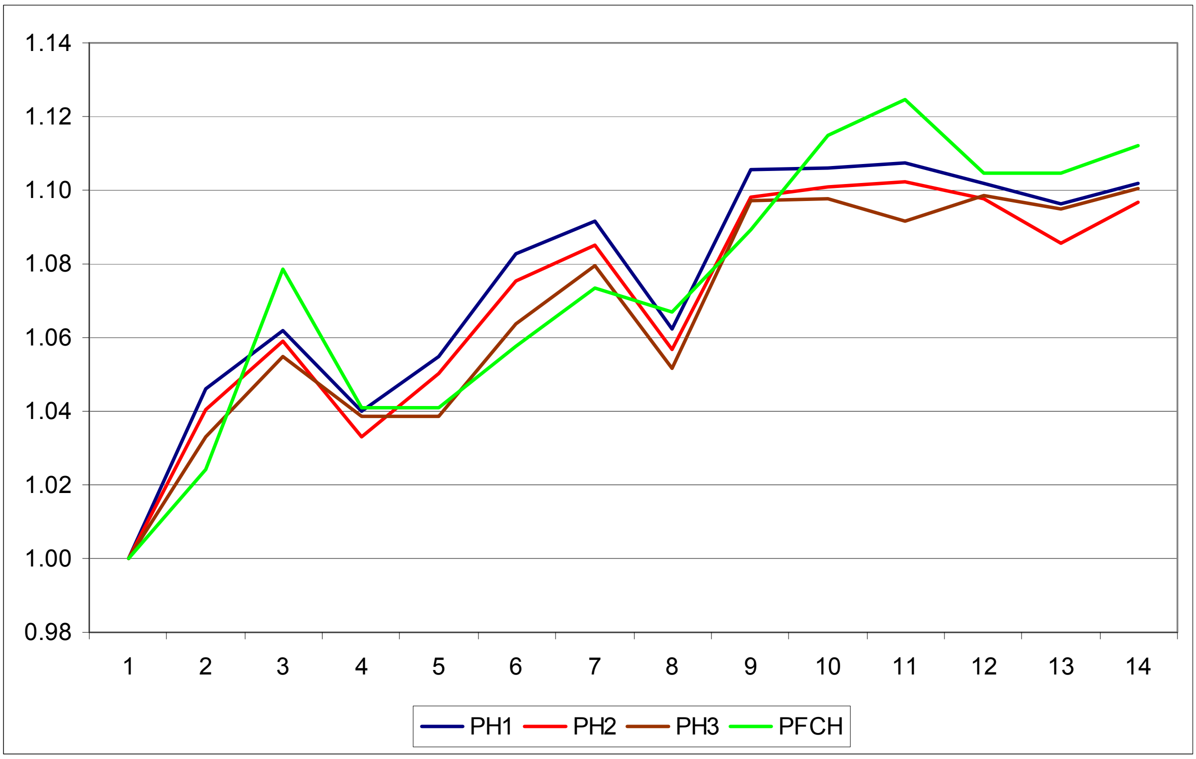


Table 6.1. Log-Linear Time Dummy Price Indexes and the Chained Stratified Sample Mean Fisher Price Index

Quarter	P_{H1}	P_{H2}	P_{H3}	P_{FCH}
1	1.00000	1.00000	1.00000	1.00000
2	1.04609	1.04059	1.03314	1.02396
3	1.06168	1.05888	1.05482	1.07840
4	1.04007	1.03287	1.03876	1.04081
5	1.05484	1.05032	1.03848	1.04083
6	1.08290	1.07532	1.06369	1.05754
7	1.09142	1.08502	1.07957	1.07340
8	1.06237	1.05655	1.05181	1.06706
9	1.10572	1.09799	1.09736	1.08950
10	1.10590	1.10071	1.09786	1.11476
11	1.10722	1.10244	1.09167	1.12471
12	1.10177	1.09747	1.09859	1.10483
13	1.09605	1.08568	1.09482	1.10450
14	1.10166	1.09694	1.10057	1.11189

Although model (6.27) performs the best of the simple hedonic regression models considered thus far, it has the unsatisfactory feature that the quantities of land and of quality adjusted structures determine the price of a property in a *multiplicative manner*. It is more likely that house prices are determined by a weighted *sum* of their land and quality adjusted structures amounts. In the following section, an additive time dummy model will therefore be estimated. The expectation is that this model will fit the data better.

6.7 Time Dummy Hedonic Regression Models using Price as the Dependent Variable

The Linear Time Dummy Hedonic Regression Model

There are reasons to believe that the selling price of a property is linearly related to the plot area of the property plus the area of the structure due to the competitive nature of the house building industry.[24] If the age of the structure is treated as another characteristic that has an importance in determining the price of the property, then the following *linear time dummy hedonic regression model* might be an appropriate one:

$$p_n^t = \alpha + \beta L_n^t + \gamma S_n^t + \delta A_n^t + \tau^t + \varepsilon_n^t, \qquad t = 1,\ldots,14;\ n = 1,\ldots,N(t);\ \tau^1 \equiv 0. \qquad (6.28)$$

The above linear regression model was run using the data for the town of "A". The R^2 for this model was .8687, much higher than those obtained in the previous regressions[25] and the log likelihood was −10790.4 (which cannot easily be compared to the previous log likelihoods since the dependent variable has changed from the logarithm of price to just price[26]).

Using the linear model defined by equations (6.28) to form an overall house price index is a bit more difficult than using the previous log-linear or log log time dummy regression models. In section 6.6, holding characteristics constant and neglecting error terms, the *relative price* for the same house over any two periods turns out to be constant, leading to an unambiguous overall index. In the present situation, holding characteristics constant and neglecting error terms, the *difference in price* for the same house turns out to be constant, but the *relative prices* for different houses will not in general be constant. Therefore, an overall index will be constructed which uses the prices generated by the estimated parameters for model (6.28) and

[24] See Clapp (1980), Francke and Vos (2004), Gyourko and Saiz (2004), Bostic, Longhofer and Redfearn (2007), Davis and Heathcote (2007), Francke (2008), Diewert (2009b), Koev and Santos Silva (2008), Statistics Portugal (2009), Diewert, de Haan and Hendriks (2010), Diewert (2010) and Chapter 9 below.

[25] However, recall that the levels adjusted measure of fit for the log log model described by (6.27) was .8880, which is higher than .8687.

[26] Marc Francke has pointed out that it is possible to compare log likelihoods across two models where the dependent variable has been transformed by a known function in the second model; see Davidson and McKinnon (1993; 491) where a Jacobian adjustment makes it possible to compare log likelihoods across the two models.

evaluated at the sample average amounts of L, S and the sample average age of a house A.[27] The resulting quarterly prices for this "average" house were converted into an index, P_{H4}, which is listed in Table 6.2 and charted in Figure 6.2.

The hedonic regression model defined by (6.28) is perhaps the simplest possible one but it is a bit too simple since it neglects the fact that the interaction of age with the selling price of the property takes place via a multiplicative interaction with the structures variable and not via a general additive factor. In what follows, model (6.28) is re-estimated using quality adjusted structures as an explanatory variable rather than just entering age A as a separate stand alone characteristic.

The Linear Time Dummy Model with Quality Adjusted Structures

The *linear time dummy hedonic regression model with quality adjusted structures* is described by

$$p_n^t = \alpha + \beta L_n^t + \gamma(1-\delta A_n^t)S_n^t + \tau^t + \varepsilon_n^t, \qquad t = 1,...,14;\ n = 1,...,N(t);\ \tau^1 \equiv 0. \qquad (6.29)$$

This is the most plausible hedonic regression model so far. It works with quality adjusted (for age) structures S^* equal to $(1-\delta A)S$ instead of having A and S as completely independent variables that enter into the regression in a linear fashion.

The results for this model were a clear improvement over the results of model (6.28). The log likelihood increased by 92 to −10697.8 and the R^2 increased to .8789 from the previous .8687. The estimated decade depreciation rate was $\hat{\delta} = 0.1119$ (0.00418), which is reasonable as usual. This linear regression model has the same property as the model (6.28): house price *differences* are constant over time for all constant characteristic models but house price *ratios* are not constant. So again an overall index will be constructed which uses the prices generated by the estimated parameters in (6.29) and evaluated at the sample average amounts of L, S and the average age of a house A. The resulting quarterly house prices for this "average" model were converted into an index, P_{H5}, which is listed in Table 6.2 and charted in Figure 6.2. For comparison purposes, P_{H3} (the time dummy Log Log model index) and P_{FCH} (the chained stratified sample mean Fisher index) will be charted along with P_{H4} and P_{H5}. The preferred indexes thus far are P_{FCH} and P_{H5}.

It can be seen that again, all four indexes capture the same trend but there can be differences of over 2 percent between the various indexes for some quarters. Note that all of the indexes move in the same direction from quarter to quarter with decreases in quarters 4, 8, 12 and 13, except that P_{H3} increases in quarter 12.

[27] The sample average amounts of L and S were 257.6 m^2 and 127.2 m^2 respectively and the average age of the detached dwellings sold over the sample period was 1.85 decades.

A major problem with the hedonic time dummy regression models considered thus far is that the prices of land and quality adjusted structures are not allowed to change in an unrestricted manner from period to period. The class of hedonic regression models to be considered in the following section does not suffer from this problem.

Figure 6.2. Linear Time Dummy Price Indexes, the Log Log Time Dummy Price Index and the Chained Stratified Sample Mean Fisher Price Index

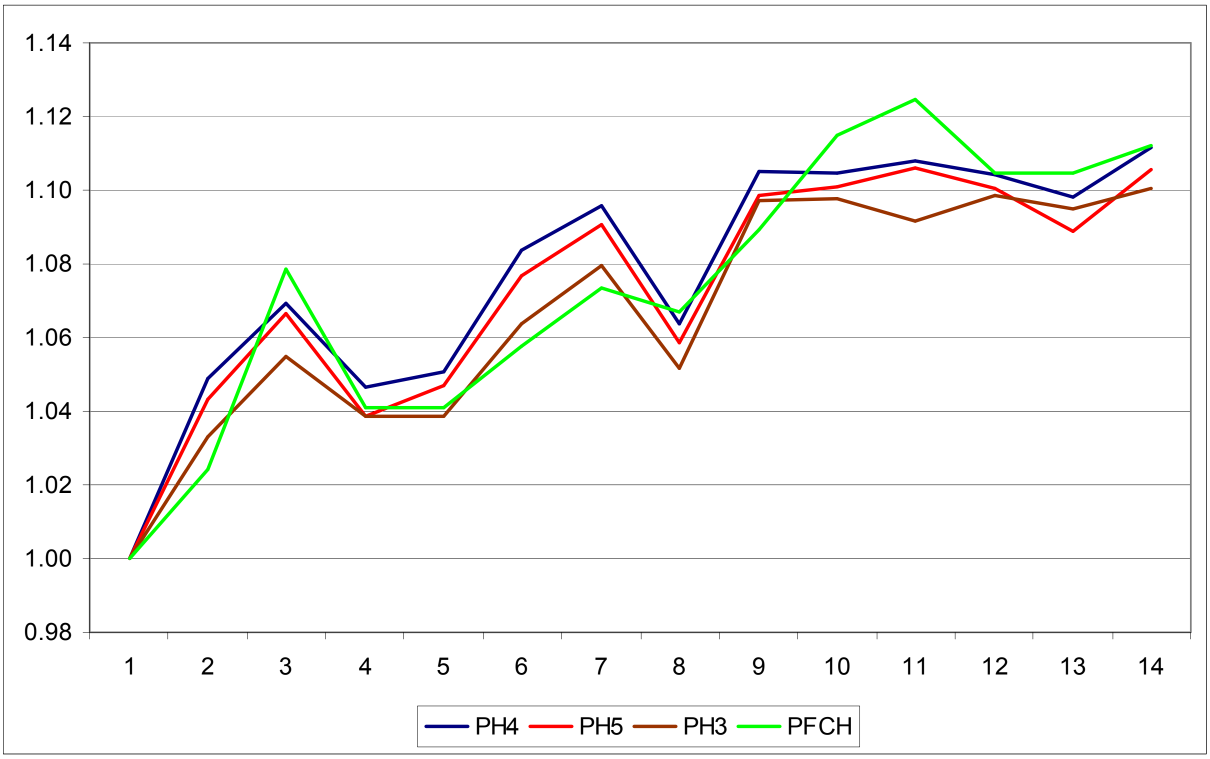


Table 6.2: Linear Time Dummy Price Indexes, the Log Log Time Dummy Price Index and the Chained Stratified Sample Mean Fisher Price Index

Quarter	P_{H4}	P_{H5}	P_{H3}	P_{FCH}
1	1.00000	1.00000	1.00000	1.00000
2	1.04864	1.04313	1.03314	1.02396
3	1.06929	1.06667	1.05482	1.07840
4	1.04664	1.03855	1.03876	1.04081
5	1.05077	1.04706	1.03848	1.04083
6	1.08360	1.07661	1.06369	1.05754
7	1.09593	1.09068	1.07957	1.07340
8	1.06379	1.05864	1.05181	1.06706
9	1.10496	1.09861	1.09736	1.08950
10	1.10450	1.10107	1.09786	1.11476
11	1.10788	1.10588	1.09167	1.12471
12	1.10403	1.10044	1.09859	1.10483
13	1.09805	1.08864	1.09482	1.10450
14	1.11150	1.10572	1.10057	1.11189

6.8 Hedonic Imputation Regression Models

The theory of *hedonic imputation indexes* was explained in section 6.3 and is applied to the present situation as follows. For each period, run a linear regression of the following form:

$$p_n^t = \alpha^t + \beta^t L_n^t + \gamma^t (1 - \delta^t A_n^t) S_n^t + \varepsilon_n^t , \qquad t = 1,...,14;\ n = 1,...,N(t). \qquad (6.30)$$

Using the data for the town of "A", there are only four parameters to be estimated for each quarter: α^t, β^t, γ^t and δ^t for t = 1,...,14. Note that (6.30) is similar in form to the model defined by equations (6.29), but with some significant differences:

- Only one depreciation parameter is estimated in the model defined by (6.29) whereas in the present model, there are 14 depreciation parameters; one for each quarter.
- Similarly, in model (6.29), there was only one α, β and γ parameter whereas in (6.30), there are 14 α^t, 14 β^t and 14 γ^t parameters to be estimated. On the other hand, model (6.29) had an additional 13 time shifting parameters (the τ^t) that required estimation.

Thus the hedonic imputation model involves the estimation of 56 parameters, the time dummy model only 17, so it is likely that the hedonic imputation model will fit the data much better.

As usual, in the housing context, precisely matched models across periods do not exist; there are always depreciation and renovation activities that make a house in the exact same location not quite comparable over time. This lack of matching, say between quarters t and t+1, can be overcome in the following way: take the parameters estimated using the quarter t+1 hedonic regression and price out all of the housing models (i.e., sales) that appeared in quarter t. This generates *predicted quarter t+1 prices for the quarter t models*, $\hat{p}_n^{t+1}(t)$, as follows:

$$\hat{p}_n^{t+1}(t) \equiv \hat{\alpha}^{t+1} + \hat{\beta}^{t+1} L_n^t + \hat{\gamma}^{t+1} (1 - \hat{\delta}^{t+1} A_n^t) S_n^t , \qquad t = 1,...,13;\ n = 1,...,N(t). \qquad (6.31)$$

where $\hat{\alpha}^t$, $\hat{\beta}^t$, $\hat{\gamma}^t$ and $\hat{\delta}^t$ are the parameter estimates for model (6.30) for t = 1,...,14. Now we have a set of pseudo matched quarter t+1 prices for the models that appeared in quarter t and the following *Laspeyres type hedonic imputation (or pseudo matched model) index*, going from quarter t to t+1, can be formed:[28]

$$P_{HIL}^{t,t+1} \equiv \frac{\sum_{n=1}^{N(t)} 1 \hat{p}_n^{t+1}(t)}{\sum_{n=1}^{N(t)} 1 p_n^t} . \qquad t = 1,...,13. \qquad (6.32)$$

[28] Due to the fact that the regressions defined by (6.30) have a constant term and are essentially linear in the explanatory variables, the sample residuals in each of the regressions will sum to zero. Hence the sum of the predicted prices will equal the sum of the actual prices for each period. Thus the sum of the actual prices in the denominator of (6.32) will equal the sum of the corresponding predicted prices and similarly, the sum of the actual prices in the numerator of (6.34) will equal the corresponding sum of the predicted prices. This was explained earlier in section 6.3.2.

As mentioned earlier, the quantity that is associated with each price is 1 as each housing unit is basically unique and can only be matched through the use of a model.

The same method can be applied going backwards from the housing sales that took place in quarter t+1; take the parameters for the quarter t hedonic regression and price out all of the housing models that appeared in quarter t+1 and generate predicted prices, $\hat{p}_n^t(t+1)$, for these t+1 models:

$$\hat{p}_n^t(t+1) \equiv \hat{\alpha}^t + \hat{\beta}^t L_n^{t+1} + \hat{\gamma}^t (1-\hat{\delta}^t A_n^{t+1}) S_n^{t+1}; \qquad t = 1,\ldots,13;\ n = 1,\ldots,N(t+1). \quad (6.33)$$

Now we have a set of "matched" quarter t prices for the models that appeared in period t+1 and we can form the following *Paasche type hedonic imputation (or pseudo matched model) index*, going from quarter t to t+1:

$$P_{HIP}^{t,t+1} \equiv \frac{\sum_{n=1}^{N(t+1)} 1 p_n^{t+1}}{\sum_{n=1}^{N(t+1)} 1 \hat{p}_n^t(t+1)}; \qquad t = 1,\ldots,13. \quad (6.34)$$

Once the above Laspeyres and Paasche imputation price indexes have been calculated, the corresponding *Fisher type hedonic imputation index* going from period t to t+1 can be formed by taking the geometric average of the two indexes defined by (6.32) and (6.34):

$$P_{HIF}^{t,t+1} \equiv \left[P_{HIL}^{t,t+1} P_{HIP}^{t,t+1}\right]^{1/2}. \qquad t = 1,\ldots,13. \quad (6.35)$$

The resulting chained Laspeyres, Paasche and Fisher imputation price indexes, P_{HIL}, P_{HIP} and P_{HIF}, based on the data for the town of "A", are plotted below in Figure 6.3 and are listed in Table 6.3. The three imputation indexes are amazingly close. The Fisher imputation index is our preferred hedonic price index thus far; it is better than the time dummy indexes in sections 6.6. and 6.7 because imputation allows the price of land and of quality adjusted structures to change independently over time, whereas the time dummy indexes shift the hedonic surface in a parallel fashion. The empirical results indicate that, at least for the present data set for the town of "A", the Laspeyres imputation index provides a close approximation to the preferred Fisher imputation index.

To conclude: our two "best" indexes are the Fisher imputation index P_{HIF} and the stratified chained Fisher index P_{FCH}. Overall, the hedonic imputation index P_{HIF} should probably be preferred to P_{FCH} since the stratified sample indexes will have a certain amount of unit value bias which will most likely be greater than any functional form bias in P_{HIF}. These two "best" indexes are plotted in Figure 6.4 along with the log-log time dummy index P_{H3} and the linear time dummy index with quality adjusted structures P_{H5}. All of the price indexes except P_{H3} show downward movements in quarters, 4, 8, 12 and 13 and upward movements in the other quarters; P_{H3} moves up in quarter 12 instead of falling like the other indexes.

Figure 6.3. Chained Laspeyres, Paasche and Fisher Hedonic Imputation Price Indexes

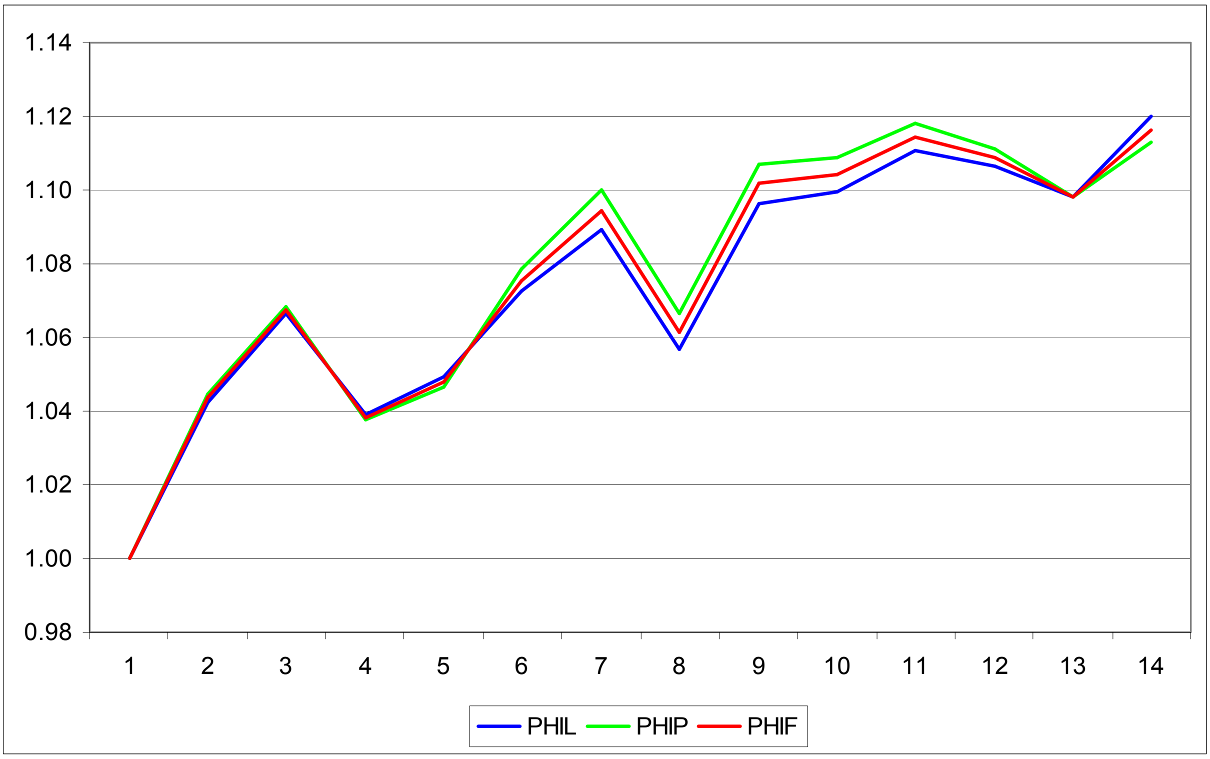
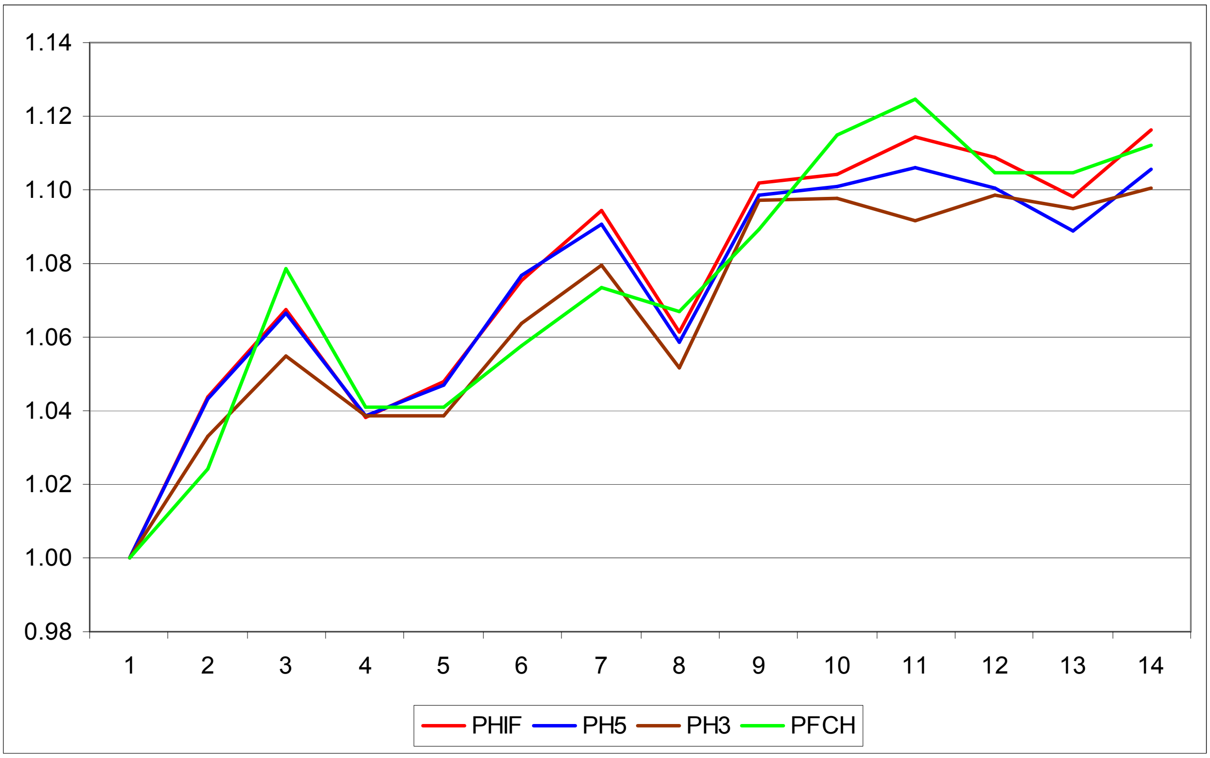


Table 6.3. Chained Laspeyres, Paasche and Fisher Hedonic Imputation Price Indexes

Quarter	P_{HIL}	P_{HIP}	P_{HIF}
1	1.00000	1.00000	1.00000
2	1.04234	1.04479	1.04356
3	1.06639	1.06853	1.06746
4	1.03912	1.03755	1.03834
5	1.04942	1.04647	1.04794
6	1.07267	1.07840	1.07553
7	1.08923	1.10001	1.09460
8	1.05689	1.06628	1.06158
9	1.09635	1.10716	1.10174
10	1.09945	1.10879	1.10411
11	1.11062	1.11801	1.11430
12	1.10665	1.11112	1.10888
13	1.09830	1.09819	1.09824
14	1.11981	1.11280	1.11630

Figure 6.4. The Fisher Imputation Price Index, the Chained Stratified Sample Mean Fisher Price Index, the Linear Time Dummy Price Index and the Log Log Time Dummy Price Index



Chapter 7: Repeat Sales Methods

7.1 The Basic Repeat Sales Model

The repeat sales method was initially proposed by Bailey, Muth and Nourse (1963). They saw their procedure as a generalization of the *chained matched model methodology* applied by the pioneers in the construction of real estate price indexes like Wyngarden (1927) and Wenzlick (1952). The best-known repeat sales indexes are the Standard and Poor's/Case-Shiller Home Price Indexes in the US, which are computed for 20 cities (Standard and Poor's, 2009). The Federal Housing Finance Agency (FHFA) also computes a repeat sales index for the US,[1] using a slightly different approach. Residex and the UK Land Registry compute repeat sales indexes for Australian cities and for the UK, respectively.[2]

As the name indicates, the method utilizes information on properties which have been sold more than once. Because it is a matched-properties type of method, there is no change in the quality mix to control for. However, because of the low incidence of resales, it would not be very useful to compute a repeat sales RPPI using the standard matched model methodology and conventional index number formulae. Therefore, a stochastic model is postulated which "explains" the price changes of houses that have been sold repeatedly. This (dummy variable) regression model is then estimated on the pooled data, i.e. on the pooled price changes, across the sample period.

The only information required to estimate a standard repeat sales regression equation is price, sales date and address of the properties, so the repeat sales method is much less data intensive than hedonic methods. Also, the repeat sales method automatically controls for location at the most detailed (address) level, something which hedonic regression methods are unable to do.[3] One problem is that a dwelling unit at two different points in time is not necessarily the same due to depreciation and quality improvements done to it. The longer the time interval between the first and second sale, the more questionable the constant-quality assumption underlying the repeat sales approach becomes.

The following stochastic model explaining the logarithm of the value (price) p_n^t for property n in period t can be found in the literature:

[1] The FHFA was established in 2008 as a combination of the former US Office of Federal Housing Enterprise Oversight (OFHEO), who published the repeat sales index until then, and the Federal Housing Finance Board (FHFB).

[2] The Dutch Land Registry computed a repeat sales index for the Netherlands until 2007 when they changed over to a SPAR index, which is published jointly with Statistics Netherlands. For the SPAR method, see Chapter 8.

[3] However, the use of geospatial data to allow for spatial dependence in the hedonic equation could ameliorate the omitted locational variables problem; see Chapter 6 and Hill (2011) for more details.

$$\ln p_n^t = P^t + H_n^t + \varepsilon_n^t, \tag{7.1}$$

where P^t is a common term for all properties (the log of "price level" in some region or city), H_n^t is a Gaussian random walk that represents the drift in individual housing value over time, and ε_n^t is a random error term or white noise. Model (7.1) is often taken as the starting point for deriving the estimating repeat sales equation.

Another point of departure could be the constrained log-linear hedonic model (6.4), where the parameters β_k of the price-determining characteristics are constrained to be fixed over time. As "identical" properties are assumed to be compared, there is a second restriction involved: the (amounts of the) characteristics of an individual property are also assumed to be constant. Denoting the *k*'th characteristic for property *n* by z_{nk}, the constrained log-linear model now becomes

$$\ln p_n^t = \beta_0^t + \sum_{k=1}^{K} \beta_k z_{nk} + \varepsilon_n^t. \tag{7.2}$$

A model for the logarithm of the change in value of property *n* between two periods, say *s* and *t* $(0 \le s < t \le T)$, is found by subtracting (7.2) for those periods. It follows that

$$\ln p_n^t - \ln p_n^s = \ln(p_n^t / p_n^s) = (\beta_0^t - \beta_0^s) + (\varepsilon_n^t - \varepsilon_n^s) = \ln P^{st} + (\varepsilon_n^t - \varepsilon_n^s). \tag{7.3}$$

What model (7.3) essentially says is that, apart from the error term $\varepsilon_n^t - \varepsilon_n^s$, the (logarithm of) the value change is the same for all properties, denoted by P^{st}, independent of their quality attributes.

Now suppose we have a sample of houses that have been sold more than once over the sample period $t = 0,...,T$ for which we have data on transaction prices, hence on their price changes. The (holding) period between subsequent sales will differ among those properties. However, given that in model (7.3) all individual property prices are expected to change with the same rate, excluding random disturbances, we can pool the repeat sales data and estimate the model through the standard *repeat sales equation*

$$\ln(p_n^t / p_n^s) = \sum_{t=0}^{T} \gamma^t D_n^t + \mu_n^t, \tag{7.4}$$

where D_n^t is a dummy variable with the value 1 in the period that the resale occurs, -1 in the period that the previous sale occurs, and 0 otherwise; μ_i^t is again an error term.[4] Under the classical error assumptions, in particular that they have a zero mean and constant variance, the repeat sales model (7.4) can be estimated by OLS regression. Multicollinearity may be present in the data, but there is not much one can do about that.

[4] Multiple resales are treated as independent observations. As noted by Shiller (1991), this should not be overly problematic because there is no overlap between the holding periods of multiple resales.

The repeat sales index going from period 0 to period t is obtained by exponentiating the corresponding regression coefficients $\hat{\gamma}^t$:

$$P_{RS}^{0t} = \exp(\hat{\gamma}^t) . \tag{7.5}$$

The simplicity of the standard repeat sales model, which only includes dummy variables, the fact that no characteristics data is needed other than address,[5] and the straightforward way to compute the price index numbers could perhaps explain part of the popularity of the method in the real estate and housing literature.

Wang and Zorn (1997) derived an analytical expression for the repeat sales index. It appears to have a rather complex geometric structure. Thus, despite the fact that the idea of matching is easily understood, the method may be difficult to explain in detail. Moreover, as mentioned earlier, a geometric property price index may be undesirable as a target, especially for a stock RPPI. A solution could be the use of an arithmetic version of the repeat sales method, which was suggested by Shiller (1991). Standard and Poor's (Case-Shiller) Home Price Indexes are based on the arithmetic repeat sales method (see Standard and Poor's, 2009).

7.2 Issues and Improvements to the Basic Model

In this section we will discuss a number of issues related to the repeat sales method and give a brief overview of extensions and improvements to the basic model that have been proposed in the literature.

Data Cleaning

In practical applications, properties that were resold very rapidly, as well as those that were not resold for long periods, have sometimes been excluded from the repeat sales regressions as such transactions might be "unusual" and therefore bias the resulting price index. Clapp and Giacotto (1998) and Steele and Goy (1997) suggested eliminating very short holds from the dataset as these could be distressed sales arising from, for example, divorce or job loss, or speculative transactions. Jansen et al. (2008), using data from the Dutch Land Registry, found that houses resold within 12 months showed relatively strong price increases.

Reproducibility is one of the strengths of the repeat sales approach. But if the procedure for excluding "unusual" observations differs from time to time, then the reproducibility might be endangered.

---

[5] In some countries, such as the UK and the Netherlands, the Land Registry collects all transaction price data but only a very limited number of characteristics, like type of dwelling and of course address. It is therefore not surprising to see that in those countries repeat sales indexes have been computed from the Land Registry data. Note that the FHFE's repeat sales index in the US is based on data obtained from Fannie Mae and Freddie Mac for mortgages.

Heteroskedasticity

Case and Shiller (1987, 1989) argued that changes in house prices include components whose variances increase with the interval of sales, so that the assumption of a constant variance of the errors does not hold. They propose a Weighted Least Squares (WLS) approach to correct for this type of heteroskedasticity. The weights are derived by regressing the squared residuals from the standard (OLS) repeat sales regression on an intercept and the time interval between sales. A modified version of their weighted repeat sales approach is used by the US Federal Housing Finance Agency to construct quarterly price indexes for single-family homes. It can be argued that the error variance will be non-linear in time intervals (Calhoun, 1996), hence the squared OLS residuals are regressed on an intercept term, the time interval and the square of the time interval.

Some studies found ambiguous results for heteroskedasticity adjustment. Leishman, Watkins and Fraser (2002), using Scottish data, and Jansen et al. (2008), using Dutch data, applied the standard (OLS) repeat sales method and various weighted methods and both concluded that the standard method was not inferior.

Sample Selection Bias

An important problem with repeat sales regressions is the possibility of sample selection bias. The problem is that some types of houses may trade more frequently on the market than other types so that they will be over-represented in the repeat sales sample (with respect to the stock of houses or the sales during some period). When these types of houses exhibit different price changes, then the repeat sales index tends to be biased. For example, if low quality houses sell more frequently than high quality houses but high quality houses rise in price at a slower rate, a repeat sales index will tend to have an upward bias.

There are various reasons why the holding duration of properties can be unevenly distributed. Life-cycle theories on property holding periods suggest that less expensive houses are traded more frequently; when people move up the property ladder they will tend to move home less often. Lower transaction costs for less expensive properties, for instance due to lower stamp duties, may also result in a higher turnover rate of less expensive homes. In addition, the Buy-to-Let market in some countries is more active in lower price segments.

Quite a few studies addressed the issue of holding duration and sample selection bias in repeat sales price indexes, e.g., Case, Pollakowski and Wachter (1991) (1997), Cho (1996), Clapp, Giacotto and Tirtiroglu (1991), Gatzlaff and Haurin (1997), Hwang and Quigley (2004), and Steele and Goy (1997). Not all these studies found strong evidence of sample selection bias. Clapp, Giacotto and Tirtiroglu (1991) did not find systematic differences between the repeat sales sample and the full sample of transactions over the long run. They argued that arbitrage typically forces prices for the repeat sample to grow at the same rate as the prices for the full sample. Wallace and Meese (1997) concluded that their repeat sales sample was sufficiently representative of all sales during the sample period in question. However, the "sample" of all housing sales themselves may not be representative of the total housing stock.

Potential sample selection problems are inherent to the repeat sales method. To some extent they can be corrected for by stratifying the repeat sales sample. A problem in this context is that the sub-samples may become very small and produce volatile indexes. Thus there may be a case for smoothing the index numbers. Moreover, it can be argued that selling prices do not always exactly represent the market values of the properties, which can be viewed as a latent variable. There may be *transaction noise* involved that causes volatility of the measured price indexes. Francke (2010) proposed a smoothing procedure that takes into account the fact that selling prices of repeatedly sold properties depend on the time interval between subsequent sales.

Inefficiency and Revision

The repeat sales method is often criticized for being inefficient as, by construction, it "throws away a lot of data". This is true compared with the multi-period time dummy hedonic method: the repeat sales prices data set is usually a lot smaller than the prices data set for all sales over the sample period. On the other hand, the longer the sample period, the more data will be used by the repeat sales method (as more and more houses will have been resold). Hence, when the sample period grows and more data is added, the efficiency of the repeat sales method will increase faster than that of the hedonic approach. Besides, the repeat sales method is efficient in the sense that it does not use any other housing characteristics than address.

It is possible to augment a repeat sales dataset by using assessment data (also referred to as appraisals) as approximations for past or current values of houses that have not been resold during the sample period. Some of the data on which the repeat sales index would then be based would be pseudo rather than genuine repeat data. Most empirical studies on this issue are based on appraisals of dwellings that are about to be re-financed. It has been suggested that appraisals tend to over-predict the actual selling price of the property. But the magnitude of the bias could depend on the purpose for which assessment information is collected. De Vries et al. (2009) investigated the reliability of the Dutch appraisals, which are collected on the government's behalf for income and local tax purposes, and concluded that the quality was satisfactory and improving over time. For more on the use of assessment information in a repeat sales index and the removal of appraisal bias, see e.g. Geltner (1996), Edelstein and Quan (2006) and Leventis (2006).

Similar to the multi-period time dummy method, the repeat sales method suffers from revision of previously computed figures: when additional repeat sales information becomes available, re-estimation will result in changes in the coefficients estimated and thus in the price indexes inferred. There have been few empirical studies on this issue to date, e.g. Clapp and Giaccotto (1999), Butler, Chang and Crews Cutts (2005), and Clapham et al. (2006). The last authors found evidence to suggest that repeat-sales indexes are relatively less stable than time dummy hedonic indices. Note that revisions may be related to sample selection bias; when the sample period is extended and the coefficients re-estimated, sample selection bias might be getting smaller as more and more repeat sales are observed.

Quality Change

Repeat sales indexes are estimated on the premise that the quality of the individual properties (as measured by their characteristics) has not changed over time. It is sometimes argued that in the aggregate, renovations are approximately equal to depreciation. For individual dwelling units, however, this cannot be true because over time, many units are demolished. One way to control for this is to use a sub-sample of repeat sales, where quality is thought to be relatively constant. Case and Shiller (1989), for example, "extracted […] data on houses sold twice for which there was no apparent quality change". The problem is that the price changes inferred may not be indicative of the price changes for the full sample repeat sales and may exacerbate the sample selection bias problem.[6]

If information on maintenance and renovation expenditures was available at the micro level, this could be used in the context of estimating a repeat sales (or hedonic) regression model for housing. In practice this kind of information is often lacking. Abraham and Schauman (1991) suggested adjusting the repeat sales index from aggregate data on renovation expenditures and make an adjustment for depreciation of the structures; see also Palmquist (1980) (1982). This approach to measuring net depreciation seems too crude and arbitrary to be suitable for the compilation of official statistics, however

Shimizu, Nishimura and Watanabe (2010) recently developed a repeat sales method that takes net depreciation into account. Their method relies on an unknown taste parameter for which a guesstimate has to be made. While making an adjustment seems to be better than completely ignoring the (net) depreciation problem, making guesses might not be an attractive option for statistical agencies.

Shiller (1993a) developed a repeat sales method that accounts for possible changes in housing characteristics between first and second sales. The method involves including characteristics in a traditional repeat sales model. Clapp and Giaccotto (1998) advocated the use of assessed values at time of first and second sales as a parsimonious control for quality changes of the properties. Goetzmann and Spiegel (1997) suggested including a constant term in the repeat-

[6] Meese and Wallace (1997) report that repeat sales units with changed characteristics tend to be larger and in worse condition than the average of units with single transactions.

sales regression to capture average quality change across all characteristics over the average holding period.

Case and Quigley (1991) were the first to advocate *hybrid models*. Hybrid models exploit all sales data by combining repeat sales and hedonic regressions and address not only the quality change problem but also sample selection bias and inefficiency problems. Case and Quigley (1991) and Quigley (1995) used samples of single-sale and repeat-sale properties to jointly estimate price indexes using generalized least squares regression. Hill, Knight and Sirmans (1997) undertook a similar though more general exercise. Their model stacks two equations, a time dummy hedonic model (including age of the dwelling) and a repeat sales model, which are jointly estimated using maximum likelihood. They used a characteristics prices method to derive the price indexes; see Chapter 6, equation (6.9).[7]

The rationale for hybrid methods is to try and combine the best features of the repeat sales and hedonic approaches. By combining both approaches, no data are discarded while repeat sales are still allowed to play a prominent role in the index construction methodology. However, we agree with Hill (2011) who has difficulty accepting that a repeat-sales price relative should be preferred to a (say double) imputation hedonic price relative. He notes that: “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”. We add that the complexity of hybrid models most likely makes them unsuitable for implementation by statistical agencies.

7.3 Main Advantages and Disadvantages

Below, the main advantages and disadvantages of the repeat sales method are listed. The main advantages are:

- The repeat sales method in its basic form needs no characteristics other than address of the properties that are transacted more than once over the sample period. Source data may be available from administrative records such as those from the Land Registry.
- Standard repeat sales regressions are easy to run and the price indexes easy to construct.
- The repeat sales method is a matched-model type of method without any imputations. By construction, location is automatically controlled for.
- The results are essentially reproducible provided that the treatment of outliers and possible corrections for heteroskedasticity (as well as the choice between a geometric or arithmetic method) are clearly described.

The main disadvantages of the repeat sales method are:

[7] Other papers on the use of hybrid models include Clapp and Giaccotto (1992), Knight, Dombrow and Sirmans (1995), Englund, Quigley and Redfearn (1998), and Hwang and Quigley (2004).

- The method is inefficient in the sense that it does not use all of the available selling prices; it uses only information on units that have sold more than once during the sample period.
- The basic version of the method ignores (net) depreciation of the dwelling unit.[8]
- There may be a sample selection bias problem in repeat sales data.
- The method cannot provide separate price indexes for land and for structures.
- The method cannot be used if indexes are required for very fine classifications of the type of property sold. In particular, if monthly property price indexes are required, the method may fail due to a lack of market sales for smaller categories of property.
- In principle, estimates for past price change obtained by the repeat sales method should be updated as new transaction information becomes available. Thus the repeat sales property price index could be subject to never ending revision.[9]

Haurin and Hendershott (1991) summarize the disadvantages of the repeat sales method as follows:

"The method is subject to many criticisms: (1) it does not separate house price change from depreciation, (2) renovation between sales is ignored, (3) the sample is not representative of the stock of housing, (4) attribute prices may change over time, and (5) a large number of sales are required before a reasonable repeat-sales sample is obtained." Donald R. Haurin and Patric H. Hendershott (1991; 260)

The fifth criticism in this quotation – the large number of sales required to obtain a reasonable data set with repeat sales– was not mentioned thus far. In section 7.4 a basic OLS repeat sales index will be constructed using the data for the town of "A" that was used earlier in Chapters 5 and 6 to show the effect of having a very small repeat sales data set.

7.4 An Example Using Data for the Town of "A"

Recall that, after deleting houses which were older than 50 years at the time of sale and also deleting observations which had land areas greater than $1200\,m^2$, we were left with 2289 sales in the 14 quarter sample period, starting in the first quarter of 2005 and ending in the second quarter of 2008. That is, we had an average of 163.5 single sales of detached dwelling units per quarter for the Dutch town of "A". A few houses were sold twice during the same quarter, and we deleted those short holds for the estimation of the repeat sales index (as they could be distressed sales). We ended up with only 85 repeat sales over the 14 quarter period. The OLS

[8] As mentioned in section 7.2, there are ways to deal with this problem but they all appear to be too crude or too complex to be used for the compilation of official statistics.

[9] In practice, this is not necessarily a big problem. A similar problem occurs when monthly scanner data are used in a CPI; a moving window of observations can be used to construct a monthly CPI component where only the incremental inflation rate for the last month is used to update the index; see Ivancic, Diewert and Fox (2011) and de Haan and van der Grient (2011).

repeat sales index computed using this small data set, labeled as P_{RS}, is plotted in Figure 7.1 along with the chained stratified sample mean Fisher index, P_{FCH}, described in Chapter 5 and the hedonic imputation Fisher index, P_{HIF}, described in Chapter 6. These three price series are listed in Table 7.1.

Figure 7.1. Repeat Sales Price Index, Chained Stratified Sample Fisher Price Index and Hedonic Imputation Fisher Price Index

Table 7.1. Repeat Sales Price Index, Chained Stratified Sample Mean Fisher Price Index and Hedonic Imputation Fisher Price Index

Quarter	P_{RS}	P_{FCH}	P_{HIF}
1	1.00000	1.00000	1.00000
2	1.00650	1.02396	1.04356
3	1.02802	1.07840	1.06746
4	1.02473	1.04081	1.03834
5	1.03995	1.04083	1.04794
6	1.04206	1.05754	1.07553
7	1.08663	1.07340	1.09460
8	1.07095	1.06706	1.06158
9	1.14474	1.08950	1.10174
10	1.15846	1.11476	1.10411
11	1.12709	1.12471	1.11430
12	1.13689	1.10483	1.10888
13	1.14903	1.10450	1.09824
14	1.12463	1.11189	1.11630

Compared to the other two price indexes, the repeat sales index turns out to be highly erratic during the second half of the sample period. In quarter 14, the repeat sales index shows a price decrease whereas the hedonic imputation and stratified sample means indexes measure a price increase. Of course we cannot draw any definitive conclusions from this simple example, but it does confirm that repeat sales methods require a large number of observations to estimate price indexes with acceptable precision.

Chapter 8: Appraisal-Based Methods

8.1 Introduction

As was mentioned in previous chapters, the matched model methodology to construct price indexes, where prices of the exact same item are compared over time, cannot be applied in the housing context. One of the reasons is the low incidence of re-sales and the resulting change in the composition of the properties sold. The repeat sales method, which was discussed in Chapter 7, attempts to deal with the quality mix problem by looking at properties that were sold more than once over the sample period. However, using only repeat-sales data could be very inefficient since all single sales data are "thrown away" and could also lead to sample selection bias.

In several countries assessed values or appraisals of properties are available, which might be useful as proxies for selling prices or, more generally, market values. In countries where they have been collected for tax purposes, appraisals will typically be available for all properties at a particular reference period. In a number of studies assessed values were used in addition to sale prices in a repeat sales framework to reduce the problem of inefficiency and the potential problem of sample selection bias. For example, Gatzlaff and Ling (1994) used sale prices as the first measure and appraisals as the second measure in a repeat "sales" regression. Clapp and Giaccotto (1998) did the reverse and used appraisals as the first and selling prices as the second measure. Both studies found that these methods produced price indexes similar to a standard repeat sales index.

The above assessed-values repeat sales methods are based on pseudo price relatives in which the appraisals do not necessarily pertain to the same period. But when assessed values for all properties are available that do relate to a single valuation period or reference date, then it will be possible to use the standard matched model methodology. For each property sold in some comparison period for which we have a sale price, a base period "price" – the assessed value – is now available also. Price relatives with a common base period – the valuation period – can then be constructed and these so-called sale price appraisal ratios can be aggregated using a standard index number formula, though some re-scaling appears necessary.

The use of a standard matched model index number formula simplifies the computation of the index because there is no need to use econometric techniques to estimate the index or to adjust for compositional change, as is the case with hedonic and repeat sales methods; see Chapters 6 and 7. Another feature of the *sale price appraisal ratio method*, or SPAR method for short, discussed in the present chapter is that it is free from revisions because there is no modeling and pooling of data involved. Thus, in contrast to the repeat sales method and the multiperiod time dummy hedonic method, previously computed price indexes are not re-estimated when new sales data becomes available.

The SPAR method has been used in New Zealand since the early 1960s and is currently also used in several European countries, notably in Denmark, the Netherlands and Sweden. Given that a few countries around the world are actually using the SPAR method, it is not surprising that there is only a small though expanding literature available. As far as we know, Bourassa, Hoesli and Sun (2006) were the first to publish a paper on this method. According to them, "the advantages and the relatively limited drawbacks of the SPAR method make it an ideal candidate for use by government agencies in developing house price indexes". Rossini and Kershaw (2006) found that the SPAR method outperformed several other methods in terms of reduced volatility of weekly index numbers. De Vries et al. (2009) reported a higher precision of monthly SPAR indexes for the Netherlands compared with monthly repeat sales indexes. Shi, Young and Hargreaves (2009) compared SPAR and repeat sales indexes for New Zealand and found a rather low correlation on a monthly basis.

When the properties are reassessed and new appraisal data become available, the SPAR index can, and probably should be, rebased. A long-term index series is obtained by "splicing" the existing and new series. Properties in the Netherlands are currently being re-valued each year, which makes it possible to construct an annually chained RPPI, where the valuation period (which is January) serves as the link month. Shi, Young and Hargreaves (2009) argued that bias can arise from frequent reassessments. De Vries et al. (2009) did not find any chain-link bias but observed that the standard error of the chained SPAR index increases each time new appraisals are introduced because an additional source of sampling error is added.

8.2 The SPAR Method in Detail

Suppose that we have samples of properties sold at our disposal for the starting or base period 0 and for comparison periods t $(t = 1,...,T)$. As in earlier chapters, the samples will be denoted by $S(0)$ and $S(t)$. In each period we know the sale prices of all sampled properties; the price of property n in period t is represented by p_n^t. As mentioned before, houses that were sold in period t were generally not sold in period 0, so there is a lack of matching. However, suppose that assessed values or appraisals are available for all properties in the housing stock, and that they relate to a single valuation period. The valuation period will serve as the base period, and the appraisal for property n will be denoted by a_n^0. Thus, for each property belonging to the period t sample $S(t)$ we know both the period t selling price p_n^t and the base period assessed value a_n^0. In other words, for all $n \in S(t)$ we can establish a price relative – a *sale price appraisal ratio* – p_n^t / a_n^0, which can be used in a matched model framework to compute an RPPI.

Although it would be possible to construct geometric appraisal-based indexes, we will focus here on arithmetic indexes as these seem to be more appropriate in the housing context. The arithmetic appraisal-based index can be defined as

$$P_{AP}^{0t} = \frac{\sum_{n \in S(t)} 1 p_n^t}{\sum_{n \in S(t)} 1 a_n^0} = \sum_{n \in S(t)} w_n^0(t) \left(\frac{p_n^t}{a_n^0} \right). \tag{8.1}$$

Expression (8.1) describes a *Paasche-type* index because we are using the comparison period sample $S(t)$ in both the numerator and the denominator. The quantities are equal to 1 as every property is basically a unique good. The construction of a Laspeyres-type price index would be problematic or even impossible: period *t* price information for dwelling units belonging to the base period sample $S(0)$ is only available for those few units, if any, that were resold in period *t*. This means that the construction of a Fisher-type index will not be feasible either. As shown by the second expression, (8.1) can be written as a value-weighted average of the sale price appraisal ratios p_n^t / a_n^0, where the weights $w_n^0(t) = a_n^0 / \sum_{n \in S(t)} a_n^0$ reflect the base period assessed value shares with respect to the sample $S(t)$.

The appraisal-based Paasche-type index, P_{AP}^{0t}, given by (8.1) is obviously a matched model index. Accordingly, there is no compositional change to account for when comparing period *t* directly with period 0. However, as there is generally no overlap, the samples $S(t)$ in periods $t = 1,...,T$ will be completely different and compositional change will be present from period to period. Those period to period sample mix changes cannot be adjusted for, which suggests that short-term volatility will most likely occur. This feature is not unique to the appraisal-based index; we would expect to observe more or less the same for the Paasche-type hedonic imputation indexes discussed in Chapter 6. The similarity with the imputation Paasche index will be addressed in section 8.3.

The appraisal-based price index (8.1) does not make use of the observed selling prices in the base period. As a result, the index will differ from 1 in the base period, which is problematic. However, this problem can easily be resolved by normalizing the indexes by dividing them by the base period value. We then obtain the following arithmetic *SPAR index*:

$$P_{SPAR}^{0t} = \frac{\sum_{n \in S(t)} p_n^t}{\sum_{n \in S(t)} a_n^0} \left[\frac{\sum_{n \in S(0)} p_n^0}{\sum_{n \in S(0)} a_n^0} \right]^{-1} = \frac{\sum_{n \in S(t)} p_n^t / N(t)}{\sum_{n \in S(0)} p_n^0 / N(0)} \left[\frac{\sum_{n \in S(0)} a_n^0 / N(0)}{\sum_{n \in S(t)} a_n^0 / N(t)} \right], \tag{8.2}$$

where $N(0)$ and $N(t)$ denote the number of properties sold in periods 0 and *t* (the respective sample sizes).

The second expression on the right-hand side of (8.2) writes the SPAR index as the product of the ratio of sample means and a bracketed factor. Since the SPAR method is a matched model method (with respect to period 0), the bracketed factor adjusts the ratio of sample means for compositional changes occurring between each period *t* and the base period 0. So, while short-term volatility is likely to be present due to period to period mix changes, the SPAR method is expected to exhibit less volatility than the ratio of sample means.

The arithmetic SPAR index can be interpreted as a proxy for a sales based Paasche RPPI[1]. But many countries, including EU member states, are typically aiming at a Laspeyres index rather than a Paasche index. Stratification could be used as a means to approximate this target index while using the SPAR method. The SPAR (Paasche) indexes at the stratum level will then be aggregated using base period expenditure share weights to obtain the overall "Laspeyres-type" index. The RPPI in the Netherlands is an example of such a stratified SPAR approach, where region and type of house are used as stratification variables. The index is compiled monthly and published jointly with the Dutch Land Registry Office. Stratification might also help to account for any systematic differences between appraisals and market values across regions or different types of houses (de Vries et al., 2009; de Haan, van der Wal and de Vries, 2009).

The SPAR index can alternatively be interpreted as a sample estimator of a stock RPPI. If in each period the properties sold are viewed as random samples from the base period housing stock, then the SPAR index is an estimator of the Laspeyres stock RPPI. Properties sold that were added to the stock after the base period should in this case be excluded.[2] As mentioned in earlier chapters, the sample of houses sold may not be representative of the total stock so that sample selection bias could arise. Stratification will again be a helpful tool to mitigate this problem.

8.3 Methodological and Practical Issues

Quality Change

Since the appraisals relate to the base period, in general the properties will have been valued at their base period characteristics. But for the SPAR index (8.1) to be a constant quality price index, the appraisals should be evaluated at characteristics of the comparison period. Thus, if housing characteristics change over time, the SPAR method will not adjust for those changes, similar to the repeat sales method. This is an important drawback.

Yet in practice there could be some implicit adjustment for quality changes. In the case of the New Zealand SPAR index, Bourassa, Hoesli and Sun (2006) note that "the base appraisal is

---

[1] Administrative data sets, particularly those from the land registry, typically contain all sales (excluding newly-built properties) in each period. From a sales point of view there is no sampling involved. In this interpretation, the SPAR index has no sampling error, but it does have error due to the use of appraisals, which are estimates of the "true" market values.

[2] It may seem that properties which are new to the stock cannot even be used because the necessary appraisals are lacking. However, this depends on the appraisal system. If former rental houses have been sold and are thus added to the stock of owner occupied housing, then they will have a base period appraisal value if rental houses are also assessed. Moreover, if property taxes are uniformly based on period 0 valuations for a number of years, then the authorities would need those values for newly built houses as well. The difficulty is of course that the authorities would have to "invent" an assessed value for a new house in period 0, even if it did not exist in that period. Such assessments might be problematic and hence should probably be excluded from the computation of the index.

adjusted for subsequent improvements to the property that require a building permit". If this is done in real time, adjustments for major quality improvements will indeed be made. However, apart from the fact that not all property improvements require a building permit, it is unlikely that these adjustments adequately deal with the *net effect* of improvements and depreciation of the structures.

In the Netherlands there may also be some implicit quality adjustment in the SPAR index. The assessments are typically carried out some time after the official appraisal reference month and may take into account major improvements to the properties. Furthermore, as mentioned above, the assessments are nowadays performed every year. Annual chaining by itself could alleviate the problem of quality change if the updated appraisals properly account for changes in the characteristics. Of course this will depend on the exact way the properties are valued, which may not be known to the index compilers.

Quality of the Assessment Information

Abstracting now from quality adjustment issues, the SPAR method is obviously dependent on the quality of the assessment information. There are three broad ways in which assessments of (not traded) properties can be carried out: by using hedonic regression, by comparing them to similar traded properties, and by expert judgment. The methods used differ among countries and sometimes even within a particular country. In various countries, private companies are engaged in mass appraisal. Although the details of the methods used are often not publicly available, some of those companies appear to combine hedonic regression with local market information or expert judgment.

Bourassa, Hoesli and Sun (2006) noted that the appraisals in New Zealand are derived from hedonic regressions, but unfortunately they did not present the exact method. In Chapter 6 it was explained that there are different hedonic approaches and that the predicted prices – in this case the appraisals – depend on the type of data used and the number of observations, the specified functional form, the variables included and other choices made. Thus, even though hedonic regression is the least arbitrary of the three assessment methods mentioned above, there can still be a lot of uncertainty and error involved, which has an unknown impact on the sale price appraisal ratios and the resulting SPAR index.

The use of comparable properties seems to be widespread. Chinloy, Cho and Megbolugbe (1997) compared a sample of U.S. private sector appraisals to selling prices. They suspected that the reliance on a relatively small number of comparable houses leads to more volatility than can be observed in market-wide selling prices. More importantly perhaps, they found that appraisals exceeded sale prices in approximately 60 percent of the cases, leading to an average upward bias of two percent.

In countries where official assessments are designed for property taxation purposes, like in the Netherlands, the assessed values may not to be too far off the mark since the government has an incentive to make the assessments as large as possible in order to maximize tax revenue while taxpayers have the opposite incentive to have the assessments as small as possible. In the Netherlands the municipalities are responsible for making assessments. The methods used differ across the municipalities. Some of them, for example the capital city of Amsterdam, use the comparable house method whereas others apparently use some kind of hedonic regression method. De Vries et al. (2009) argued that Dutch authorities may in fact have an incentive to make the assessments not too high to avoid court procedures because households who feel the appraised value is too high can lodge an appeal.

Other Issues

The advantage of the SPAR method as compared to hedonic regression methods is that data on only a few property attributes is needed: appraised values (relating to a common reference period), possibly some stratification variables, and addresses to merge the data files in case the selling prices and appraisals come from different sources. In the Netherlands, for example, the transaction prices and a limited number of stratification variables are recorded by the Land Registry whereas the appraisals are from a second administrative data source. It is well known that merging data files by address can be difficult, although in the Netherlands this seems to be a minor issue.

Data cleaning is another important practical issue. The SPAR method is dependent on the quality of the appraisals. Some of the sale price appraisal ratios might be found implausible, perhaps because the appraisals are deemed "wrong", and deleted from the data set.[3] Deleting erroneous observations, such as obvious entry errors, is good practice. However, a cautious approach is called for as deleting price relatives can lead to biased results. At least a rule for deleting outliers should be explicitly formulated to inform users.

8.4 A Regression-Based Imputation Interpretation

In this section we will show that the SPAR method is essentially an imputations approach in which the "missing" base period prices are estimated from a linear regression of selling prices on appraisals. Recall first that the base period prices of the properties belonging to the period t sample $S(t)$ cannot be observed directly because those properties were generally not traded in period 0. We can try to estimate the "missing" prices to obtain the imputation Paasche price index

[3] The example for the town of "A" in section 8.6 shows that the removal of a relatively low number of outliers can have a substantial effect on the SPAR index.

$$P_{IP}^{0t} = \frac{\sum_{n \in S(t)} 1 p_n^t}{\sum_{n \in S(t)} 1 \hat{p}_n^0(t)}. \tag{8.3}$$

The imputed value $\hat{p}_n^0(t)$ in (8.3) should predict the period 0 price for property *n*, evaluated at its period *t* characteristics. Keeping the (quantities of the) characteristics fixed is necessary to adjust for quality change. The use of *hedonic* imputation was discussed in Chapter 6. Hedonic regression models explain the selling price of a property in terms of a set of price-determining characteristics that relate to the structure and the location. This section addresses a different type of regression-based imputation.

Consider the following two-variable regression model for the base period:

$$p_n^0 = \beta_0 + \beta_1 a_n^0 + \varepsilon_n^0. \tag{8.4}$$

Equation (8.4) is a simple *descriptive* model where selling prices are regressed on appraisals. We assume that this model is estimated by Ordinary Least Squares (OLS) on the data of the base period sample $S(0)$. The predicted prices for $n \in S(0)$ are

$$\hat{p}_n^0 = \hat{\beta}_0 + \hat{\beta}_1 a_n^0, \tag{8.5}$$

where $\hat{\beta}_0$ is the estimated intercept term and $\hat{\beta}_1$ the estimated slope coefficient. We expect to find $\hat{\beta}_0 \cong 0$ and $\hat{\beta}_1 \cong 1$ if the appraisal system works well.[4] Equation (8.5) will be used below to predict the "missing prices" in the denominator of the imputation Paasche index (8.3).

For convenience we first rewrite (8.3) as

$$P_{IP}^{0t} = \frac{\sum_{n \in S(t)} p_n^t / N(t)}{\sum_{n \in S(0)} p_n^0 / N(0)} \frac{\sum_{n \in S(0)} p_n^0 / N(0)}{\sum_{n \in S(t)} \hat{p}_n^0(t) / N(t)} = \frac{\sum_{n \in S(t)} p_n^t / N(t)}{\sum_{n \in S(0)} p_n^0 / N(0)} \frac{\sum_{n \in S(0)} \hat{p}_n^0 / N(0)}{\sum_{n \in S(t)} \hat{p}_n^0(t) / N(t)}. \tag{8.6}$$

In the second step of (8.6) we have used $\sum_{n \in S(0)} p_n^0 / N(0) = \sum_{n \in S(0)} \hat{p}_n^0 / N(0)$, which holds true because the OLS regression residuals sum to zero. The first problem we face is that the housing characteristics should be kept fixed when predicting the base period prices $\hat{p}_n^0(t)$ for $n \in S(t)$. This is obviously not possible using equation (8.5). Thus, the *first assumption* is that of no quality change, and we accordingly replace $\hat{p}_n^0(t)$ in (8.6) by $\hat{p}_n^0(0) = \hat{p}_n^0$. Using (8.5) for both $n \in S(0)$ and $n \in S(t)$, equation (8.6) becomes

$$P_{IP}^{0t} = \frac{\sum_{n \in S(t)} p_n^t / N(t)}{\sum_{n \in S(0)} p_n^0 / N(0)} \left[\frac{\hat{\beta}_0 + \hat{\beta}_1 \sum_{n \in S(0)} a_n^0 / N(0)}{\hat{\beta}_0 + \hat{\beta}_1 \sum_{n \in S(t)} a_n^0 / N(t)} \right]. \tag{8.7}$$

[4] If the selling prices would be used as official valuations, then of course the values 0 and 1 would exactly hold and we would find a perfect fit of (8.4) to the period 0 data.

Notice that if $\hat{\beta}_0 = 0$, that is, if the regression line passes through the origin, (8.7) simplifies to the SPAR index (8.2), irrespective of the slope coefficient $\hat{\beta}_1$. So, if the aim is to estimate an imputation Paasche index, the *second assumption* underlying the SPAR method seems to be that the intercept term $\hat{\beta}_0$ is negligible.

The *third assumption* is that equation (8.5) holds for $n \in S(t)$: the linear relationship between base period selling prices and appraisals postulated and estimated for the properties actually sold during the valuation or base period 0 (for $n \in S(0)$) is assumed to hold also for properties that were not sold. But this is a very restrictive assumption. While the linear relation can be tested for $n \in S(0)$,[5] it would be difficult if not impossible to test it for $n \in S(t)$ as the selling prices are "missing". The presence of appraisal bias, in the sense that the appraisals over- or underestimate the unknown market values (the prices at which the properties would have been sold), can bias the SPAR index. Bias in the SPAR index will particularly arise if the "true" value of β_1 for $n \in S(t)$ would be very different from β_1 for $n \in S(0)$.

In this section we focused on the SPAR index as a sales RPPI. A related approach, where the appraisals serve as auxiliary information in a "generalized regression" (GREG) framework in order to estimate a stock based RPPI, was described by de Haan (2010b). The GREG method uses population information on the appraisals instead of sample information. He showed that the SPAR index is a straightforward estimator of the GREG stock based index which, when applied to Dutch data, turned out to be almost as efficient.

8.5 Main Advantages and Disadvantages

The merits of the SPAR method are listed below. The main advantages are:

- The SPAR method is essentially based on the standard matched model methodology and links up with traditional index number theory.
- The method is computationally simple.
- Information on housing characteristics is not required in order to implement this method; the only information required is data on sale prices and appraisals. In some countries the data is available from administrative sources such as the land registry, and usually covers all transactions (for resold properties).
- This method uses much more data than the repeat sales method and hence there are fewer problems due to sparse data. In particular, sample selection bias is likely to be smaller.

[5] Van der Wal, ter Steege and Kroese (2006) and de Vries et al. (2009) compared Dutch government appraisals to selling prices. In the latter study the linear relationship (8.4) was explicitly tested (for the properties traded in the valuation month) for various valuation months. It turned out that the constant term was indeed very small and that the slope coefficient did not significantly differ from 1.

Also, the SPAR method does not suffer from revision of previously calculated figures when new data becomes available.

- Conditional on the data cleaning rules, the SPAR method is reproducible.

The main disadvantages of the SPAR method are:

- The method cannot deal adequately with quality changes (major repairs or renovations and depreciation) of the dwelling units.[6]
- The SPAR method is dependent on the quality of the base period assessment information. The exact way the valuations are carried out may not always be clear and has an unknown impact on the results.
- The method cannot decompose the overall property price index into land and structures components.[7]

8.6 An Example on Data for the Town of "A"

Using the data set for the town of "A", which was described in Chapter 5, a SPAR index was computed. Recall that this data set contained sales of detached houses for 14 quarters, starting in the first quarter of 2005 and ending in the second quarter of 2008. After some data cleaning – in particular deleting houses that were older than 50 years at the time of sale were – a total of 2289 sales remained.

To compute SPAR index numbers we also need assessed values for the properties sold. Our appraisal data relate to the first quarter (i.e., January) of 2005. Matching the sales data set and the appraisal data set was quite successful; 99.3% of the selling prices could be matched with the corresponding appraisals; i.e. for only 15 observations we could not find an appraisal, so these were deleted. The resulting SPAR index, P_{SPAR}, is plotted in Figure 8.1 and listed in Table 8.1, along with the hedonic imputation Fisher index, P_{HIF}, described in Chapter 6, and the repeat sales index, P_{RS}, estimated in Chapter 7. The trend of P_{SPAR} is very similar to that of P_{HIF}, but P_{SPAR} is slightly more volatile.

A potential drawback of the SPAR method is that is entirely dependent on the accuracy of the appraisal data. An inspection of the distribution of the sale price appraisal ratios indicated a

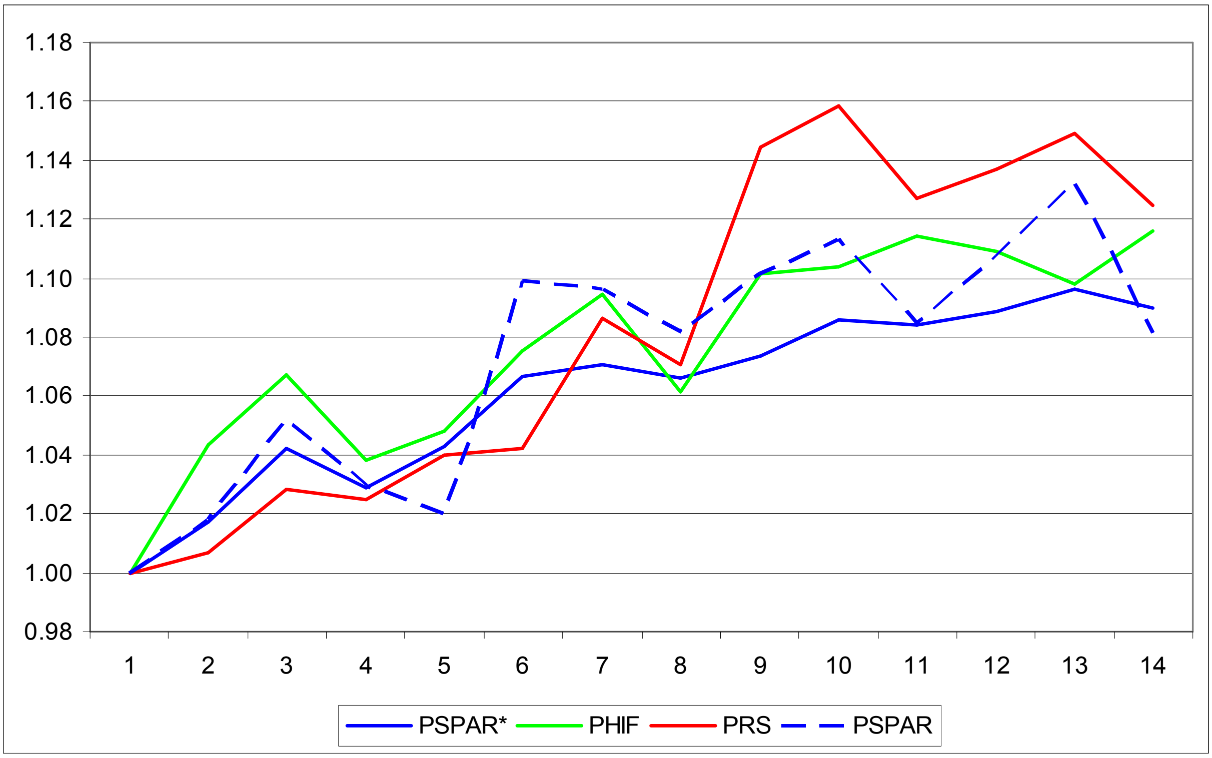
[6] In countries where the assessments provide separate information on the value of the structures and the value of the land, the SPAR index could in principle be adjusted by using exogenous information on the net depreciation of houses of the type being considered.

[7] If fresh property assessment information appeared every month or quarter, this information could be used to form separate price indexes for both land and structures, provided that the assessments decomposed the total assessed value of the property into land and structures components. Unfortunately, official assessments generally are made only once a year or once every few years. This low frequency information could however be used to check the land and structures price indexes generated by hedonic regression methods.

number of big outliers. Specifically, there were several observations with very high sale price appraisal ratios (up to 10.5), in most instances as a result of unusually low appraised values. It is most likely that a significant proportion of these outliers were recording errors. Hence, we decided to delete the biggest outliers. Following Statistics Netherlands data cleaning methods at the time, based on the distribution of the natural logarithm of the sale price appraisal ratios, 26 observations were removed for which the log of price ratio differed more than 5 standard deviations from the mean.[8] We ended up with 2248 observations.

The improved SPAR index, labeled P_{SPAR*}, computed on the cleaned data set is also shown in Figure 8.1 and Table 8.1. As can be seen, cleaning of the data had a substantial impact on the result: P_{SPAR*} is much less volatile than the index P_{SPAR} that was computed on the initial data set. The trend was also affected: P_{SPAR*} is generally lower than P_{SPAR} due to the fact that most of the deleted observations had unusually high sale price appraisal ratios. Figure 8.1 confirms that – using a relatively small data set which covers a short time period – the SPAR method generates more credible results than the standard repeat sales method, especially after cleaning the data.

Figure 8.1. SPAR Index, Hedonic Imputation Fisher Price Index and Repeat Sales Index



[8] As a first step in the data cleaning procedure, Statistics Netherlands removed all properties with selling prices or appraisals below 10,000 or above 5,000,000 Euros. In our data set, however, there were no such properties. Note that Statistics Netherlands recently changed the outlier detection and removal procedures.

Table 8.1. SPAR Index, Hedonic Imputation Fisher Price Index and Repeat Sales Index

Quarter	P_{SPAR}	P_{HIF}	P_{RS}	P_{SPAR*}
1	1.00000	1.00000	1.00000	1.00000
2	1.01769	1.04356	1.00650	1.01693
3	1.05196	1.06746	1.02802	1.04204
4	1.02958	1.03834	1.02473	1.02883
5	1.02040	1.04794	1.03995	1.04273
6	1.09938	1.07553	1.04206	1.06655
7	1.09635	1.09460	1.08663	1.07076
8	1.08169	1.06158	1.07095	1.06604
9	1.10173	1.10174	1.14474	1.07378
10	1.11333	1.10411	1.15846	1.08609
11	1.08477	1.11430	1.12709	1.08396
12	1.10742	1.10888	1.13689	1.08869
13	1.13206	1.09824	1.14903	1.09642
14	1.08132	1.11630	1.12463	1.09003

A comparison of P_{SPAR*} with the hedonic imputation Fisher index P_{HIF} reveals that in several periods, for example in the last four quarters, the price changes according to the two methods are in opposite directions. Also, P_{SPAR*} is generally lower than P_{HIF}; at the end of the sample period, in quarter 14, the difference amounts to 0.026 index points. At first sight this seems to suggest that P_{SPAR*} has a downward bias. However, a difference of the same magnitude (0.027 points) is already found in quarter 2. So if we had normalized both series to equal 1 in quarter 2, the two methods would have produced approximately the same index value in quarter 14. This is an illustration of a general *starting problem* encountered when comparing volatile time series: the choice of starting or base period affects the average difference during the sample period.

Chapter 9: Decomposing an RPPI into Land and Structures Components

9.1 Introduction

In Chapter 4 of this Handbook it was mentioned that for national accounts and CPI purposes, it will be useful or necessary to have a decomposition of the residential property price index (RPPI) into two components: a quality adjusted price index for structures and a price index for the land on which the house is built. The present chapter outlines how hedonic regression can be utilized to derive such a decomposition. Hedonic regression methods were discussed in Chapter 6.

Some economic reasoning will be helpful to derive an appropriate hedonic regression model. Think of a property developer who is planning to build a structure on a particular property. He or she will likely determine the selling price of the property after the structure is completed by first calculating the total expected cost. This cost will be equal to the floor space area of the structure, say S square meters, times the building cost per square meter, γ say, plus the cost of the land, which will be equal to the cost per square meter, β say, times the area of the land site, L. We follow a *cost of production approach* to modeling the property price. That is, the functional form for the hedonic price function is assumed to be determined by the supply side of the market, i.e., by independent contractors.[1]

Now consider a sample of properties of the same general type, which have structure areas S_n^t and land areas L_n^t in period t for $n = 1,...,N(t)$; the prices p_n^t are equal to costs of the above types plus error terms ε_n^t which are assumed to have means 0. This gives rise to the following hedonic regression model for period t where α^t and β^t are the parameters to be estimated:[2]

$$p_n^t = \beta^t L_n^t + \gamma^t S_n^t + \varepsilon_n^t; \qquad t = 1,...,T; \; n = 1,...,N(t). \qquad (9.1)$$

The quantity of land L_n^t and the quantity of structures S_n^t associated with the sale of property n in period t are the only two property characteristics included in this very simple model; the corresponding prices in period t are the price of a square meter of land β^t and the price of a

[1] McMillen (2003) discusses a Cobb Douglas *demand side* model. On identification issues in hedonic regression models, see Rosen (1974).

[2] Following Muth (1971), Thorsnes (1997; 101) has a related cost of production model. He assumed that the value of the property under consideration in period t, p^t, is equal to the price of housing output in period t, ρ^t, times the quantity of housing output $H(L,K)$ where the production function H is a CES function. Thus Thorsnes assumed that $p^t = \rho^t H(L,K) = \rho^t [\alpha L^\sigma + \beta K^\sigma]^{1/\sigma}$ where ρ^t, σ, α and β are parameters, L is the lot size of the property and K is the amount of structures capital (in constant quality units). Our problem with this model is that there is only one independent time parameter ρ^t whereas our model has two, β^t and γ^t for each t, which allow the price of land and structures to vary freely between periods.

square meter of structure floor space γ^t. Separate linear regressions of the form (9.1) can be performed for each time period *t* in the sample.

The "builder's model" (9.1) essentially relates to newly-built dwellings. To make it applicable to existing or resold houses we should account for the fact that older structures will be worth less than newer structures due to *depreciation* of the structures. Thus, information on the age of the structure will be needed. Section 9.2 shows how depreciation can be incorporated into the model, similar to what was done in the examples for the town of "A" presented in Chapter 6. It will also be shown how additional land and structures characteristics can be included as explanatory variables.

9.2 Accounting for Depreciation and Additional Characteristics

Depreciation

Suppose that in addition to information on the selling price of property *n* at time period *t*, p_n^t, the land area of the property, L_n^t, and the structure area, S_n^t, information on the age of the structure at time *t*, say A_n^t, is available. If straight line depreciation is assumed, the following model is a straightforward extension of (9.1) to include "existing" houses:

$$p_n^t = \beta^t L_n^t + \gamma^t (1 - \delta A_n^t) S_n^t + \varepsilon_n^t, \qquad t = 1,...,T;\ n = 1,...,N(t), \qquad (9.2)$$

where the parameter δ reflects the (straight line) *depreciation rate* as the structure ages one additional period. If structure age is measured in years, δ will probably be between 0.5% and 2%. This will be an underestimate of "true" depreciation because it will not account for major renovations or additions to the structure. The estimated straight line depreciation rate in (9.2) should therefore be interpreted as a *net depreciation rate*; i.e., a gross depreciation rate less the rate of renovations and additions to the structure. Model (9.2) will not work for very old structures since, if they are still in use, they will likely have been extensively renovated.[3]

Notice that (9.2) is a *nonlinear* regression model whereas (9.1) is a linear regression model.[4] Because the depreciation parameter δ is regarded as fixed over time, (9.2) would have to be estimated as one nonlinear regression over all time periods in the sample, whereas model (9.1) can be run as a period by period linear regression. The period *t* price of land in model (9.2) will be the estimate for the parameter β^t and the price of a unit of a newly built structure for period *t* will be the estimate for γ^t. The period *t* quantity of land for property *n* is L_n^t and the period *t* quantity of structures for property *n*, expressed in equivalent units of a new structure, is $(1 - \delta A_n^t) S_n^t$, where S_n^t is the floor space area of property *n* in period *t*.

---

[3] See for example Meese and Wallace (1991; 320) who found that the age variable in their hedonic regression model had the wrong sign.

[4] The model defined by (9.2) can be converted into a linear regression model.

Expensive properties probably have relatively large absolute errors compared to inexpensive properties, so it might be better to assume multiplicative rather than additive errors. However, we prefer an additive model specification because the purpose is to decompose the aggregate value of housing (in the sample of properties transacted) into the *sum* of structures and land components; the use of additive errors facilitates this decomposition. When there is evidence of heteroskedasticity, weighted regressions can be considered. Several researchers suggested hedonic regression models that lead to additive decompositions of a property price into land and structures components.[5]

There is a potential problem with the above builder's model, namely multicollinearity. Large structures are generally built on large plots of land, so that S_n^t and L_n^t could be highly collinear (i.e., the land-structure ratios L_n^t / S_n^t could be centered around a constant). This could give rise to unstable estimates of the quality adjusted prices β^t and γ^t for land and structures. As will be seen in section 9.5, using data for the Dutch town of "A", this problem of multicollinearity and instability does indeed occur. In general, multicollinearity is not a major problem if the goal is to produce an overall house price index, but it is problematic if the goal is to produce separate price indexes for land and structures components. Some possible methods for overcoming the multicollinearity problem will be suggested in sections 9.6 and 9.7.

The hedonic regression model (9.2) has the implication that the parameters would have to be re-estimated whenever the data for a new period became available. To overcome this problem, a "rolling window" approach could be applied. A suitable window length T would be chosen,[6] the model defined by (9.2) or (9.3) would be estimated using the data for the last T periods, and the existing series for price of land and for price of structures would be updated using the chain link factors β^T / β^{T-1} and γ^T / γ^{T-1}. This approach will be illustrated in section 9.8.

Adding More Characteristics

The above basic nonlinear hedonic regression framework can be generalized to encompass the traditional array of characteristics used in real estate hedonic regressions. Suppose that we can associate with each property n traded in period t a list of K characteristics $X_{n1}^t, X_{n2}^t, ..., X_{nK}^t$ that are price determining characteristics for the land on which the structure was built and a similar list of M characteristics $Y_{n1}^t, Y_{n2}^t, ..., Y_{nM}^t$ that are price determining characteristics for the type of structure. The following equations generalize (9.2) to the present setup:[7]

[5] See Clapp (1980), Francke and Vos (2004), Gyourko and Saiz (2004), Bostic, Longhofer and Redfearn (2007), Diewert (2007), Francke (2008), Koev and Santos Silva (2008), Statistics Portugal (2009), Diewert, de Haan and Hendriks (2010) (2011) and Diewert (2010).

[6] The model becomes a modified adjacent period hedonic regression model for $T = 2$.

[7] This generalization was suggested by Diewert (2007).

$$p_n^t = \beta^t\left[1+\sum_{k=1}^{K}X_{nk}^t\eta_k\right]L_n^t + \gamma^t(1-\delta A_n^t)\left[1+\sum_{m=1}^{M}Y_{nm}^t\lambda_m\right]S_n^t + \varepsilon_n^t,\; t=1,\ldots,T\;;\; n=1,\ldots,N(t)\,, \quad (9.3)$$

where the parameters to be estimated are now the *K* quality of land parameters, $\eta_1,\ldots,\eta_K$, the *M* quality of structures parameters, $\lambda_1,\ldots,\lambda_M$, the period *t* quality adjusted price for land β^t and the period *t* quality adjusted price for structures γ^t. The quality adjusted amount of land, L_n^{t*}, and the corresponding quality adjusted amount of structures, S_n^{t*}, for property *n* in period *t* are defined as follows:

$$L_n^{t*} \equiv \left[1+\sum_{k=1}^{K}X_{nk}^t\eta_k\right]L_n^t\;;\; S_n^{t*} \equiv \left[1+\sum_{m=1}^{M}Y_{nm}^t\lambda_m\right]S_n^t \quad , \qquad t=1,\ldots,T\;;\; n=1,\ldots,N(t)\,. \qquad (9.4)$$

To illustrate how *X* and *Y* variables can be formed, consider the list of explanatory variables in the hedonic housing regression model reported by Li, Prud'homme and Yu (2006; 23). The following variables in their list of explanatory variables can be viewed as variables that affect structures quality; i.e., they are *Y* type variables: number of bedrooms, number of bathrooms, number of garages, number of fireplaces, age of the unit, age squared of the unit, exterior finish is brick or not, dummy variable for new units, unit has hardwood floors or not, heating fuel is natural gas or not, unit has a patio or not, unit has a central built in vacuum cleaning system or not, unit has an indoor or outdoor swimming pool or not, unit has a hot tub unit or not, unit has a sauna or not, and unit has air conditioning or not. The following variables can be assumed to affect the quality of the land; i.e., they are *X* type location variables: unit is at the intersection of two streets or not (corner lot or not), unit is at a cul-de-sac or not, shopping center is nearby or not, and various suburb location dummy variables.

Equations (9.3) and (9.4) show how the quality adjusted amounts of land and structures would be calculated if the goal is to construct price indexes for the *sales* of properties of the type that are included in the hedonic regression model. If the goal is to construct price indexes for the *stock* of properties of the type included in the regression, then the construction of appropriate weights becomes more complex. These weighting problems will be discussed in section 9.3 and illustrated in section 9.9.

9.3 Aggregation and Weighting Issues: Indexes for Sales versus Stocks of Housing

As was explained in Chapter 6, the construction of an RPPI for the sales of property using standard hedonic regression techniques is fairly straightforward. Typically, a separate hedonic regression of the type defined by (9.3) will be run for each locality or region in a country.[8]

[8] Separate hedonic regressions may also be run for different types of property as well as for different locations. However, cost considerations may mean that a comprehensive system of regressions covering all properties in the country cannot be implemented so that there will only be a sample of representative hedonic regressions. The aggregation issues in the sampling case are too complex to be considered here; the exact details for constructing a national index would depend on the nature of the sampling design.

Recall that once a particular regression has been run, period t quality adjusted prices for land, P_L^t, and for structures, P_S^t, for the region under consideration can be defined in terms of the estimated parameters for the model as follows:

$$P_L^t \equiv \beta^t ; \qquad t = 1,...,T ; \qquad (9.5)$$

$$P_S^t \equiv \gamma^t ; \qquad t = 1,...,T . \qquad (9.6)$$

The corresponding quality adjusted quantities of land and structures for the region, say Q_L^t and Q_S^t can also be defined in terms of the estimated parameters using definitions (9.4) above as follows:

$$Q_L^t \equiv \sum_{n=1}^{N(t)} L_n^{t*} = \sum_{n=1}^{N(t)} \left[1 + \sum_{k=1}^{K} X_{nk}^t \eta_k \right] L_n^t ; \qquad t = 1,...,T ; \qquad (9.7)$$

$$Q_S^t \equiv \sum_{n=1}^{N(t)} S_n^{t*} = \sum_{n=1}^{N(t)} \left[1 + \sum_{m=1}^{M} Y_{nm}^t \lambda_m \right] S_n^t ; \qquad t = 1,...,T . \qquad (9.8)$$

If hedonic regressions, for say R regions, of the type defined by (9.3) have been run for the T periods of data, then the algebra associated with (9.5)-(9.8) can be repeated for each region r. Denote the resulting prices and quantities for region r that are the counterparts to (9.5)-(9.8) by P_{Lr}^t, P_{Sr}^t, Q_{Lr}^t and Q_{Sr}^t for $r = 1,...,R$ and $t = 1,...,T$. Now Fisher (sales) RPPIs for land can be constructed using the regional price and quantity data for land, $P_L^t \equiv [P_{L1}^t,...,P_{LR}^t]$ and $Q_L^t \equiv [Q_{L1}^t,...,Q_{LR}^t]$, for each time period t $(t = 1,...,T)$. Similarly, Fisher (sales) RPPIs for structures can be constructed using the price and quantity data for structures in each period t, $P_S^t \equiv [P_{S1}^t,...,P_{SR}^t]$ and $Q_S^t \equiv [Q_{S1}^t,...,Q_{SR}^t]$, for $t = 1,...,T$.[9]

As was the case with stratification methods, it is now necessary to consider how to construct a RPPI for the *stock* of residential properties when hedonic regression methods are used. The period t hedonic cell prices P_{Lr}^t and P_{Sr}^t defined by the region r counterparts to (9.5) and (9.6) can still be used as cell prices to construct stock price indexes for land and structures, but the counterpart quantities Q_{Lr}^t and Q_{Sr}^t defined by (9.7) and (9.8) are no longer appropriate; these quantities need to be replaced by estimates that apply to the *total stock of dwelling units* in the region (or some other reference population) for regression r at time t, say Q_{Lr}^{t*} and Q_{Sr}^{t*}, for $r = 1,...,R$. Thus, the counterpart summations in (9.7) and (9.8) are now taken over the entire stock of dwellings in region r in period t instead of just the dwelling units that were sold in period t. Period t information on the quantity of land L_{nr}^t for every unit n in the region that is in scope for the hedonic regression model m is now required, along with the accompanying characteristics information X_{nrk}^t for every land characteristic k, as well as data on the quantity of the structures S_{nr}^t, along with the accompanying characteristics information Y_{nrm}^t for every

[9] As was the case for stratification methods, fixed base or chained indexes could be constructed. Rolling window hedonic regressions could also be run. The rolling window approach will be explained in section 9.8.

structures characteristic *m*. With these new population quantity weights, the rest of the details of the index construction are the same as was the case for the sales RPPI.

In order to construct appropriate period *t* population stock weights, it will be necessary for the country to have census information on the housing stock with enough details on each dwelling unit in the stock so that the required information on the quantity of land and structures and the accompanying characteristics can be calculated. If information on new house construction (plus the required characteristics data) and on demolitions is available in a timely manner, the census information can be updated and period *t* estimates for the constant quality amounts of land and structures, the Q_{Lr}^{t*} and Q_{Sr}^{t*}, can be approximated in a timely manner. Hence, stock RPPIs for land and structures can be constructed using Fisher indexes, as was the case for the sales RPPI. If timely data on new construction and demolitions is unavailable, it may only be possible to construct fixed base Laspeyres type price indexes using the quantity weights from the last available housing census.

If census information is not available at all (or if data on the characteristics of the dwelling units is missing), it still may be possible to approximate RPPIs for land and structures using hedonic regression techniques. If characteristics data on the residential properties that are sold in each period is stored over a large period of time, an approximate distribution of dwelling units by type can be constructed. This information may then be used to approximate a stock based RPPI in the manner explained above. An example using this technique will be given in section 9.9.

9.4 Main Advantages and Disadvantages

This section summarizes the main advantages and disadvantages of using hedonic regression methods to construct an RPPI for land and structure components. The main advantages are:

- If the list of available property characteristics is sufficiently detailed, the method adjusts for both sample mix changes and quality changes of the individual houses.
- Price indexes can be constructed for different types of dwellings and locations through a proper stratification of the sample. Stratification has a number of additional advantages.
- The method is probably the most efficient method for making use of the available data.
- The method is virtually the only method that can be used to decompose the overall price index into land and structures components.

The main disadvantages of the hedonic regression approach explained in this chapter are:

- The method is data intensive since it requires data on all relevant property characteristics (in particular, the age, the type and the location of the properties in the sample as well as information on the structure and lot size) so it is relatively expensive to implement.

- The method may not lead to reasonable results due to multicollinearity problems.
- While the method is essentially reproducible, different choices can be made regarding the set of characteristics entered into the regression, the functional form for the model, the stochastic specification, possible transformations of the dependent variable, etc., which could lead to varying estimates of overall price change.
- The general idea of the hedonic method is easily understood but some of the technicalities may not be easy to explain to users.

9.5 Application on Data for the Town of "A": Preliminary Approaches

The general techniques explained in this chapter will now be illustrated using the data set for the Dutch town of "A", which was described at the end of Chapter 5. We have data on sales of detached dwellings for 14 quarters, starting in the first quarter of 2005. Recall the notation used above and in Chapters 5 and 6: there were $N(t)$ sales of detached houses in quarter *t*, where p_n^t is the selling price of house *n*. There is information on three characteristics: area of the plot in square meters, L_n^t; floor space area of the structure in square meters, S_n^t; and age in decades of house *n* in period *t*, A_n^t.

The Simple Case

The simple hedonic regression model defined by (9.2) will be estimated on this data set and is repeated here for convenience:

$$p_n^t = \beta^t L_n^t + \gamma^t (1 - \delta A_n^t) S_n^t + \varepsilon_n^t, \qquad t = 1,\ldots,14\,;\; n = 1,\ldots,N(t)\,. \qquad (9.9)$$

The parameters to be estimated are β^t (i.e., the price of land in quarter *t*), γ^t (the price of constant quality structures in quarter *t*) and δ (the common depreciation rate for all quarters). Model (9.9) has 14 unknown β^t parameters, 14 unknown γ^t parameters and one unknown δ or 29 unknown parameters in all.[10]

The R^2 for this model was equal to .8847, which is the highest yet for regressions using the data set for the town of "A". The log likelihood was −10642.0, which is considerably higher than the log likelihoods for the two time dummy regressions that used prices as the dependent variable; recall the regression results associated with the construction of indexes P_{H4} and P_{H5} defined in Chapter 6 where the log likelihoods were −10790.4 and −10697.8. The estimated decade straight line net depreciation rate was 0.1068 (0.00284).

[10] Model (9.9) is similar in structure to the hedonic imputation model described in Section 6.8 except that the present model is more parsimonious; there is only one depreciation rate, as opposed to 14 depreciation rates in the imputation model defined by equations (6.25), and there is no constant term. The important factor in both models is that the prices of land and quality adjusted structures are allowed to vary independently across time periods.

The estimated land price series $\hat{\beta}^1,...,\hat{\beta}^{14}$ (rescaled to equal 1 in quarter 1), labeled P_{L1}, and quality adjusted price series for structures $\hat{\gamma}^1,...,\hat{\gamma}^{14}$ (rescaled also), labeled P_{S1}, are plotted in Figure 9.1 and listed in Table 9.1. Using these price series and the corresponding quantity data for each quarter t, i.e., the amount of land transacted, $L^t \equiv \sum_{n=1}^{N(t)} L_n^t$, and the quantity of constant quality structures, $S^{t*} \equiv \sum_{n=1}^{N(t)} (1-\hat{\delta}A_n^t)S_n^t$, an overall property price index has been constructed using the Fisher formula. This overall index, labeled P_1, is also plotted in Figure 9.1 and listed in Table 9.1. For comparison purposes, the Fisher hedonic imputation index constructed in section 6.8, P_{HIF} is also presented.

It can be seen that the new overall hedonic price index based on a cost of production approach to the hedonic functional form, P_1, is very close to the Fisher hedonic imputation index P_{HIF}. However, the price series for land, P_{L1}, and the price series for quality adjusted structures, P_{S1}, are not credible at all: there are large random fluctuations in both series. Notice that when the price of land spikes upwards, there is a corresponding dip in the price of structures. This is a clear sign of multicollinearity between the land and quality adjusted structures variables, which leads to highly unstable estimates for the prices of land and structures. In section 9.6 an attempt will be made to cure this volatility problem.

Figure 9.1. The Price of Land (P_{L1}), the Price of Quality Adjusted Structures (P_{S1}), the Overall Cost of Production House Price Index (P_1) and the Fisher Hedonic Imputation House Price Index

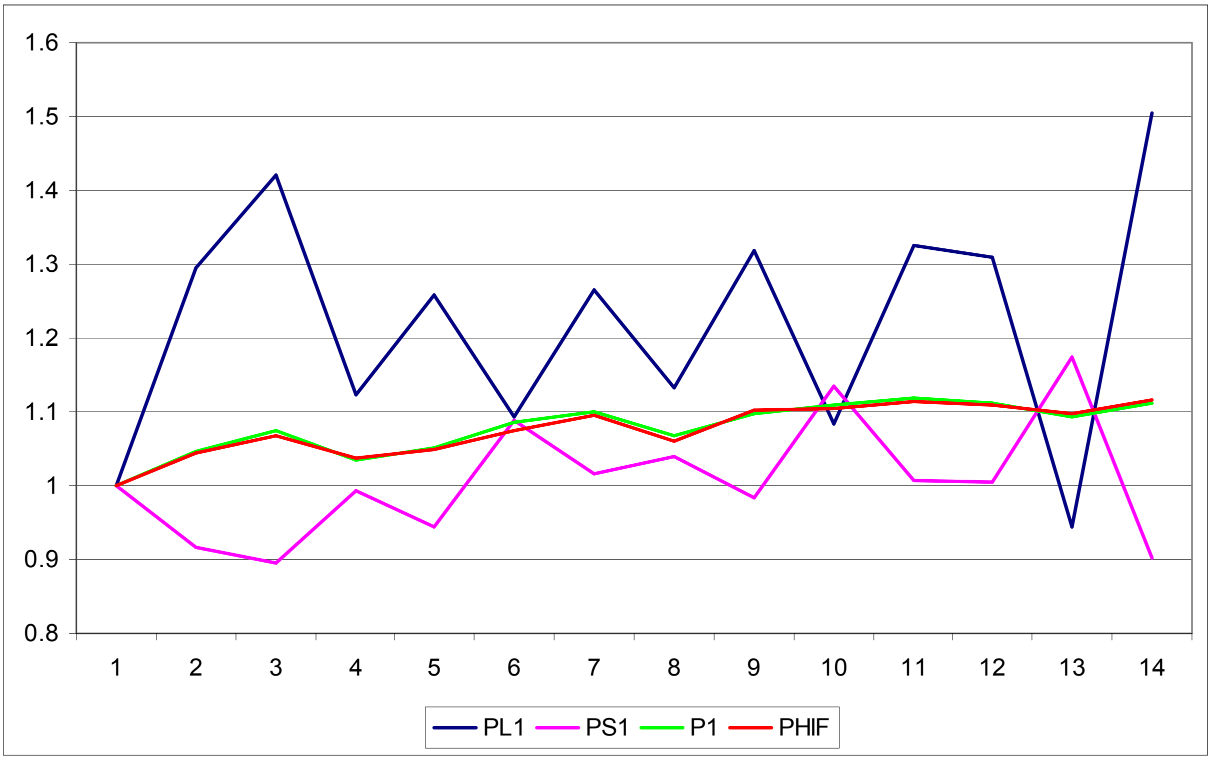


Table 9.1. The Price of Land (P_{L1}), the Price of Quality Adjusted Structures (P_{S1}), the Overall Cost of Production House Price Index (P_1) and the Fisher Hedonic Imputation House Price Index

Quarter	P_{L1}	P_{S1}	P_1	P_{HIF}
1	1.00000	1.00000	1.00000	1.00000
2	1.29547	0.91603	1.04571	1.04356
3	1.42030	0.89444	1.07482	1.06746
4	1.12290	0.99342	1.03483	1.03834
5	1.25820	0.94461	1.05147	1.04794
6	1.09346	1.08879	1.08670	1.07553
7	1.26514	1.01597	1.09941	1.09460
8	1.13276	1.03966	1.06787	1.06158
9	1.31816	0.98347	1.09713	1.10174
10	1.08366	1.13591	1.11006	1.10411
11	1.32624	1.00699	1.11782	1.11430
12	1.30994	1.00502	1.11077	1.10888
13	0.94311	1.17530	1.09373	1.09824
14	1.50445	0.9032	1.11147	1.11630

The Use of Linear Splines

There is a tendency for the price of land per meter squared to decrease for large lots. In order to account for this, a *linear spline model for the price of land* will be used.[11] For lots that are less than 160 m^2, it is assumed that the cost of land per meter squared is β_S^t in quarter *t*. For properties that have lot sizes between 160 m^2 and 300 m^2, it is assumed that the cost of land changes to a price of β_M^t per additional square meter in quarter *t*. Finally, for plots above 300 m^2, the marginal price of an additional unit of land is set equal to β_L^t per square meter in quarter *t*. Let the sets of sales of small, medium and large plots be denoted by $S_S(t)$, $S_M(t)$ and $S_L(t)$, respectively, for $t = 1,...,14$. For sales *n* of properties that fall into the small land size group during quarter *t*, the hedonic regression model is given by (9.10); for the medium group by (9.11) and for the large land size group by (9.12):

$$p_n^t = \beta_S^t L_n^t + \gamma^t (1 - \delta A_n^t) S_n^t + \varepsilon_n^t, \qquad t = 1,...,14;\ n \in S_S(t); \qquad (9.10)$$

$$p_n^t = \beta_S^t[160] + \beta_M^t[L_n^t - 160] + \gamma^t (1 - \delta A_n^t) S_n^t + \varepsilon_n^t, \qquad t = 1,...,14;\ n \in S_S(t); \qquad (9.11)$$

$$p_n^t = \beta_S^t[160] + \beta_M^t[140] + \beta_L^t[L_n^t - 300] + \gamma^t (1 - \delta A_n^t) S_n^t + \varepsilon_n^t, \quad t = 1,...,14;\ n \in S_S(t); \qquad (9.12)$$

Estimating the model defined by (9.10)-(9.12) on the data for the town of "A", the estimated decade depreciation rate was $\hat{\delta} = 0.1041$ (0.00419). The R^2 for this model was .8875, which is an increase over the previous no-splines model where the R^2 was .8847. The log likelihood was −10614.2 (an increase of 28 from the previous model's log likelihood.) The first period parameter values for the three marginal prices for land were $\hat{\beta}_S^1 = 281.4$ (55.9), $\hat{\beta}_M^1 = 380.4$ (48.5) and $\hat{\beta}_L^1 = 188.9$ (27.5). In other words, in quarter 1, the marginal cost per m^2 of small

[11] This approach follows that of Diewert, de Haan and Hendriks (2010) (2011). The use of linear splines to model nonlinearities in the price of land as a function of lot size is due to Francke (2008).

lots is estimated to be 281.4 Euros per m^2, for medium sized lots, the estimated marginal cost is 380.4 Euros/m^2, and for large lots, the estimated marginal cost is 188.9 Euros/m^2. The first period parameter value for quality adjusted structures is $\hat{\gamma}^1 = 978.1$ Euros/m^2 with a standard error of 82.3. The lowest t statistic for all of the 57 parameters was 3.3, so all of the estimated coefficients in this model are significantly different from zero.

Once the parameters for the model have been estimated, then in each quarter t, the predicted value of land for small, medium and large lot sales, V_{LS}^t, V_{LM}^t and V_{LL}^t, respectively, can be calculated along with the associated quantities of land, L_{LS}^t, L_{LM}^t and L_{LL}^t, as follows:

$$V_{LS}^t \equiv \sum_{n \in S_S(t)} \hat{\beta}_S^t L_n^t \,, \qquad t = 1,...,14 ; \qquad (9.13)$$

$$V_{LM}^t \equiv \sum_{n \in S_M(t)} \{\hat{\beta}_S^t [160] + \hat{\beta}_M^t [L_n^t - 160]\} \,, \qquad t = 1,...,14 ; \qquad (9.14)$$

$$V_{LL}^t \equiv \sum_{n \in S_L(t)} \{\hat{\beta}_S^t [160] + \hat{\beta}_M^t [140] + \hat{\beta}_L^t [L_n^t - 300]\} \,, \qquad t = 1,...,14 ; \qquad (9.15)$$

$$L_{LS}^t \equiv \sum_{n \in S_S(t)} L_n^t \,, \qquad t = 1,...,14 ; \qquad (9.16)$$

$$L_{LM}^t \equiv \sum_{n \in S_M(t)} L_n^t \,, \qquad t = 1,...,14 ; \qquad (9.17)$$

$$L_{LL}^t \equiv \sum_{n \in S_L(t)} L_n^t \,, \qquad t = 1,...,14 ; \qquad (9.18)$$

The corresponding *average quarterly prices*, P_{LS}^t, P_{LM}^t and P_{LL}^t, for the three types of lot are defined as the above values divided by the above quantities:

$$P_{LS}^t \equiv V_{LS}^t / L_{LS}^t \,; \; P_{LM}^t \equiv V_{LM}^t / L_{LM}^t \,; \; P_{LL}^t \equiv V_{LL}^t / L_{LL}^t \,. \qquad t = 1,...,14 ; \qquad (9.19)$$

The average land prices for small, medium and large lots defined by equation (9.19) and the corresponding quantities of land defined by (9.16)-(9.18) can be used to construct a chained Fisher land price index, which is denoted by P_{L2}. This index is plotted in Figure 9.2 and listed in Table 9.2. As before, the estimated quarter *t* price per meter squared of quality adjusted structures is $\hat{\gamma}^t$ and the quantity of constant quality structures is $S^{t*} \equiv \sum_{n=1}^{N(t)} (1 - \hat{\delta} A_n^t) S_n^t$. The structures price and quantity series $\hat{\gamma}^t$ and S^{t*} were combined with the three land price and quantity series to form a chained overall Fisher house price index P_2, which is also graphed in Figure 9.2 and listed in Table 9.2. The constant quality structures price index P_{S2} (which is a normalization of the series $\hat{\gamma}^1,...,\hat{\gamma}^{14}$) is presented as well.

The overall house price index resulting from the spline model, P_2, is fairly close to the Fisher hedonic imputation index P_{HIF}. However, the spline model does not generate sensible series for the price of land, P_{L2}, and the price of structures, P_{S2}: both series are extremely volatile but in opposite directions. As was the case with the previous cost of production model, the present model suffers from a multicollinearity problem.

Comparing Figures 9.1 and 9.2, it can be seen that in Figure 9.1 the price index for land is above the overall price index for the most part and the price index for structures is below the overall index while in Figure 9.2, this pattern *reverses*. This instability is again an indication of multicollinearity. In the following section an attempt to cure this problem will be made by imposing monotonicity restrictions on the prices of the constant quality structures.

Figure 9.2. The Price of Land (P_{L2}), the Price of Structures (P_{S2}), the Overall Price Index Using Splines on Land (P_2) and the Fisher Hedonic Imputation Price Index

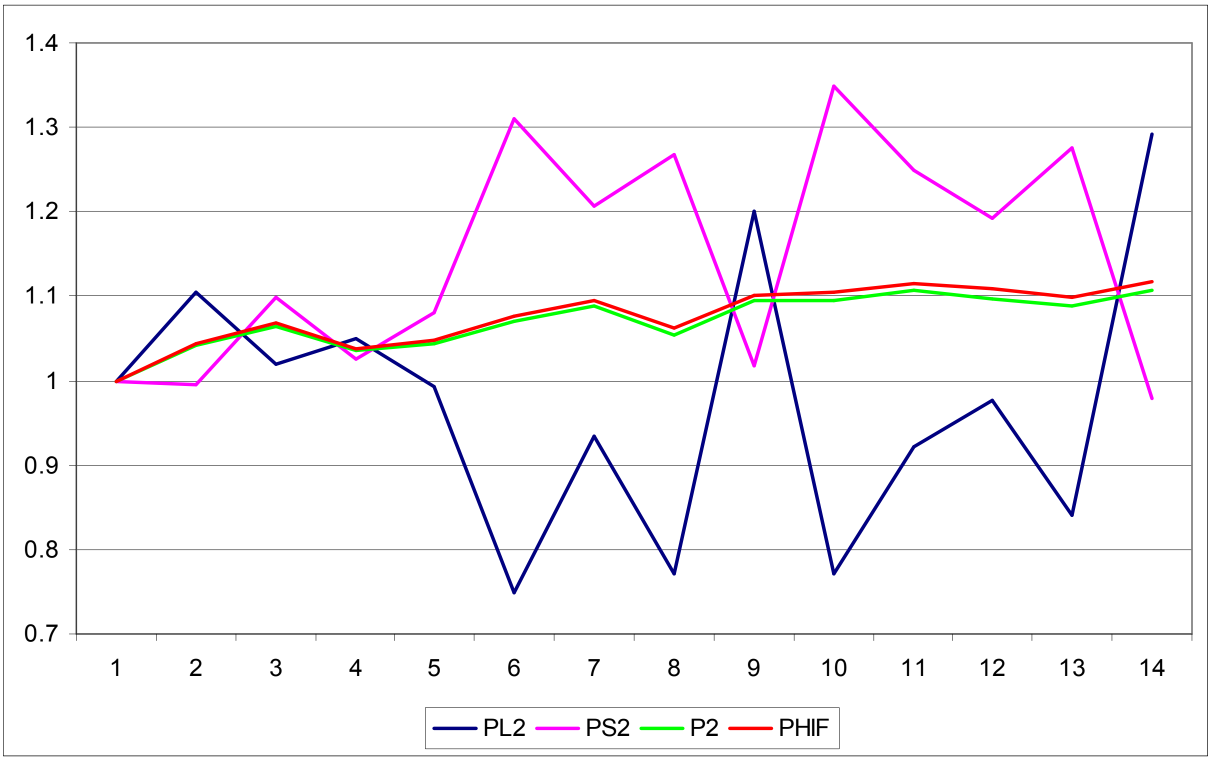


Table 9.2. The Price of Land (P_{L2}), the Price of Structures (P_{S2}), the Overall Price Index Using Splines on Land (P_2) and the Fisher Hedonic Imputation Price Index

Quarter	P_{L2}	P_{S2}	P_2	P_{HIF}
1	1.00000	1.00000	1.00000	1.00000
2	1.10534	0.99589	1.04137	1.04356
3	1.02008	1.09803	1.06465	1.06746
4	1.05082	1.02542	1.03608	1.03834
5	0.99379	1.08078	1.04294	1.04794
6	0.74826	1.31122	1.06982	1.07553
7	0.93484	1.20719	1.08912	1.09460
8	0.77202	1.26718	1.05345	1.06158
9	1.19966	1.01724	1.09425	1.10174
10	0.77139	1.34813	1.09472	1.10411
11	0.92119	1.24884	1.10596	1.11430
12	0.97695	1.19188	1.09731	1.10888
13	0.84055	1.27531	1.08811	1.09824
14	1.29261	0.97875	1.10613	1.11630

9.6 An Approach Based on Monotonicity Restrictions

It is likely that Dutch construction costs did not fall significantly during the sample period.[12] If this is indeed the case, *monotonicity restrictions* on the quarterly prices of quality adjusted structures, $\gamma^1, \gamma^2, \gamma^3, ..., \gamma^{14}$, can be imposed on the hedonic regression model (9.10)-(9.12) in section 9.5 by replacing the constant quality quarter *t* structures price parameters γ^t by the following sequence of parameters for the 14 quarters: γ^1, $\gamma^1 + (\phi^2)^2$, $\gamma^1 + (\phi^2)^2 + (\phi^3)^2, ...,$ $\gamma^1 + (\phi^2)^2 + (\phi^3)^2 + ... + (\phi^{14})^2$, where $\phi^2, \phi^3, ..., \phi^{14}$ are scalar parameters.[13] For each quarter *t* starting at quarter 2, the price of a square meter of constant quality structures γ^t is thus equal to the previous period's price γ^{t-1} plus the square of a parameter ϕ^{t-1}, $(\phi^{t-1})^2$. Now replace this reparameterization of the structures price parameters γ^t in (9.10)-(9.12) in order to obtain a linear spline model for the price of land with monotonicity restrictions on the price of constant quality structures.

Implementing this new model using the data for the Dutch town of "A", the estimated decade depreciation rate was $\hat{\delta} = 0.1031$ (0.00386). The R^2 for this model was .8859, a drop from the previous unrestricted spline model where the R^2 was .8875. The log likelihood was −10630.5, a decrease of 16.3 over the previous unrestricted model. Eight of the 13 new parameters ϕ^t are zero in this monotonicity restricted hedonic regression. The first period parameter values for the three marginal land prices are $\hat{\beta}_S^1 = 278.6$ (37.2), $\hat{\beta}_M^1 = 380.3$ (41.0) and $\hat{\beta}_L^1 = 188.0$; these values are almost identical to the corresponding estimates in the previous unrestricted model. The first period parameter estimate for quality adjusted structures is $\hat{\gamma}^1 = 980.5$ (49.9) Euros/m^2., which is little changed from the previous unrestricted estimate of 978.1 Euros/m^2.

Once the parameters for the model have been estimated, convert the estimated ϕ^t parameters into estimated γ^t parameters using the following recursive equations:

$$\hat{\gamma}^{t+1} \equiv \hat{\gamma}^t + (\hat{\phi}^t)^2; \qquad t = 2,...,14. \quad (9.19)$$

Now use equations (9.13)-(9.19) in the previous section in order to construct a chained Fisher index of land prices, which is denoted by P_{L3}. This index is plotted in Figure 9.3 and listed in Table 9.3. As in the previous two models, the estimated period *t* price for a squared meter of quality adjusted structures is $\hat{\gamma}^t$ and the corresponding quantity of constant quality structures is $S^{t*} \equiv \sum_{n=1}^{N(t)} (1 - \hat{\delta} A_n^t) S_n^t$. The price and quantity series $\hat{\gamma}^t$ and S^{t*} were combined with the three land price and quantity series to construct a chained overall Fisher house price index P_3 which is also graphed in Figure 9.3 and listed in Table 9.3. The constant quality structures

[12] Some direct evidence on this assertion will be presented in the following section.

[13] This method for imposing monotonicity restrictions was used by Diewert, de Haan and Hendriks (2010) with the difference that they imposed monotonicity on both structures and land prices, whereas here, monotonicity restrictions are imposed on structures prices only.

price index P_{S3} (a normalization of the series $\hat{\gamma}^1,\ldots,\hat{\gamma}^{14}$) may be found in Figure 9.3 and Table 9.3 as well.

Figure 9.3. The Price of Land (P_{L3}), the Price of Quality Adjusted Structures (P_{S3}), the Overall House Price Index with Monotonicity Restrictions on Structures (P_3) and the Overall House Price Index Using Splines on Land (P_2)

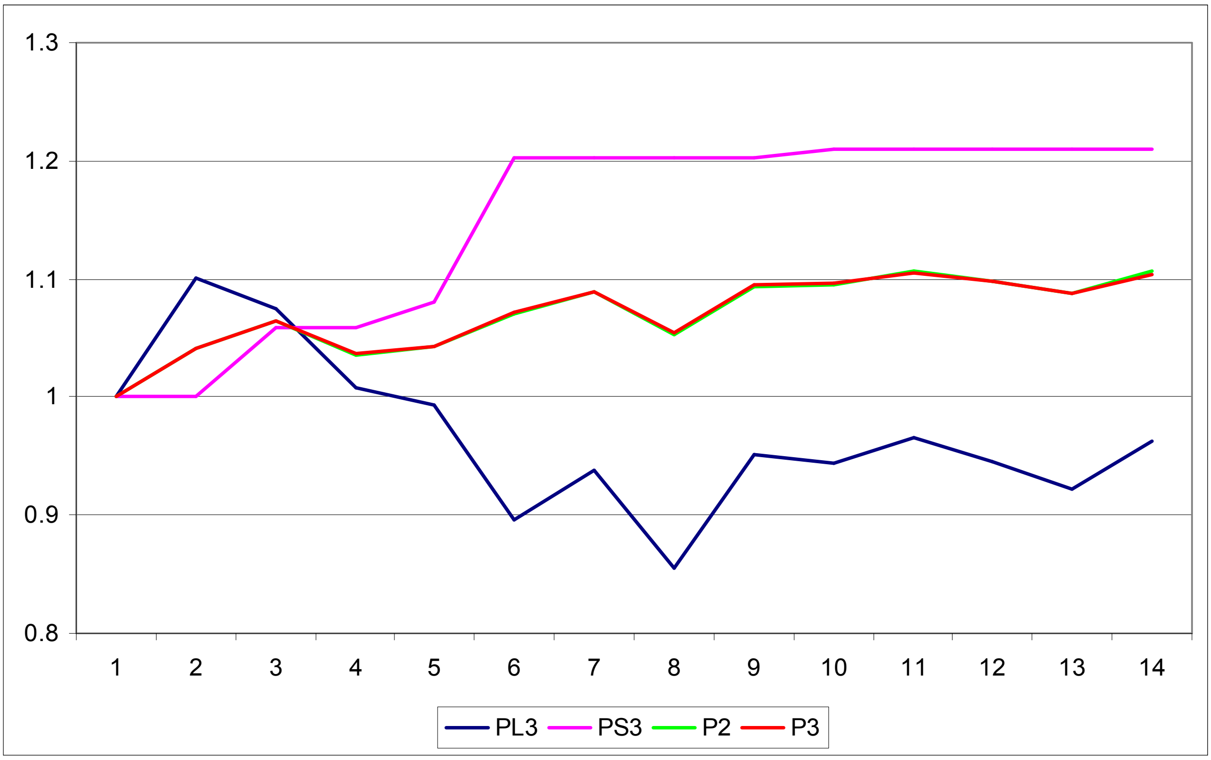


Table 9.3. The Price of Land (P_{L3}), the Price of Quality Adjusted Structures (P_{S3}), the Overall House Price Index with Monotonicity Restrictions on Structures (P_3) and the Overall House Price Index Using Splines on Land (P_2)

Quarter	P_{L3}	P_{S3}	P_3	P_2
1	1.00000	1.00000	1.00000	1.00000
2	1.10047	1.00000	1.04148	1.04137
3	1.07431	1.05849	1.06457	1.06465
4	1.00752	1.05849	1.03627	1.03608
5	0.99388	1.08078	1.04316	1.04294
6	0.89560	1.20300	1.07168	1.06982
7	0.93814	1.20300	1.08961	1.08912
8	0.85490	1.20300	1.05408	1.05345
9	0.95097	1.20300	1.09503	1.09425
10	0.94424	1.21031	1.09625	1.09472
11	0.96514	1.21031	1.10552	1.10596
12	0.94596	1.21031	1.09734	1.09731
13	0.92252	1.21031	1.08752	1.08811
14	0.96262	1.21031	1.10427	1.10613

The new overall house price index P_3 that imposed monotonicity on the quality adjusted price of structures in Figure 9.3 can hardly be distinguished from the previous overall house price index P_2, which was based on a similar hedonic regression model except that the movements in the price of structures were not restricted. The fluctuations in the price of land and quality adjusted structures are no longer violent.

While the above results seem "reasonable", the early rapid rise in the price of structures and the slow growth in structures prices from quarter 6 to 14 are not very likely. In the following section, one more method for extracting separate structures and land components out of real estate sales data will therefore be tried.

9.7 An Approach Based on Exogenous Information on the Price of Structures

Using Information on Construction Costs

Many countries have new construction price indexes available on a quarterly basis. This is the case for the Netherlands.[14] If one is willing to make the assumption that construction costs for houses have the same rate of growth over the study period across all cities in the Netherlands, the information on construction costs can be used to eliminate the multicollinearity problem encountered in the previous sections.

Recall equations (9.10)-(9.12) in section 9.5 above. These are the estimating equations for the unrestricted hedonic regression model based on costs of production. In the present section, the constant quality price parameters for the structures, the γ^t for $t = 2,...,14$ in (9.10)-(9.12), are replaced by the following numbers, which involve only the single unknown parameter γ^1:[15]

$$\gamma^t = \gamma^1 \mu^t ; \qquad t = 2,...,14. \quad (9.20)$$

where μ^t is the statistical agency's *construction cost price index* for the location and the type of house under consideration, normalized to equal 1 in quarter 1. The new hedonic regression model is again defined by equations (9.10)-(9.12) except that the 14 unknown γ^t parameters are now defined by (9.20), so that only γ^1 needs to be estimated. The number of parameters to be estimated in this new restricted model is 44 whereas the old number was 57.

Using the data for the town of "A", the estimated decade depreciation rate was $\hat{\delta} = 0.1028$ (0.00433). The R^2 for this model was .8849, a small drop from the previous restricted spline

[14] From the Statistics Netherlands (2010) online source, Statline, the following series was downloaded for the New Dwellings Output Price Index for the 14 quarters in our sample of house sales: 98.8, 98.1, 100.3, 102.7, 99.5, 100.5, 100.0, 100.3, 102.2, 103.2, 105.6, 107.9, 110.0, 110.0. This series was normalized to 1 in the first quarter by dividing each entry by 98.8. The resulting series is denoted by μ^1 (=1), μ^2,...,μ^{14}.

[15] The technique suggested here for decomposing property prices into land and structures components can be viewed as a variant of a technique used by Davis and Heathcote (2007) and Davis and Palumbo (2008).

model, where the R^2 was .8859, and a larger drop from the unrestricted spline model R^2 in section 9.5, which was .8875. The log likelihood was −10640.1, a decrease of 10 over the monotonicity restricted model. The first period parameter estimates for the 3 marginal prices for land are now $\hat{\beta}_S^1 = 215.4$ (30.0), $\hat{\beta}_M^1 = 362.6$ (46.7) and $\hat{\beta}_L^1 = 176.4$ (28.4). They differ slightly from the previous figures. The first period parameter estimate for the quality adjusted structures is $\hat{\gamma}^1 = 1085.9$ (22.9) Euros/m^2, which is significantly higher than the unrestricted estimate of 980.5 Euros/m^2. So the imposition of a (nationwide) growth rate on the change in the price of quality adjusted structures has had some effect on the estimates for the levels of land and structures prices.

As usual, equations (9.13)-(9.19) were used in order to construct a chained Fisher index of land prices, which is denoted by P_{L4}. This index is plotted in Figure 9.4 and listed in Table 9.4. As for the previous three models, the estimated price in quarter *t* for a square meter of quality adjusted structures is $\hat{\gamma}^t$ (which now equals $\hat{\gamma}^1\mu^t$) and the corresponding quantity is $S^{t*} \equiv \sum_{n=1}^{N(t)}(1-\hat{\delta}A_n^t)S_n^t$. These structures price and quantity series were again combined with the three land price and quantity series to form a chained overall Fisher house price index P_4, which is graphed in Figure 9.4 and listed in Table 9.4. The constant quality structures price index P_{S4} (a normalization of the series $\hat{\gamma}^1,...,\hat{\gamma}^{14}$) is also presented.

Figure 9.4. The Price of Land (P_{L4}), the Price of Quality Adjusted Structures (P_{S4}) and the Overall House Price Index using Exogenous Information on the Price of Structures (P_4)

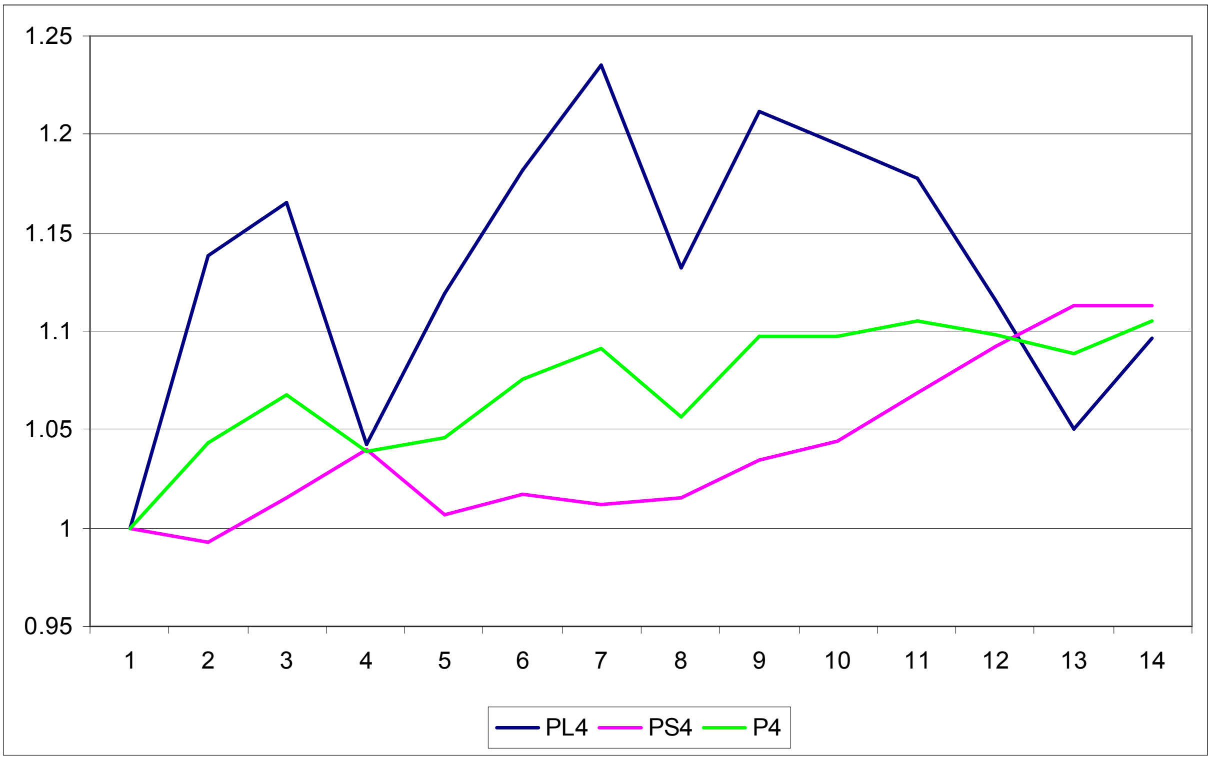


Table 9.4. The Price of Land (P_{L4}), the Price of Quality Adjusted Structures (P_{S4}) and the Overall House Price Index using Exogenous Information on the Price of Structures (P_4)

Quarter	P_{L4}	P_{S4}	P_4
1	1.00000	1.00000	1.00000
2	1.13864	0.99291	1.04373
3	1.16526	1.01518	1.06752
4	1.04214	1.03947	1.03889
5	1.11893	1.00709	1.04628
6	1.18183	1.01721	1.07541
7	1.23501	1.01215	1.09121
8	1.13257	1.01518	1.05601
9	1.21204	1.03441	1.09701
10	1.19545	1.04453	1.09727
11	1.17747	1.06883	1.10564
12	1.11588	1.09211	1.09815
13	1.05070	1.11336	1.08863
14	1.09648	1.11336	1.10486

A comparison of Figures 9.3 and 9.4 shows that the imposition of the national growth rates for new dwelling construction costs has changed the nature of the land and structures price indexes: in Figure 9.3, the price series for land lies below the overall house price series for most of the sample period while in Figure 9.4, the pattern is reversed: the price series for land lies above the overall house price series for most of the sample period (and vice versa for the price of structures). But which model is best? Although the previous model can be preferred on statistical grounds because the log likelihood is somewhat higher, we would nevertheless prefer the present model that uses of exogenous information on structures prices because it yields a more plausible pattern of price changes for land and structures.

Choosing the "Best" Overall Index

This section is concluded by listing and charting our four "best" *overall* indexes: the chained stratified sample Fisher index P_{FCH} constructed in Chapter 5, the chained hedonic imputation Fisher index P_{HIF} studied in Chapter 6, the index P_3 that resulted from the cost based hedonic regression model with monotonicity restrictions constructed in section 9.6, and the index P_4 that resulted from the cost based hedonic regression model using exogenous information on the price of structures studied in the present section. As can be seen from Figure 9.5, all four indexes paint much the same picture. Note that P_3 and P_4 are virtually identical.

All things considered, the hedonic imputation index P_{HIF} is our preferred index since it has fewer restrictions than the other indexes and seems closest to a matched model index in spirit, followed by the two cost of production hedonic indexes P_4 and P_3, followed by the stratified sample index P_{FCH}. The latter likely suffers from some unit value bias. Hedonic indexes can be biased too (if important explanatory variables are omitted or if an "incorrect" functional form is chosen), but in general we would prefer hedonic regression methods over stratification methods. If separate land and structures indexes are required, we are in favour of the cost based hedonic regression model that uses exogenous information on the price of structures.

Figure 9.5. House Price Indexes Using Exogenous Information (P_4) and Using Monotonicity Restrictions (P_3), the Chained Fisher Hedonic Imputation Index and the Chained Fisher Stratified Sample Index

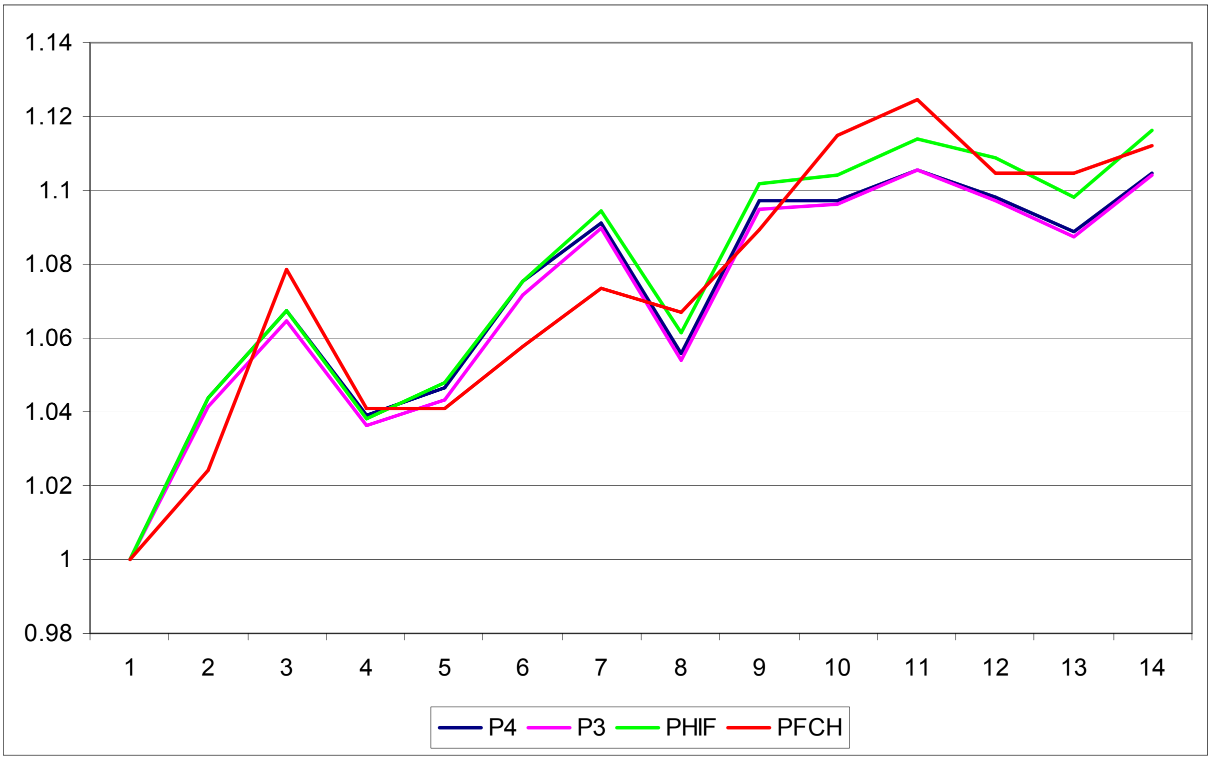


Table 9.5. House Price Indexes Using Exogenous Information (P_4) and Using Monotonicity Restrictions (P_3), the Chained Fisher Hedonic Imputation Index and the Chained Fisher Stratified Sample Index

Quarter	P_4	P_3	P_{HIF}	P_{FCH}
1	1.00000	1.00000	1.00000	1.00000
2	1.04373	1.04148	1.04356	1.02396
3	1.06752	1.06457	1.06746	1.07840
4	1.03889	1.03627	1.03834	1.04081
5	1.04628	1.04316	1.04794	1.04083
6	1.07541	1.07168	1.07553	1.05754
7	1.09121	1.08961	1.09460	1.07340
8	1.05601	1.05408	1.06158	1.06706
9	1.09701	1.09503	1.10174	1.08950
10	1.09727	1.09625	1.10411	1.11476
11	1.10564	1.10552	1.11400	1.12471
12	1.09815	1.09734	1.10888	1.10483
13	1.08863	1.08752	1.09824	1.10450
14	1.10486	1.10427	1.11630	1.11189

9.8 Rolling Window Hedonic Regressions

A problem with the hedonic regression model discussed in the previous section (and all other hedonic models discussed in this Handbook with the exception of hedonic imputation models) was mentioned in Chapter 6: when more data are added, the indexes generated by the model change. This feature of these regression based methods makes these models unsatisfactory for statistical agency use, where users expect the official numbers to remain unchanged as time passes. Users may tolerate a few revisions to recent data but typically, they would not like all the numbers to be revised back into the indefinite past as new data become available. As was noted in section 6.2, a simple solution to this problem is available, however. This so-called *rolling window approach* will be outlined in more detail and applied to the cost based hedonic regression model that uses exogenous information on the price of structures.

First, one chooses a "suitable" number of time periods (equal to or greater than two) where it is thought that the hedonic model yields "reasonable" results; this will be the *window length* (say M periods) for the sequence of regression models which will be estimated. Secondly, an initial regression model is estimated and the appropriate indexes are calculated using data pertaining to the first M periods in the data set. Next, a second regression model is estimated where the data consist of the initial data less the data for period 1 but adding the data for period M+1. Appropriate price indexes are calculated for this new regression model but only the rate of increase of the index going from period M to M+1 is used to update the previous sequence of M index values. This procedure is continued with each successive regression dropping the data of the previous earliest period and adding the data for the next period, with one new update factor being added with each regression. If the window length is a year, then this procedure is called a *rolling year hedonic regression model*; for a general window length, it is called a *rolling window hedonic regression model*.[16]

Using the data for the town of "A", the rolling window procedure for the model in section 9.7 was applied with a window length of 9 quarters. The initial hedonic regression model defined by equations (9.10)-(9.12) and (9.20) was thus initially estimated for the first 9 quarters. The resulting price indexes for land and for constant quality structures and the overall index are denoted by P_{RWL4}, P_{RWS4} and P_{RW4} and are listed in the first 9 rows of Table 9.6.[17] Next, a regression covering quarters 2-10 was run and the resulting land, structures and overall price indexes were used to update the initial indexes; i.e., the price of land in quarter 10 of Table 9.6 is equal to the price of land in quarter 9 times the price relative for land (quarter 10 land

[16] This procedure was recently used by Shimizu, Nishimura and Watanabe (2010) and Shimizu, Takatsuji, Ono and Nishimura (2010) in their hedonic regression models for Tokyo house prices. An analogous procedure has also been recently applied by Ivancic, Diewert and Fox (2011) and de Haan and van der Grient (2011) in their adaptation of the GEKS method for making international comparisons to the scanner data context.

[17] We imposed the restrictions (33) on the rolling window regressions and so the rolling window constant quality price index for structures, P_{RWS}, is equal to the constant quality price index for structures listed in Table 9.4, P_{S4}.

index divided by the quarter 9 land index) obtained from the regression covering quarters 2-10, etc. Similar updating was done for the next 4 quarters using regressions covering quarters 3-11, 4-12, 5-13 and 6-14.

The rolling window indexes can be compared to the corresponding indexes based on the data pertaining to all 14 quarters constructed in section 9.7 by looking at Table 9.6. Recall that the estimated depreciation rate and the estimated quarter 1 price of quality adjusted structures for the last model section 9.7 were $\hat{\delta} = 0.1028$ and $\hat{\gamma}^1 = 1085.9$, respectively. If by chance the 6 rolling window hedonic regressions generated the exact same estimates for δ and γ, then the indexes resulting from the rolling window regressions would coincide with the indexes P_{L4}, P_{S4} and P_4 described in section 9.7. The estimates for δ generated by the 6 rolling window regressions are 0.10124, 0.10805, 0.11601, 0.11103, 0.10857 and 0.10592. The estimates for γ^1 generated by the 6 rolling window regressions are 1089.6, 1103.9, 1088.1, 1101.0, 1123.5 and 1100.9. While these estimates are not identical to the corresponding estimates of 0.1028 and 1085.9 for P_4, they are fairly close. So we can expect the rolling window indexes to be close to their counterparts for the last model in section 9.7. The R^2 values for the 6 rolling window regressions were .8803, .8813, .8825, .8852, .8811 and .8892.

Table 9.6. The Price of Land (P_{L4}), the Price of Quality Adjusted Structures (P_{S4}), the Overall House Price Index using Exogenous Information on the Price of Structures (P_4) and their Rolling Window Counterparts (P_{RWL}) and (P_{RW})

Quarter	P_{RWL}	P_{L4}	P_{RW}	P_4	P_{S4}
1	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.14073	1.13864	1.04381	1.04373	0.99291
3	1.16756	1.16526	1.06766	1.06752	1.01518
4	1.04280	1.04214	1.03909	1.03889	1.03947
5	1.12055	1.11893	1.04635	1.04628	1.00709
6	1.18392	1.18183	1.07542	1.07541	1.01721
7	1.23783	1.23501	1.09123	1.09121	1.01215
8	1.13408	1.13257	1.05602	1.05601	1.01518
9	1.21417	1.21204	1.09698	1.09701	1.03441
10	1.19772	1.19545	1.09738	1.09727	1.04453
11	1.18523	1.17747	1.10718	1.10564	1.06882
12	1.11889	1.11588	1.09779	1.09815	1.09201
13	1.05191	1.05070	1.08893	1.08863	1.11335
14	1.09605	1.09648	1.10436	1.10486	1.11335

The rolling window series for the price of quality adjusted structures, P_{RWS}, is not listed in Table 9.6 since it is identical to the series P_{S4} described in section 9.7.[18] The rolling window price series for land, P_{RWL}, is extremely close to its counterpart in section 9.7, P_{L4}, and the overall rolling window price series for detached dwellings in the town of "A", P_{RW}, is also

[18] By construction, P_{S4} and P_{RWS} are both equal to the official Statistics Netherlands construction price index for new dwellings, μ^t/μ^1 for $t = 1,...,14$.

close to its counterpart section 9.7, P_4. The corresponding series in Table 9.6 are so close to each other that we decided not to provide a chart.

Using the data for the town of "A', rolling window hedonic regressions gave much the same results as a hedonic regression that covers the whole sample period. This supports our view that the rolling window approach can be used by statistical agencies to compile an RPPI based on hedonic regressions, including a decomposition into land and structures components.

9.9 The Construction of Price Indexes for the Stock of Dwelling Units

This section shows how hedonic regression models can be used to form an approximate RPPI for the *stock* of dwelling units. Before turning to the model estimated in section 9.7, we will first look at the hedonic imputation model discussed in section 6.8 and compare the resulting index with an approximate stock based index using the stratification approach.

The Hedonic Imputation Model

Recall that the hedonic imputation model was defined by equations (6.25), where L_n^t, S_n^t and A_n^t denoted, respectively, the land area, structure area, and age (in decades) of property *n* sold in period *t*. In order to form a price index for the stock of detached houses in the town of "A", it would in principle be necessary to know *L*, *S* and *A* for all detached houses in "A" during some base period. This information is not available to us, but we can treat the total number of detached houses sold over the sample period as an *approximation* to the stock of this type.[19] In our data set there were $N(1)+N(2)+...+N(14)=2289$ of such transactions.[20]

As in section 6.8, the estimated parameters for land size, structure size and depreciation in quarter *t* are denoted by $\hat{\beta}^t$, $\hat{\gamma}^t$ and $\hat{\delta}^t$; $\hat{\alpha}^t$ denotes the constant term. Our approximation to *the total value of the housing* stock for quarter *t*, V^t, is defined as

$$V^t \equiv \sum_{s=1}^{14}\sum_{n=1}^{N(s)}[\hat{\alpha}^t + \hat{\beta}^t L_n^s + \hat{\gamma}^t(1-\hat{\delta}^t A_n^s)S_n^s], \qquad t=1,...,14. \quad (9.21)$$

That is, V^t is (approximated by) the imputed value of all houses traded during the 14 quarters in our sample, where the regression coefficients from the quarter *t* hedonic imputation model given by (6.25) serve as weights for the characteristics of each house. Dividing the V^t series

[19] This approximation would probably be an adequate one if the sample period were a decade or so. Obviously, our sample period of 14 quarters is too short to be accurate and there are also sample selectivity problems, i.e., newer houses will be over represented. However, the method we are suggesting here can be illustrated using this rough approximation.

[20] We did not delete the observations for houses that were transacted multiple times over the 14 quarters since a particular house transacted during two or more of the quarters is not actually the same house due to depreciation and renovations.

by the value for quarter 1, V^1, is our first estimated stock price index, P_{Stock1}, for the town of "A".[21] This is a form of a Lowe index; see the CPI Manual (2004) for the properties of Lowe indexes. In Table 9.7 and Figure 9.6 this price index for the stock of houses is compared with the corresponding sales based Fisher hedonic imputation price index, P_{HIF}, from section 6.8.

An additional approximate stock price index based on *stratification*, P_{Stock2} is also graphed in Figure 9.6 and listed in Table 9.7. This index uses the unit value prices for the nonempty cells in the stratification scheme in each quarter, as explained in section 5.5, and uses the imputed prices based on the hedonic imputation regressions from section 6.8 for the empty cells in each quarter. The quantity vector used for P_{Stock2} is the (sample) total quantity vector by cell, which makes P_{Stock2} an alternative Lowe price index. It can be seen that while P_{Stock2} has the same general trend as P_{Stock1} and P_{HIF}, it differs substantially from these hedonic imputation indexes during several quarters. These differences are due to the existence of some unit value bias in the stratification indexes. Thus, although stratification indexes can be constructed for the stock of dwelling units of a certain type and location (with the help of hedonic imputation for empty cells), it appears that the resulting stock indexes will not be as accurate as indexes that are entirely based on the use of hedonic regressions.[22]

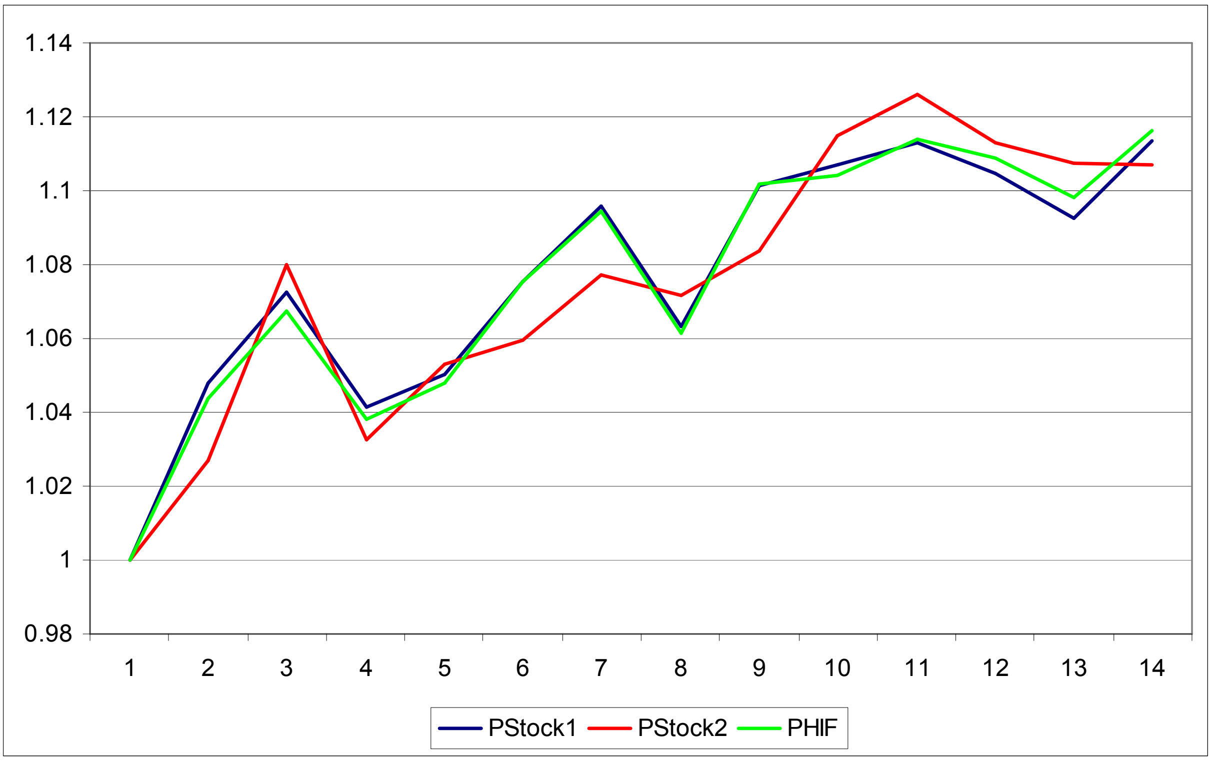
Table 9.7. Approximate Stock Price Indexes and Based on Hedonic Imputation (P_{Stock1}) and Stratification (P_{Stock2}) and the Fisher Hedonic Imputation Sales Price Index

Quarter	P_{Stock1}	P_{Stock2}	P_{HIF}
1	1.00000	1.00000	1.00000
2	1.04791	1.02712	1.04356
3	1.07255	1.07986	1.06746
4	1.04131	1.03257	1.03834
5	1.05040	1.05290	1.04794
6	1.07549	1.05934	1.07553
7	1.09594	1.07712	1.09460
8	1.06316	1.07172	1.06158
9	1.10137	1.08359	1.10174
10	1.10708	1.11482	1.10411
11	1.11289	1.12616	1.11430
12	1.10462	1.11291	1.10888
13	1.09278	1.10764	1.09824
14	1.11370	1.10686	1.11630

[21] Since V^t is a value, it does not appear to be a price series at first glance. But in each quarter, the quantity vector which underlies this value is a vector of ones of dimension 2289, which is constant over the 14 quarters. Hence V^t can also be interpreted as a price series, which is normalized to equal one in quarter 1.

[22] If the imputed prices described in section 5.5 are used for every one of the 45 cell prices for each period (instead of just for the zero transaction cells as was the case for the construction of $P_{Sctock2}$) and the same total sample quantity vector is used as the approximate stock quantity vector, then the resulting Lowe index turns out to be exactly equal to P_{Stock1}. Thus these two different ways for constructing a stock index turn out to be equivalent. The fact that P_{Stock1} is not equal to P_{Stock2} is clear evidence that there is unit value bias in the cells of the stratification scheme: the cells are simply not defined narrowly enough.

Figure 9.6. Approximate Stock Price Indexes and Based on Hedonic Imputation (P_{Stock1}) and Stratification (P_{Stock2}) and the Fisher Hedonic Imputation Sales Price Index



The Use of Exogenous Information on the Price of Structures

The same kind of construction of an approximate stock price index can be applied to the other hedonic regression models discussed in this chapter. Here we will show how this works for the cost based model that used exogenous information on the price of structures, which was explained in section 9.7. This model was defined by equations (9.10)-(9.12) and (9.20). Recall that the sets of period t sales of small, medium and large lot houses were denoted by $S_S(t)$, $S_M(t)$ and $S_L(t)$, respectively; the total number of sales in period t was denoted by $N(t)$ for $t = 1,...,14$. The estimated model parameters are $\hat{\delta}^t$, $\hat{\gamma}^t$ and $\hat{\beta}_S^1$, $\hat{\beta}_M^1$ and $\hat{\beta}_L^1$ for $t = 1,...,14$. The estimated *period t values of all small, medium and large lot houses* traded over the 14 quarters, V_{LS}^t, V_{LM}^t and V_{LL}^t, respectively, are defined by (9.22)-(9.24):

$$V_{LS}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_S(s)} \hat{\beta}_S^t L_n^s \,; \qquad t = 1,...,14 \,; \qquad (9.22)$$

$$V_{LM}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_M(s)} \{\hat{\beta}_S^t [160] + \hat{\beta}_M^t [L_n^s - 160]\} \,; \qquad t = 1,...,14 \,; \qquad (9.23)$$

$$V_{LL}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_L(s)} \{\hat{\beta}_S^t [160] + \hat{\beta}_M^t [140] + \hat{\beta}_L^t [L_n^s - 300]\} \,; \qquad t = 1,...,14 \,; \qquad (9.24)$$

The estimated period *t* value of quality adjusted structures, V_S^t, is defined by

$$V_S^t \equiv \sum_{s=1}^{14} \sum_{n=1}^{N(s)} \hat{\gamma}^1 \mu^t (1 - \hat{\delta} A_n^s) S_n^s \; ; \qquad t = 1,...,14 , \qquad (9.25)$$

where all structures traded during the 14 quarters are included.

The *quantities* that correspond to the above period *t* valuations of the three land stocks and the stock of structures are defined as follows:[23]

$$Q_{LS}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_S(s)} L_n^s \; ; \qquad t = 1,...,14 ; \qquad (9.26)$$

$$Q_{LM}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_M(s)} L_n^s \; ; \qquad t = 1,...,14 ; \qquad (9.27)$$

$$Q_{LL}^t \equiv \sum_{s=1}^{14} \sum_{n \in S_L(s)} L_n^s \; ; \qquad t = 1,...,14 ; \qquad (9.28)$$

$$Q_S^t \equiv \sum_{s=1}^{14} \sum_{n=1}^{N(s)} (1 - \hat{\delta} A_n^s) S_n^s \; ; \qquad t = 1,...,14 . \qquad (9.29)$$

Approximate stock prices, P_{LS}^t, P_{LM}^t, P_{LL}^t and P_S^t, that correspond to the values and quantities defined by (9.22)-(9.29), can be computed in the usual way:

$$P_{LS}^t \equiv V_{LS}^t / Q_{LS}^t ; \; P_{LM}^t \equiv V_{LM}^t / Q_{LM}^t ; \; P_{LL}^t \equiv V_{LL}^t / Q_{LL}^t ; \; P_S^t \equiv V_S^t / Q_S^t , \qquad t = 1,...,14 . \qquad (9.30)$$

Using the above prices and quantities, an approximate stock index of land prices, P_{LStock}, is formed by aggregating the three types of land and an approximate constant quality stock price index for structures, P_{SStock}, is simply formed by normalizing the series P_S^t. The approximate *overall* stock index, P_{Stock}, is obtained by aggregating the three types of land with the constant quality structures (or, equivalently, by aggregating P_{LStock} and P_{SStock}). Since the quantities are constant over all 14 quarters, the Laspeyres, Paasche and Fisher price indexes are all equal.[24] The stock price indexes P_{LStock}, P_{SStock} and P_{Stock} are charted in Figure 9.7 and listed in Table 9.8. For comparison purposes, the corresponding price indexes based on *sales* of properties for the model presented in Section 9.7, P_{L4}, P_{S4} and P_4, are also listed in Table 9.8. As can be seen from Table 9.8, the approximate stock price index for structures P_{SStock} coincides with the sales based price index for constant quality structures P_{S4}, so P_{S4} is not charted in Figure 9.7.

[23] The quantities defined by (9.26)-(9.29), which are constant over the 14 quarters, are equal to 77455, 258550, 253590 and 238476 for small lots, medium size lots, large lots and structures, respectively.

[24] Fixed base and chained Laspeyres, Paasche and Fisher indexes are also equal under these circumstances.

Figure 9.7. Approximate Price Indexes for the Stock of Houses (P_{Stock}), the Stock of Land (P_{LStock}), the Stock of Structures (P_{SStock}) and the Corresponding Sales Indexes (P_{L4} and P_4)

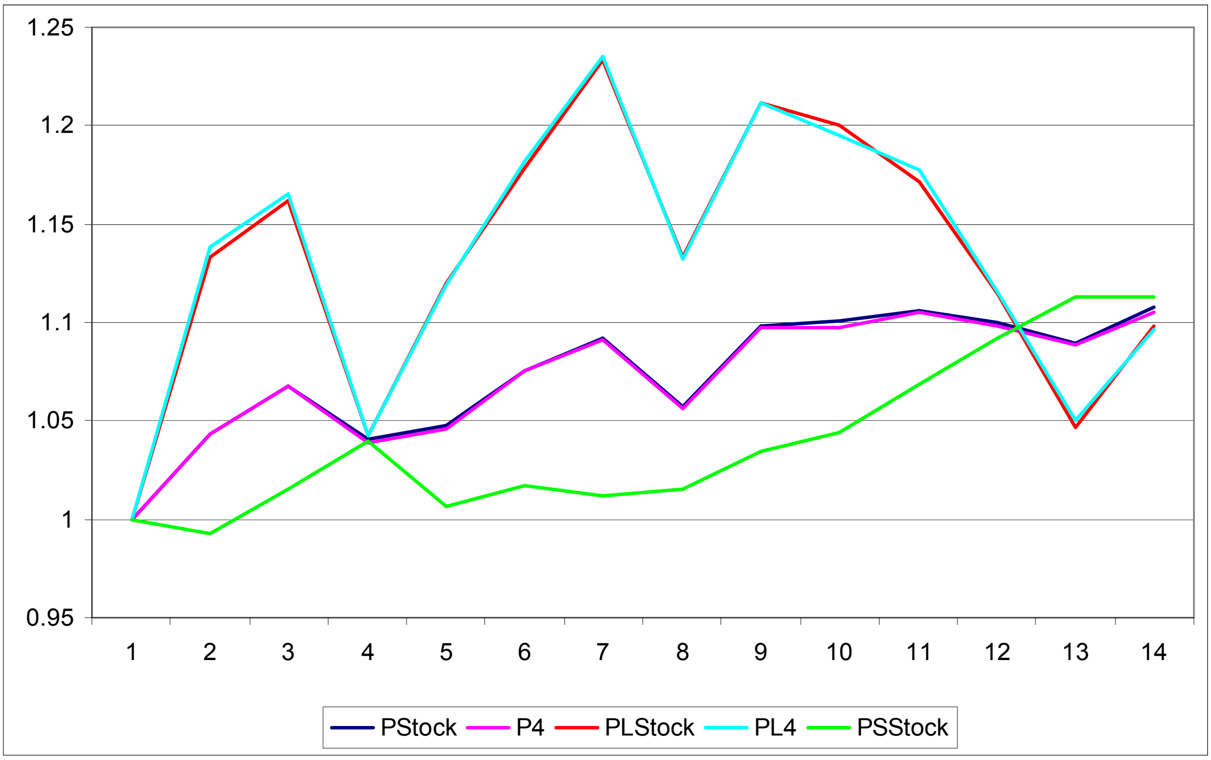


Table 9.8. Approximate Price Indexes for the Stock of Houses (P_{Stock}), the Stock of Land (P_{LStock}), the Stock of Structures (P_{SStock}) and the Corresponding Sales Indexes (P_{L4} and P_4)

Quarter	P_{Stock}	P_4	P_{LStock}	P_{L4}	P_{SStock}	P_{S4}
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.04331	1.04373	1.13279	1.13864	0.99291	0.99291
3	1.06798	1.06752	1.16171	1.16526	1.01518	1.01518
4	1.04042	1.03889	1.04209	1.04214	1.03947	1.03947
5	1.04767	1.04628	1.11973	1.11893	1.00709	1.00709
6	1.07540	1.07541	1.17873	1.18183	1.01721	1.01721
7	1.09192	1.09121	1.23357	1.23501	1.01215	1.01215
8	1.05763	1.05601	1.13299	1.13257	1.01518	1.01518
9	1.09829	1.09701	1.21171	1.21204	1.03441	1.03441
10	1.10065	1.09727	1.20029	1.19545	1.04453	1.04453
11	1.10592	1.10564	1.17178	1.17747	1.06883	1.06883
12	1.10038	1.09815	1.11507	1.11588	1.09211	1.09211
13	1.08934	1.08863	1.04668	1.05070	1.11336	1.11336
14	1.10777	1.10486	1.09784	1.09648	1.11336	1.11336

The overall approximate price index for the total stock of detached houses in the town of "A" (P_{Stock}) can hardly be distinguished from the corresponding overall sales price index (P_4) in Figure 9.7. Similarly, the approximate price index for the stock of land in "A" (P_{LStock}) can barely be distinguished in Figure 9.7 from the corresponding sales price index for land (P_{L4}).

Nevertheless, there are small differences between the stock and sales indexes, as Table 9.8 shows.

Our conclusion is that the hedonic regression models for the sales of houses can readily be adapted to compute Lowe type price indexes for the stock of houses. There do not appear to be major differences between the two index types when using our data set, but this result may not hold for other data sets.

Chapter 10: Data Sources

10.1 Introduction

In practice, because of the high cost of undertaking purpose-designed surveys of house prices, the methods adopted by statistical agencies and others to construct residential property price indexes have mainly made use of administrative data, the latter usually being a function of the house price data sets generated by a country's legal and administrative processes associated with buying a house. The indexes so constructed can vary according to the point in the house purchasing process at which the price is measured. For example, the final transaction price or the earlier valuation used for securing a loan could be used as the "price" of the property. Furthermore, different administrative data sets will generally collect information on different sets of characteristics associated with the sales of the properties. These differing information sets will generally affect index compilation methods, often acting as a constraint on the techniques available to quality adjust for houses of different sizes, locations, etc. Thus data sets have historically acted as a constraint on index construction.

This chapter examines the different sources of data used for constructing residential property prices indexes. Although it focuses mainly on price data, the chapter also considers how the choice of weighting scheme can be constrained by the information generated from the house-purchasing process. Different weighting schemes, notably whether an index is stock or sales weighted, produce price indexes which measure different concepts. In these circumstances it is important that there is a clear understanding of what the target measure is so that the indexes compiled can be evaluated against the target measure to determine fitness-for-purpose.

10.2 Prices

10.2.1 The Process of Buying and Selling a House

The process of buying and selling a property normally takes place over a period of several months or more. The particular stage in this process at which the price is entered into an index will depend on the source of the data and this has consequences for what is being measured and for the comparability of different indexes. Price data for a residential property price index may be taken at the following stages:

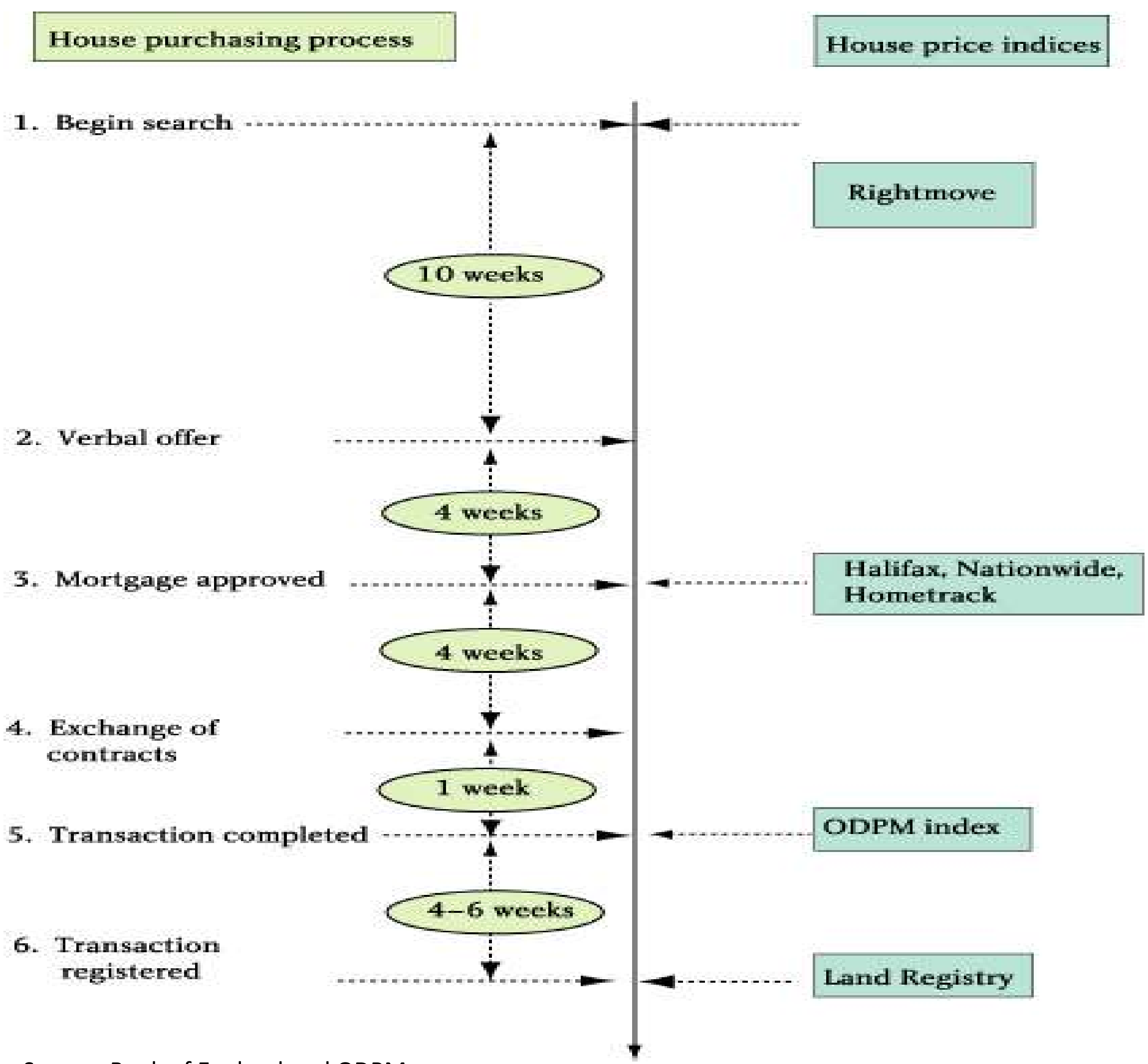
- As soon as the property is on the market. Typical data sources: newspapers, real estate agents.
- Mortgage applications. Typical data source: mortgage lenders.

- Mortgage approved. Typical data source: mortgage lenders.
- Signing of binding contract. Typical data source: lawyers.
- Transaction completed. Typical data sources: land registries, tax authorities.

Each source of price data has its advantages and disadvantages. For example, a disadvantage of advertised prices and prices on mortgage applications and approvals is that not all of the advertised prices will end in transactions, and the price may differ from the final negotiated transaction price. These prices are likely to be available sometime before the final transaction price. Indexes that measure the price earlier in the purchase process are able to detect price changes first, but they will measure final prices with error because prices can be renegotiated extensively before the deal is finalized.

It should be noted that the availability of different sources of price information at different points in the buying and selling process can be an advantage. For instance, changes in the relationship between asking price and selling price may provide an early indication of a change in the housing market. The diagram below illustrates the situation in the UK; see also the case study for the UK in Chapter 11.

Diagram: House purchase timeline and house price indexes



Source: Bank of England and ODPM

Most data sources are susceptible to all the disadvantages of using administrative systems for statistics. The use of administrative data in economic statistics has been associated with four challenges: definitions, coverage, quality, and timeliness – with expected trade-offs against compilation costs. Definitions and coverage are sometimes placed under the one heading of "coverage": to embrace the types of units covered and the degree of coverage. For example, cash sales could be recorded but properties bought with a mortgage may not be covered or some cash sales may not be recorded if, for example, they are under the threshold for tax liability.

The underlying problem arises from the fact that the data are primarily recorded as a step in the administrative process and not as an input into a statistical system. The data are not under the control of the statistician. The inherent weaknesses in administrative data need to be taken into account when using the data and in interpreting the results, most particularly when they are used as a substitute for statistical data rather than as a supplement to or in conjunction with purpose-designed statistics. Some of the weaknesses may be overcome by an appropriate methodology, such as combining complementary data sources, and possibly by using some form of modeling.

A number of basic characteristics come into play in considering the suitability of different data sources.

- *Definition.* This is closely associated with conceptual issues and what the target measure of an index is.

- *Coverage.* Issues relating to coverage will be determined by the operational boundaries of the agency or business providing the housing data. For example, the agency could cover country-wide property sales or just cover a particular region or the transactions covered could relate only to cash purchases or to properties purchased using a mortgage loan. For a government agency, the operational boundaries will be dictated by the regulations and legal processes relating to the purchase of residential property. Inevitably, for public and private data providers coverage will also be heavily dependent on the resources at the disposal of the agency or business and its efficiency in providing data. All these factors are outside the control of the index compiler and can impact on data quality and on any divergence between intended coverage of the residential property price index and actual coverage.

- *Quality.* When considering the issue of data quality, it should be borne in mind that the administrative authority is likely to focus on validating the information which is pertinent to the sale and to the execution of its duties and which reflect the laws and regulations which it is required to comply with. There may be other information which is collected which is of interest to the statistical agency, but which is only of limited relevance to the administrative authority. For instance, this may be the case for some house characteristics which the statistical agency may wish to use for quality adjustment. At the end of the day, the reliability of administrative data will depend on the incentive for data suppliers to give correct information and complete information. There can be mutual advantage to both parties from the statistical agency helping the administrative authority to improve the quality of its data. This can be done by giving feedback on the consistency of data entries and from advising on more general weaknesses. Some statistical agencies provide the administrative agency with an incentive to improve their data collection by compiling custom-designed statistics for the data supplier in return for access to the raw data.

- *Timeliness.* The timeliness of administrative data will depend on who is responsible for reporting to the administrative authority and on the incentive for timely reporting. For instance, there may be a big incentive for a buyer to obtain approval from the mortgage company, for a house loan and for the mortgage company, to quickly get an accurate and up-to-date valuation so that the sale can go through, with all parties safeguarded, before another potential purchaser takes an interest in the property. On the other hand, there may be less of an incentive to register the sale quickly with the official land registry once completed.

One of the keys to the successful use of administrative data is to have an intimate and detailed knowledge of the data collection processes and associated operational systems.

Each source of price data is considered separately below. Where more than one data source is available to the index compiler, the opportunity arises for consistency checks and for data from different sources to be combined. For instance, it may be possible to use the property valuations carried out for the approval of loans to predict the final transaction price recorded much later on by the land registry. This depends of course on the stability of any correlation found between the two.

10.2.2 Seller's Asking Price: Estate Agents, Newspapers, Etcetera

Information on the seller's asking price can be collected through surveys of real estate agents or from an examination of advertisements in newspapers, magazines or online. One of the main advantages of indexes constructed from such information is their timeliness. By taking asking prices, indexes constructed using this information can provide a timelier estimate of house prices than those indexes that are based on subsequent transactions. They also have an advantage over house price indexes based on information from mortgage lenders, as the latter are limited to transactions involving mortgages. However, indexes based on initial asking prices have a major drawback. Houses can be withdrawn from sale and the agreed selling price may not equal the seller's asking price. These indexes ignore reductions in prices that sellers subsequently make, for example when the housing market is on a downturn, or offer prices above the asking price when the housing market is buoyant. Such indexes can therefore present an over-optimistic outlook when the housing market becomes depressed and an over-pessimistic outlook when the housing market is recovering. The fact that they cannot be relied upon to present an accurate picture of the housing market in the short term devalues their usefulness to most users, most particularly those interested in the early detection of turning points in the housing market or an advanced indicator of the future direction of house prices. It should also be noted that the differences between initial asking price compared to actual transaction price also means that the calculation of "average house price estimates" can also be misleading.

Information collected on a seller's asking price cannot always be easily verified and, as well as depending on a balanced and representative sample, relies on the honesty and knowledge of those being surveyed and when drawn from advertisements, the accuracy of the information, especially when it is from a website. For example, it has been argued that real estate agents are more likely to be optimistic about prices and have a vested interest in prices going up rather than down and that this may influence survey results. On the other hand, an estate agent might suggest to a seller an unrealistically low asking price in order to get the property off their books quickly to get the commission. It has also been argued that websites will tend to be biased towards properties that have a competitive asking price to entice potential sellers. All this is, of course, speculation but it does bring home some of the potential difficulties associated with these sources.

Surveys of real estate agents have some inherent advantages over surveys of advertisements. Agency surveys can be based on a more scientifically selected sample and can provide information on a representative selection of those properties on the market, including those which typically are not covered in advertisements. Data from real estate agents might include extensive information on the characteristics of the property and this information is extremely important for quality adjustment (using either hedonic regression methods or stratification methods as was seen in previous chapters). Also the survey questionnaire could collect information on issues such as: what is the average selling time or what has been the recent difference between asking prices and selling prices (e.g. "higher" or "lower") or on the number of potential buyers registering and the number of properties listed with the agent. This information can help put the price information used in compiling the index into context and can be useful for interpretation of the final results. But such surveys typically do not record the asking price of a specific property. Rather, the questionnaire would normally ask the real estate agent to give the "average asking price" for a selection of representative properties.[1] For example, this might be for each of four standard property types (flat, terraced, semi-

[1] Some surveys also ask for "achievable" price and use this to construct a house price index.

detached and detached) in a number of different locations. It is this information which is used to create an average property price for each property type in each location, which is used in turn to compile the corresponding price index. In contrast, the inherent advantage of a survey of advertisements is that the latter will collect the actual asking price for each of the advertised properties.

In summary, although a house price index based on surveys of asking prices may be more timely, the difficulties in determining exactly how the survey information was constructed and the uncertain relationship between asking price and selling price mean that care should taken if such an index is to be used as a barometer of house prices.

10.2.3 The Initial Offer Price Accepted by Seller: Mortgage Companies

Many countries turn to mortgage lenders as the main data source for their house price index. The information is stored in the lender's computer system and serves the operational business needs of the mortgage lenders. This database may include the initial offer price made by the potential purchaser, the valuation price used for authorising a loan and sometimes also the final transaction price. Information from mortgage companies can suffer from all the disadvantages of using data drawn from administrative systems, as described above, but these databases can be a rich source of timely information.

However, data from mortgage lenders suffer from a major drawback: they exclude non-financed home purchases. Research has indicated that cash buyers account for about a third of the UK market and cash buyers tend to purchase either very cheap or very expensive properties. This would not be problematic if it was not for the fact that dwellings purchased for cash can experience different price developments compared to those financed by a mortgage. This is likely to be particularly the case at turning points in the market where different ends of the housing market may be react differently to the economic circumstances and the premium for a cash-buyer increase. For instance in a down-turn, people at the top end of the market who were considering selling their homes to release equity may hold back from putting their homes on the market at a reduced price, so the supply of houses for sale falls and is mainly from owners who, for one reason or another, are very keen to sell. However, at the same time the number of active potential mortgage-based buyers could drop significantly as people are reluctant to take out larger mortgages. But some people will need to sell. In this situation a cash-buyer for a house at the upper end of the market will be in a relatively stronger position to negotiate a bargain price than in a more stable market.

10.2.4 The Valuation Price for a Loan: Mortgage Companies

Mortgage companies will obtain an independent valuation of a property before approving a loan. The valuation that the mortgage company provides the customer with at the time of the mortgage approval can be some weeks after the buyer and seller have negotiated a final price and the buyer has made the initial application for a loan. In practice there is a negotiation process between these two stages in which it is possible for the agreed purchase price of the dwelling to change. This can be the case when the independent valuation differs from the price the purchaser and buyer had agreed upon or where the purchaser has paid for a detailed survey of the property which reveals that substantial repairs are necessary. For instance, it is fairly common for a buyer to try to leverage a price reduction if the valuation by the mortgage company turns out to be significantly lower than the previously agreed price, or if a survey of the condition of the property reveals the need for new roofing. Clearly, the difference between the initial offer price and the follow-up valuation and any process of re-negotiation which takes place subsequently can result in the measured rate of house price inflation to differ from the true rate as measured by the actual transaction price.

The house price change measured by indexes based on valuations by mortgage companies can differ from the price change shown by the offer price and both may differ from the price change based on final transaction prices even when taken from the same sample of mortgage lenders. Thus it is important to understand exactly what an index is measuring.

10.2.5 The Final Transaction Price: Mortgage Companies

The time lag between the mortgage application, mortgage approval and purchase completion stages and the differences in the corresponding values of the house prices illustrate the trade-off between timeliness and accuracy. The final transaction price is not always recorded by mortgage lenders and is often extracted instead from legal records such as entries made in land registers, which additionally also include sales that did not require a mortgage. But there can be a long time lag between the completion of the transaction and the recording of the sale in the land register. One of the main advantages of data from mortgage lenders is its timeliness. Initial offer prices and valuations provide an earlier indication of current prices, as these data are available earlier, and final transaction prices may be available sooner from the mortgage lender than from the land registry. It is for this reason that the exploitation of information from mortgage lenders on final transaction price may be a preferred option. The final transaction price held by mortgage lenders can be easily verified against land registry records to alleviate any concerns regarding accuracy and credibility.

10.2.6 The Final Transaction Price: Administrative Data from Property Registers and Tax Offices

Ideally a house price index would be based on actual transaction prices at the time when the property is sold and the sale completed. The signing of the first binding contract best fits this requirement because of its timeliness but in practice there can be some ambiguity about the point at which a contract is binding, e.g. whether this is at the point where an offer is formally accepted (e.g. when sealed bids are opened), or when a contract is signed or when the contract is exchanged. Similarly, there can be a difference between when a contract is signed and when the transfer of ownership takes place and when it is recorded in the property registers or at the tax office.

In theory, information from property registers or tax offices will cover all properties, including cash purchases as well as purchases via a mortgage and thus these databases should be the most comprehensive of all the sources available to the index compiler. But, in practice, comprehensiveness cannot be guaranteed, particularly if there is a disincentive for the owner to register a property. For example, when the primary purpose of registration is for taxation purposes, properties may not get registered at all, or may be registered with some relevant detail such as square metres of floor space missing or incorrectly recorded, in order to avoid tax or reduce the tax charges.[2]

10.2.7 Valuation Price for Taxation and Payment for Local Services: Tax Offices

In many countries, the central or a local government may impose a monthly or annual tax or service charge on residential properties, for funding the provision of public services such as road maintenance, police and fire services or refuse collection. In many cases, the tax bill faced by an individual is proportional to the assessed value of property and the latter is usually based on a valuation undertaken by professional chartered surveyors either under contract or directly employed by the taxation authority. The valuations should take into account characteristics of the property, such as location and size of plot. However, they rely on accurate information about the properties and also on the chartered surveyors' assessments, which are difficult to verify. Also the updating of the valuations tends to be infrequent due to the field costs involved. Because of these drawbacks, the information collected can sometimes be of limited use in the construction of residential property price indexes. That said, this source of official valuation information has been exploited by statistical agencies; see the material on the SPAR method of index construction described in Chapter 8.

10.2.8 Other Expert Opinion Information: Surveys of Estate Agents Organisations, other Professional Bodies and their Members

In some countries, regular surveys are conducted of real estate agents, chartered surveyors or their corresponding professional bodies, asking about house prices and housing stock. These

[2] There is a related problem: the transaction price may not be a market price because the transaction, while genuine, is between relatives or friends. For example, parents may decide to pass on the family home to their children at a below market price.

"opinion" surveys are typically restricted to asking respondents to give a view on whether house prices are moving up, down or flat. These surveys do not give an indication of how much houses are worth or by how much prices are falling or rising but they can provide an up-to-date and broad-based picture on the direction of price change in the housing market to supplement and help to add credibility to the latest figures from a residential property price index. For instance, a significant change in the difference in the proportions of real estate agents who think prices are going up and those who think prices are going down might provide an early indication of a change in the housing market not yet detected by the currently available statistics on mortgage lender valuations. Contextual information of this kind adds value and is regularly used by commentators when interpreting official house price indexes.

10.3 Evaluation of Data Sources for Fitness-for-Purpose

The overall usefulness of the above sources of information on residential property prices will very much depend on their fitness-for-purpose for the particular applications to which they are being used. To gauge fitness-for-purpose requires an evaluation of the intrinsic advantages and disadvantages of the index against an agreed set of criteria, i.e. an evaluation against user needs.

Chapter 3 reviewed the many different uses of house price indexes: as a macro-economic indicator of inflation; for monetary policy targeting; as a measurement of change in wealth; as a financial stability indicator to measure risk exposure; as a deflator for the national accounts; as an input into an individual citizen's decision making on whether to invest in residential property; as an input into other price indexes, in particular the Consumer Price Index (CPI), and for use in wage bargaining or indexation.

An effective evaluation of the different sources of data on house prices is dependent on a systematic analysis of user requirements. User needs have a significant impact on decisions relating to the conceptual basis of an index and the associated statistical requirement. This may take the form of a series of questions reflecting the different reasons why users may want information on house prices. For instance, whether an index of house prices is to be used as one of a suite of general macroeconomic indicators, as an input into the measurement of consumer price inflation, as an element in the calculation of household wealth or as a direct input into an analysis of lenders' exposure. Such an analysis can then be transformed into a statistical user requirement and an associated conceptual framework by expressing the needs in statistical terms and identifying the common linkages and corresponding relationships at a micro and macro level. The different data sources can then be evaluated against the statistical need.

The following list of desirable properties for a residential property price index constitute a possible set of criteria for an evaluation of alternative data sources for fitness-for-purpose for different uses.[3] The list builds upon the discussion at the beginning of this chapter. The relative importance of each of the criteria will depend on use and in essence constitutes a statistical requirement. There will also be the usual trade-offs between fully meeting user needs and the costs of data collection.

10.3.1 Definitions and Measurement Concept

This also covers *coherence* with other statistical outputs. It represents the user requirement at the most basic level.

Consider the needs of governments and analysts looking at inflationary pressures and those with a direct investment in real estate. The primary focus of these users may be the cyclical nature of prices and the ability of real estate prices to lead to destabilising booms and slumps in the economy as a whole. For this purpose, users will be looking to a variety of indicators, including indexes of the volume and price of real estate transactions as well as macro-economic indicators for modeling the economic cycle and predicting peaks and troughs. Analysts looking at the inflationary pressures of real estate price rises in comparison to other price rises may be interested in including in a CPI the inflationary costs of owner-occupier housing costs by means of a house price index based on the net acquisition cost basis but excluding land.

For users wanting a general macro-economic indicator, an index based on all purchases – both cash and those with a mortgage – is appropriate. Taking transaction prices solely from data supplied by mortgage lenders represents a serious deficiency. Conceptually, land registry data would represent a better source as it should cover all transactions. The challenge is to find a source of price data which readily fits, or can be manipulated to meet, the requirements of users interested in the inclusion of owner-occupier housing costs in a CPI on a net acquisition cost basis, that is, excluding the price of land.[4]

In contrast, users interested in an analysis of the current value of the real estate portfolio against which outstanding mortgages are secured, will require an index of changes in the price of the properties for which mortgages were issued, weighted by the amounts loaned for each type of property at the time at which they were issued. For both of these measures, the value of the land underlying the buildings is as important as the value of the buildings themselves

[3] See also section 4.5 of Chapter 4 where a listing of user needs is presented based on discussions between users of house price indexes and the Office of National Statistics. In that section, it was pointed out that there is a trade-off between the desires of users to have a family of more detailed indexes (stratified by location and type of housing) and the quality of the indexes: more detail inevitably leads to less accurate indexes.

[4] In most countries for most transactions, land and building are purchased together as a "single package", so the two components are typically not separated in the information generated by records relating to the transfer of ownership. As such separating the prices would require a supplementary exercise. In Chapter 9 it was outlined how hedonic regression can be used to decompose the overall price index into land and structures components.

and it is the total value of the land and buildings which is of interest. For these users, data from mortgage providers on property prices and the size of new mortgages and outstanding debt will fit the purpose.

Now consider the needs of employers and trade unions when negotiating wage settlements. Their primary focus will be the effects of price changes on the standard of living of workers. For this purpose users will be looking to a CPI that includes the cost of keeping a roof over their heads – for owner-occupiers the cost of mortgage interest payments and the repairs costs. The measurement of this will require the calculation of the mortgage outlay at time of purchase and the subsequent repayment history will need a sales weighted house price index. In an ideal world re-financing would be excluded. The repairs element may be measured by the calculation of depreciation. For this, a stock-weighted smoothed house price index is most appropriate. In addition, there is the issue of land where it is often argued that in most circumstances land is an investment which appreciates and that its inclusion in a depreciation calculation is inappropriate.[5] Thus an index excluding the price of land may be required.

For the calculation of mortgage outlay, the user can again rely on information supplied by mortgage lenders, but not for the estimation of depreciation, where the value of land may again need to be separately identified.

As a final example, consider the needs of national accountants, who are seeking appropriate deflators for national accounts. Their needs again will be different. Real estate appears in the National Accounts in several ways (see also Chapter 4, section 4.2):

- The imputed rental value received by owner occupiers for buildings, as opposed to land, is part of household final consumption.
- The capital formation in buildings, again as opposed to land, is part of gross fixed capital formation, depreciation, and the measurement of the stock of fixed capital.
- Land values, which are an important part element of the national stock of wealth.

In each case the derivation of volumes from values requires price indexes for respectively: the imputed rent of owner occupied dwelling units weighted by the stock of different types of owner occupied housing; new house purchases weighted by the transactions in new houses but excluding the land component; and of the whole housing stock including land weighted by the housing stock

It can be seen that user needs will vary and that in some instances more than one measure of house price or real estate inflation may be required. It can also be seen that coherence between different measure and with other economic statistics is important and that achieving this will

[5] There are other more general issues, which are not addressed here, to do with the measurement of depreciation and its inclusion in a consumer price index.

be especially difficult as statisticians are unlikely to have an ideal set of price indicators available to them.

10.3.2 Coverage

Coverage includes not just whether all properties are covered irrespective of whether the property is owned outright or being funded by a mortgage but also whether country-wide property sales or valuations are covered or just those in a particular region and whether all price ranges are covered. It can be noted that even where the primary need is for a national index, regional indexes can be in demand for analytical purposes.

House price information from any individual mortgage lender is unlikely to be representative of the country as a whole, not only because of the exclusion of cash purchases but also because lenders often focus their business on particular regions.

10.3.4 Quality

Quality relates to the accuracy and completeness of the information, i.e. there are no serious errors and the information is what it purports to be.

Compared with other administrative data, house price information from a land registry is likely to score relatively highly in terms of accuracy due to the legal requirements to record property transactions and exchanges of ownership. However, the reliability of data from any administrative source is difficult to validate.

10.3.5 Timeliness

Indexes that measure prices earlier in the purchasing process are able sooner to detect price changes and turning points in house price inflation. This is likely to be particularly important when used, say, for macro-economic policy and monetary targeting but less important for a national accounts deflator.

Data from mortgage lenders may better suit the needs of those engaged in macro-economic policy and monetary targeting, even though cash purchases are excluded, whilst land registry data may better suit the needs of, for example, those calculating deflators.

10.3.6 Detail for Quality Adjustment and Mix-Adjustment

This relates to two (related) issues: the degree to which residential property price indexes are able to adjust for changes in the mix of properties sold and to eliminate the effect of quality changes of the individual dwellings. For this purpose, “real time” information is needed on price determining attributes such as size of plot, size of house, type of property (flat, house, semi-detached or detached), location, the condition of the property, whether it has central heating, a fully-fitted kitchen and bathroom, etc. Quality (or mix) adjustment is essential in order to construct an accurate price index for housing components.[6]

It is unlikely that any of the sources of prices data listed above will be ideal for all purposes. The amount of detailed and relevant characteristics data will depend on the individual data set.[7]

10.3.7 Frequency

Frequency essentially relates to how frequently an index can be computed, e.g. once a month or once a quarter. There is a tradeoff between frequency and accuracy. For a particular geographic area and type of housing, current information on the price of houses in a given strata will come from sales of old and new houses in that strata during the chosen time period. If the frequency is chosen to be a month as opposed to a quarter, the monthly sample size will only be approximately one third of the quarterly sample size. Thus a monthly house price index based on sales of properties in the given strata will be subject to increased sample volatility (and hence will not be as accurate) as compared to the corresponding quarterly index. Volatility of a monthly index *may* be reduced by making the strata "bigger",[8] e.g., different neighbourhoods could be combined within the same general location but this leads to another tradeoff between fineness of the strata (which many users may want) and accuracy of the index (which all users want).

It may be possible to provide smoothed monthly house price indexes that are say a three month moving average of the raw monthly indexes[9] or the statistical agency could provide both monthly and quarterly indexes and let users choose their preferred index.[10]

---

[6] The various methods available for constructing quality adjusted house price indexes were discussed in Chapters 5-9.

[7] In cases where the real estate agent data base includes the final selling price of the listed properties along with the main characteristics of the properties, this information base is probably the "best" for most purposes. However, the sample of listed properties needs to be compared with the properties listed in land registry offices to ensure that the coverage of listed properties is adequate for the purpose at hand. When constructing price indexes for the stock of housing, it will be necessary to have census information on housing stocks along with post census information on demolitions and the construction of new dwelling units.

[8] It is not certain that combining strata will reduce index volatility if house prices in the different micro strata have different trends.

[9] The Australian Bureau of Statistics makes frequent use of this technique for a wide range of its statistics. If the window length is 12 months, then the resulting smoothed index can be regarded as a seasonally adjusted index, centered in the middle of the 12 month period under consideration. For a variant of this smoothing technique, see section 5.6 in Chapter 5.

[10] There is a possibility that some users may be confused by having more than one index covering essentially the same housing strata. However, the Bureau of Labor Statistics now has two monthly published Consumer Price Indexes: their headline Lowe type CPI which is not revised and a second index which is an approximation to a

It is not possible to provide definitive advice on how frequent a house price index covering a certain stratum should be published. The issue of frequency must be decided by the national statistical agency, taking into account user needs and data availability.

10.3.8 Revisions

Revisions can refer to either revisions resulting from subsequent returns (so that the series itself is revised) or from other sources of more relevant data subsequently coming on stream (so an early indicative measure is eventually replaced by a precise measure of what needs to be measured).[11] For instance, an example of the former might be revisions arising from late registration of property sales. An example of the latter might be where an initial offer price recorded on the mortgage application form is used as an early indication of movements in transaction prices but is subsequently discarded when land registry data on actual transaction prices (which takes into account any price renegotiation before the sale is finalised) eventually comes on stream at a much later point.

The extent to which figures are revised due to the receipt of subsequent returns is partly determined by the reference point of the prices data and partly by the point in time when the particular data set is received by the statistical agency: the earlier is the data reference period in the purchasing cycle and the earlier the particular data set is received, the more the index will be subject to revision. Thus, although information from the registration of property sales is appropriately referenced and provides a definitive source of information on property prices, the time delay that can sometimes take place in some countries for the legal registration of property transfers can mean that the register is not final until, say, twelve months the sale of the property.

Valuation prices kept by tax offices for taxation and payment for local services and the final transaction price recorded by mortgage companies are least likely to be subject to revision, whilst the final transaction price based on administrative data held on property registers and tax offices could be subject to revision over a long period depending on the time-lags involved in the legal processes of recording changes in ownership.

10.3.9 Comparability

---

superlative Törnqvist index (which is revised). Users in the U.S. seem to have accepted multiple indexes in this context.

[11] A related issue is that some of the methods for constructing an RPPI, such as the multiperiod time dummy hedonic method (see Chapter 6) and the repeat sales method (Chapter 7) suffer from revision in the sense that previously computed figure will change when new data is added to the sample. In some cases, revised indexes are published while in other cases, the rolling window technique with updating due to Shimizu, Nishimura and Watanabe (2010) and Shimizu, Takatsuji, Ono and Nishimura (2010) and explained in section 9.8 of Chapter 9 is used. The rolling window with updating technique does not revise the historical index up to the current period.

Comparability refers to the degree of *inter-country comparability* between house price indexes. This is important because comparing house prices from non-harmonised national data can be problematic as differences in concept, index construction, market coverage, quality adjustment procedures, etc. can make cross country comparisons difficult. Differences in frequency, timeliness and revisions policy can also cause comparability problems.

Problems can arise at both the national and international levels:

- Users in individual countries can be confronted either with a lack of relevant statistics or with different statistics for different time periods and with varying time-lags and these statistics can be based on different data sources or compilation methods.
- For users seeking international comparisons the situation is complicated by significant differences among countries with regards to the availability of data and the challenge this represents for compiling like-for-like comparisons and interpreting relative trends among countries. The complication of aggregate price indexes covering groups of countries – a requirement for co-ordinated economic policy and monitoring across an economic area such as the Euro zone (those countries in Monetary Union) – is a further challenge.

From Chapter 11, it can be seen that the methods employed for the compilation of residential property price indicators vary considerably *between* countries, and even between alternative sources *within* individual countries.

10.4 Weights

The data sources drawn on for the weights in a residential property price index are a function both of the data needs of the target index and of the availability of the required information. Also the data needs depend not only on the conceptual basis of the index but also on detailed aspects of index construction, such as the method of quality adjustment and any subindexes that are required for analytical and other purposes. For instance, the construction of a mix adjusted property price index based on transactions requires that enough information is known about the sales in each period for them to be classified into groups sufficiently homogenous so that the unit values can be treated as prices. In the housing market, the problems are compounded by the low volumes of sales for certain house types in particular geographical areas which could lead to many cells being empty.[12]

Putting these detailed issues of construction to one side, the conceptual basis of the index is the main factor determining the data needs relating to weights. One price index cannot meet the diverse needs of users. For estimating net capital formation, for instance, only new houses

[12] The stratification or mix adjustment method was discussed in Chapter 5. In the example in section 5.5 for the Dutch town of "A", many cells were indeed empty. A "matched-model" approach was suggested to cope with this problem.

should be included while estimating the effect of price changes on capital stocks requires the index to cover all transactions.

The weights can be derived from a number of sources, in particular, from national accounts data, periodic national censuses which collect information on the housing stock, information from banks on the loans taken out for house purchase, construction statistics, official registers recording ownership, etc. There can be a lack of coherence between these different data sources resulting from the long and quite often involved processes associated with buying and selling a house and the fact that a valuation or offer price associated with an application for a mortgage will not necessarily lead to a sale and change of ownership. Other issues arise also, such as the distinction between what is being built for selling and what is being built for renting out. This sort of information is rarely readily available from one statistical source. It is for this reason that the construction of weights may draw on a multitude of different sources.

10.5 Developing Countries, Traditional Dwellings and the Informal Housing Market

For many developing countries, a significant proportion of the housing stock consists of newly constructed buildings on family owned land or of old buildings which have been significantly upgraded since they were first constructed. There can also be a significant element of owner-constructed housing. Construction may take many years and at any point in time a substantial proportion of the houses could be considered incomplete. The use of formal mortgage finance is often very limited but informal finance may be used. House construction can vary from shanties built on compacted soil with salvaged materials to substantial multi-room dwellings built on concrete foundations with concrete blocks. Amenity levels can vary from virtually none to the elaborate. Housing mobility, particularly with owner-constructed dwellings, is usually very low and consequently the markets for rental or sale of owner-constructed houses are limited and there is very little movement between the two. In principle the compilation of a house price index is the same for owner-constructed housing as for third party constructed housing, but the measurement problems are, at the least, different and are generally more difficult.[13]

The above complications mean that formal records will rarely be kept of the cost of building the new dwelling or of upgrading an old house, for example, by incorporating running water, an internal WC or additional rooms. Formal transfers of ownership sometimes do not take place, formal valuations are often not available and methods of financing can be informal

[13] In particular, the important price determining characteristics of the structure can be quite different for a developing country than for a developed country. In a developed country, there is perhaps less variation in the type of construction and the materials used whereas the quality of shanties could differ more markedly. Also land title may be missing in many instances in developing countries which again creates problems for mix adjustment and hedonic regression techniques for adjusting housing quality.

through the family or may simply not be recorded or records not kept centrally. Thus in these circumstances it will not be possible to calculate mortgage interest payments (including or excluding notional interest payments to relatives), or to estimate net acquisition costs.

The lack of such basic information often means that the rental equivalence or an imputed rent approach is the only practical option for constructing a housing price index. The price indicator for imputed rents can be derived either from a readily available price series for rents, re-weighted to reflect the current composition of the stock of owner-occupier housing, which can then be applied to the rental equivalents in the base period, or from asking an expert to provide on a monthly basis the equivalent rents for a sample of houses which is representative of the owner-occupier housing stock.

In each case, stratification by type of dwelling (house or flat), location (region or area, urban or rural), plus other characteristics which will influence rent is important so that the rents data can be combined to reflect the composition of owner-occupied property. Other stratification variables may include such things as the total size of the plot, floor area and number of rooms, whether there is mains water, an internal WC and mains electricity, the material used in construction and whether the building is of traditional design. The price statistician should seek the advice of an expert active in the field of renting domestic property, such as a housing corporation, to ascertain the most important rent-determining characteristics and should bear in mind the need to keep these to a manageable number. Weights information can be derived from the latest Housing Census or Census of Population and Housing. In practice this information may not be up-to-date due to the change in the owner-occupied housing stock which can occur in the time period between censuses. Where this is the case special surveys may need to be conducted or, particularly in urban areas including townships, use made of planning applications to update the latest census.

But the measurement problems can be significant. In summary, traditional or informal dwellings are generally built by family members or other unpaid labour. The walls can be made of less durable materials such as dried clay, bamboo or latticework and the roofs can be made from reeds, straw or palm fronds or corrugated iron. The dwellings may or may not have electricity or piped water in the dwelling, let alone other facilities. Traditional dwellings are generally located in rural areas. Some associated complications when attempting to include the owner-occupier housing costs in a consumer prices index are:

- Many such dwellings are located in or very near to large cities, such as shanty-towns. These dwellings may be rented or owner-occupied and it may be difficult to obtain details of ownership. Conducting surveys can be problematic.
- There are many such dwellings in rural areas that may be built with family labour on family or unregistered land or land in “common” ownership.

In these circumstances, the concept of "ownership" becomes a grey area. Thus the definition of owner-occupied housing and what a family actually own is subject to debate and even when there is an agreed upon definition, even basic records of the number of such owner-occupied housing may not exist let alone details of the dwellings.

Relevant characteristics for the computation of a price index, that are encountered in traditional and other dwellings in the informal market include:

- Electricity supply. This will often be electricity supplied by a generating or distribution company. However, electricity may also be generated by the household itself, e.g. from a diesel generator or wind power or be taken illegally from the distributor.
- Running water. This may be piped into the dwelling itself or the dwelling takes water from a communal standpipe or well.
- A private or communal toilet which may be either a water-flushing WC-type or a chemical toilet.

In addition there is, as with any home the issue of living space, recorded in terms of number of rooms, m^2, or both. For this there need to be relevant definitions. In particular, definitions of usable floor space (the floor area of the living room, kitchen, hall, bathroom and all adjoining rooms minus the wall thickness and door and window recesses and excluding e.g. stairs) and of the number of rooms (e.g. to whether to include or exclude hall-ways) are required.

Finally, even if information on the characteristics of these dwellings is available there may not be an "equivalent" rental unit to value the services of an owner-occupied unit. Thus the *indirect* measurement of prices may not be possible. In this situation, statisticians can put a system in place to measure input prices (construction costs) and then use this information to construct a user cost measure of the housing services as a proxy for the prices of the housing services consumed.[14] For own-account consumption, the System of National Accounts 1993 (SNA 1993) recognises that it may only be practicable to measure input prices.

The issues discussed above are considered in the case study on the compilation of residential property price indexes in South Africa, which can be found in Chapter 11.

[14] See Blades (2009) for additional material on constructing these user costs for traditional housing in developing countries.

Chapter 11: Methods Currently Used

11.1 Introduction

The methods used for constructing indexes of residential property prices can be constrained in large part by the nature of the data available. The data required to construct the target index, once defined, are not always available on a regular and timely basis, if at all. Moreover, even where suitable data are available to construct an index to meet the needs of one set of users, more often than not, the data does not fit the requirements of another set of users. For many countries setting up the required infrastructure and procedures for the collection of the data necessary for producing a property price index can sometimes be prohibitively costly. Also, changes in methodologies and in the underlying data sources can frustrate the construction of an historical series, which are often required for econometric modelling to inform policy options for the management of the economy. Last but not least, the timeliness and frequency of the data, when available, may not be suitable for producing the kind of house price index that the users want or need.

For users, this data shortcoming for the construction of house price indexes and related indicators has sometimes been a source of frustration. For example, the then Governor of the Bank of Canada, in a speech to the Conference of European Statisticians (Dodge, 2003) stated:

"Given that the investment in housing represents a big chunk of household spending, and that for most people their homes represent their most valuable asset, it is surprising that in many countries there are no comprehensive, quality-adjusted data on housing prices or rents".[1]

In addition, the data sources and the methods used to construct indexes are not always well documented, and surveys of meta-data on residential property prices confirm that there is a lack of harmonisation in the practices. This represents a further challenge for users. In particular, it compromises the possibility of making meaningful international comparisons of trends in house prices and makes any comparative economic analysis extremely difficult. This can bring into question the credibility of the results.

Data availability apart, the methods used by countries to compile residential property price indexes have also to confront some inherent problems, most particularly, that properties have unique characteristics, resulting in heterogeneity in different dimensions, many of which are difficult to measure objectively, and that transactions of individual properties are infrequent. Both of these issues make the compilation of price indices especially challenging. In addition,

[1] The position has improved since then. See the case study for Canada in section 11.3.1 below.

the fact that asking prices are negotiable means that the transaction price may differ from the initial or final asking price, the offer price and an expert valuation.

The identification of the techniques most widely used in compiling indexes of residential property prices also begs the question of whether international best practice in the methods for constructing such indexes can be identified, or whether the techniques adopted inevitably are governed and dependent on local conditions.

Other sections of this handbook provide recommendations on best practice. This chapter describes the range of available indexes by different countries and also presents some case studies. It relies on meta-data gathered by various organisations, including the European Central Bank and the Bank for International Settlements, and more recently, by a fact-finding exercise conducted by Eurostat in connection with the inclusion of owner occupied housing costs in the European Union's Harmonised Index of Consumer Prices.

Meta-data on residential property price indexes published by different countries are available from the website of the Bank for International Settlements (BIS); see www.bis.org/statistics.[2]

11.2 Index Availability

It can be seen from the available meta-data on the BIS website[3] that the methods used to compile residential property price indexes vary considerably, both *among* countries and even *within* individual countries. The latter raises a key question for users with regard to which series should be used to meet their particular needs. With regards to the former, a key issue is raised for users about the validity of available international comparisons.

The differences between the available house price indexes cover almost every aspect of price index construction. These have been referred to in earlier chapters: the conceptual basis of index (i.e., what is the appropriate target index to suite each user need); data sources (property registrations, tax records, mortgage applications and completions, real estate agents, print media such as newspapers and other forms of advertisements); market coverage (geographical coverage, type of property, mortgage/cash transactions); quality adjustment (hedonics, mix-adjustment) and weighting (stock or sales weighted). The problems caused by these different factors can be exacerbated by the fact that housing markets can be highly heterogeneous. Thus not only do properties vary in price according to their physical attributes such as floor area

[2] The property price statistics on the BIS website include data from thirty-seven countries and are available at different frequencies. The data differ significantly from country to country, for instance in terms of sources of information on prices, type of property, area covered, property vintage, priced unit, detailed compilation methods and seasonal adjustment. This reflects two facts. First, that the processes associated with buying and selling a property, and hence the data available, vary between countries and, second, that there are currently no specific international standards for property price statistics.

[3] See hhtp://bis.org/statistics/pp.htm.

and whether they are detached houses on their own plot of land or an apartment in a high-rise complex. The prices can also diverge widely depending on, for example, the region of the country, the area of the town or whether the location is classified as rural or urban. Location affects desirability which leads to different demand conditions, thus explaining why an otherwise identical house may have a different price depending on its location. For instance, a property in a region with a high GDP per capita and low unemployment and in a locality known for the quality of its schools and pleasant surroundings will command a higher price than an otherwise identical property but in an area plagued by high unemployment, low household incomes, poor quality schools, and a high crime rate.[4]

An overview of the current situation is presented below. It should be noted that the position is changing as more countries develop their residential property price indexes and review the indexes currently published. The reader should refer to the information from the BIS website mentioned above for more facts about the residential property price indexes for a particular country.

11.2.1 Responsibility for Compilation

In the majority of the OECD countries, residential property price indicators are calculated by national statistical institutes. In the EU, this is the case in Belgium, Bulgaria, Canada, the Czech Republic, Denmark, Germany, Estonia,[5] France, Latvia, Malta, Norway, Hungary, the Netherlands, Poland, Slovenia, Spain, Sweden and Finland. National central banks compile house price indexes in Belgium, Cyprus, Germany, Greece, Italy, Luxembourg and Slovakia. In Austria, the national central bank works jointly with the Vienna University of Technology, while the price index compiled by the central bank of Luxembourg is based on the data from the national statistical institute. In Ireland, France, Spain, the UK and the USA, residential property price indexes are compiled by government departments other than the statistical office. In some instances, such as in the UK, this reflects in part the fact that the statistical system is decentralised with government statisticians located in government departments and working alongside their policy and service-delivery colleagues. In some cases, responsibility for the compilation of the index resides with the department which has policy, operational or legal responsibility for the housing sector. The latter is the case with the Federal Housing Finance Agency in the USA, for example, and in the UK. The government department with policy or operational or legal responsibilities for the sector is often in a better position to gain access to administrative information for statistical purposes and should also be well-informed about the sector and may even have access to additional useful background information.

---

[4] See for example Chiodo, Hernandez-Murillo and Oryang (2010).

[5] The national statistical institute of Estonia calculates a pure average transaction price per square metre for a specific type of flat in Tallinn.

11.2.2 Data Sources[6]

In Bulgaria, Canada, the Czech Republic, Estonia, Ireland, France, Latvia, Luxembourg, Poland, Spain and the USA, data on residential property prices are collected by the statistical institutes or ministries. The source of official residential property price indexes in Denmark, Lithuania, the Netherlands, Norway, Finland, Hong Kong, Slovenia, Sweden and the UK is information gathered for registration or taxation purposes. In Germany, the Federal Statistical Office collects prices from the local expert committees for property valuation. The statistical institutes in Spain and France calculate price indexes from information provided by notaries. In Belgium, Germany, Greece, France, Italy, Portugal and Slovakia, real estate agencies and associations, research institutes or property consultancies are the sources of data. Data from newspapers or websites are collected for the compilation of residential property price indexes for Malta, Hungary ("Origo"), Austria ("Austria Immobilienbörse") and Romania – the latter are merged by the central bank of Romania with expert data from the Chamber of Notaries. The Central Bank of Cyprus combines data from the private sector (real estate agents and commercial banks) and data from the Department of Lands and Surveys. The limited number of cases of integration of different data sources to add value and produce a better index is interesting given the number of countries that report multiple sources of information on property prices. In Ireland and the UK, residential property price data are, inter alia, provided by mortgage lenders. The index compiled by the UK's Department for Communities and Local Government is based on a mortgage survey conducted by the Council of Mortgage Lenders – the long time-lag associated with the registration of property ownership transfers, undermines the use of the latter as a timely indicator.

Comparability between indexes can be very limited as a result of the different data sources listed above – mortgage versus cash purchases; urban versus rural prices; the prices of old properties versus new properties; valuations versus advertised prices versus initial offer prices versus final transaction prices. The net result is that published indexes can, in practice, measure very different aspects of the price development in the housing markets.

The deployment of different data sources and compilation practices, and the use to which the index is put (i.e., the index purpose) all explain the wide variation both in timeliness and in revisions policy.

11.2.3 Index Methodology

The inherent difficulties with price measurement and the varying data sources used, lead to an array of different methodological approaches being adopted in the construction of house price indexes.

---

[6] See also ECB (2009).

11.2.4 Quality (Mix) Adjustment

Quality adjustment, to control for compositional changes (mix-adjustment) and for changes in the quality of the individual properties, is an essential part of index methodology. It ensures that price comparisons are on a "like with like" basis and avoids the possibility of bias in the series when, for instance, the quality of the housing stock is improving as a result of, amongst other reasons, renovations to the dwelling, which can take various forms, such as the modernisation of kitchens and bathrooms, the introduction of improved insulation and central heating or air conditioning systems. Quality adjustment techniques also play an important role in the compilation of house price indexes because houses that come onto market will change from period to period.

Quality adjustment is applied in a number of different ways. For instance, Estonia uses *unit values* for compiling its residential property price index – that is the average transaction price per square metre of floor space (in this particular case, the sum of the value of all real estate transactions divided by the sum of the square metres of floor space of all real estate sales, with outliers excluded). But unit value indexes based on price per square meter of structure floor space, whilst adjusting for the size of the dwellings in each period, does not adjust for differences in the quality of construction or the age of the structure and perhaps more importantly, does not adjust for changes in the mix of plot sizes in the sample of properties sold in any particular period. Other changes to the features of the dwelling can potentially occur which, together with more general trends in the housing market, are reflected in compositional changes to the sample such as location, physical and environmental amenities, the general quality of housing, etc.

The main alternative of *mix-adjustment* (discussed in Chapter 5) utilises a classification of dwellings by what are generally recognised as important price determining characteristics to calculate individual price indexes for each cell in the classification matrix. The overall index is then calculated as the weighted average of these sub-indexes. Mix-adjustment is in essence a form of stratification. This method is adopted by, e.g., the Australian Bureau of Statistics to control for compositional change to compile quarterly house price indexes for each of the eight capital cities. Their approach stratifies houses according to two characteristics: the long-term level of prices for the suburb in which the house is located, and the neighbourhood characteristics of the suburb, as represented by the ABS Socio-Economic Indexes for Areas (SEIFA)[7]. In practice, the number of characteristics included in the classification is often limited by the number of observations that can regularly be found for each cell, i.e. by the ability to populate the "price-determining characteristics database" from the available data sources as well as by the availability of information on price-determining characteristics.

[7] See http://www.abs.gov.au/ausstats/abs@.nsf/mf/6464.0.

The most sophisticated form of quality adjustment used by countries is the *hedonic regression* approach (discussed in Chapter 6) which uses a regression model to isolate the value of each of the chosen characteristics and control for changes in the characteristics of the properties sold. But this method is usually more data intensive. It is sometimes used in conjunction with stratification (by type of structure and location). The use of hedonics in the compilation of indexes of residential property prices is, in large part, a fairly recent innovation. Countries which publish indexes that have been compiled using hedonic regression include Finland, France, Norway and the UK. The hedonic model used in the compilation of the Norwegian house price index includes only a few explanatory variables and does not adjust for housing standards and for the age of the building;[8] the index adjusts only for size and location of the dwelling. The index is likely to be biased (unless the age structure and type of dwelling sold remains stable over time). This shortcoming is acknowledged by Statistics Norway.

An additional method used in, for example, the USA and Canada, is the *repeat sales* method (described in Chapter 7); i.e., the Case-Shiller home price index in the USA and the Teranet - National Bank House Price Index ™ in Canada. This approach matches pairs of sales of the same dwellings over time. It requires a huge database of transactions and is not used by any of the European index compilers.

It is interesting to note that in Germany, data collection is limited to "good quality" dwellings, which might imply that the issue of quality adjustment is by-passed. In practice, there could be a built-in measurement problem, since it is unlikely that the market definition of "good quality" is independent of the general increase in housing standards over time. For this reason there is potential for bias in the resulting index in the longer term. This is in addition to any concerns about sampling and, in particular, the capability of "good quality" housing to be able to represent the price trend of all houses.

It can be seen from the above paragraphs that two crucial questions for all quality adjustment procedures are: (1) whether the chosen characteristics used for quality adjustment are the main determinants of price differences, and (2) whether the application of different techniques to the same data set will produce the same results (i.e., the issue of statistical robustness). In reality, while some of the price-determining characteristics – such as the size of the living area – are easy to measure, other important factors such as location[9] and the quality of construction, can be inherently difficult to capture and measure. Also, it should be noted that

[8] As was seen in previous chapters using the data for the town of "A", the age of the structure is an important price determining characteristic.

[9] The physical location of a property can be measured rather precisely but the problem with "location" is one of *grouping* of properties. Stratification and hedonic regression methods need to group together sales of properties in the same location but how exactly should the boundaries of a location be determined

the application of different quality adjustment techniques to the same data set will not necessarily produce the same results.[10]

11.2.5 The Value of Meta-Data

A number of organisations have websites providing meta-data on the residential property price indexes published by different countries. Most particularly, the Bank for International Settlements provides such information (see the earlier reference). This is in addition to any information provided by individual countries on, for instance, the websites of the national statistical institute or central bank.

As well as providing the user with guidance on the strengths and weaknesses of a particular price index and its appropriate use, a systematic and more detailed analysis of the meta-data on the currently available statistics and data sources can help to identify:

- major gaps in data provision;
- options for filling these gaps cost effectively from readily available sources;
- data coherence issues;
- the scope for further data integration and the need for new data sources.

Such an analysis of the basic meta-data also provides evidence of the compromises made in relying on readily available data and where one all-purpose house price index is used for a multitude of purposes. For example, the main official house price index published in the UK by the Department for Communities and Local Government (DCLG) uses sales weights and is appropriate for inclusion in, for example, a Consumer Price Index used for indexation of benefits but does not fully suit the needs of users who want to calculate "wealth", where stock rather than expenditure weights are most appropriate. The latter may be addressed either by a re-weighting of the official index or by reference to one of the many indexes published by lenders. However, the latter suffer from limited coverage. Thus re-weighting of the official index may provide a cost effective solution to filling this particular data gap.

A more detailed gap analysis may point to solutions involving synthetic estimates, based on the integration of data from different sources. For example, it can be noted in the context of the UK that the DCLG house price index referred to above has the advantages of being timely and not subject to revision but has the drawback that it excludes cash purchases.

A systematic approach to the construction of indexes of residential property prices in the UK might conclude that it is possible to supplement the official index with information on cash

[10] This point is illustrated by the differing indexes that resulted from the application of different methods of quality adjustment described in Chapters 5-9 above using the same data set for the town of "A". However, all of the methods did result in roughly similar trends in prices.

purchases from the Land Registry. Although the latter is less up to date due to the time-lag in registering transactions in the official registry, time series modelling may be able to address this misalignment. The Land Registry constructs a repeat sales index by tracking the average growth in house prices using multiple transactions associated with the same home in an attempt to hold quality constant.

In the next section a series of case studies are presented relating to the residential property price indexes published in a selection of countries.

11.3 Case Studies

11.3.1 Case Study: Canada

In Canada there are four house price indexes that are currently available. These are Statistics Canada's New House Price Index, the Teranet-National Bank House Price Index ™ National Bank Composite House Price Index, the Canadian Real Estate Association's measure of average house prices, and the Royal LePage Survey of Canadian House Prices. Each one will be explained in turn.

The New House Price Index

The New Housing Price Index (NHPI) is a monthly price index that measures changes over time in the builders' selling prices of new residential houses. Prices that are collected are from a survey of builders from various areas of the country. It is a constant quality price index inasmuch that the features and characteristics of the units in the sample are identical between successive months; in other words, the NHPI is a matched-model index. Separate estimates provided by the builder about the current value (evaluated at market price) of the lots are also an important part of the survey. Consequently, given this information, Statistics Canada also publishes an independent price index series for land excluding the structure. The residual value (total selling price less land value), provides an indicator of the trend in the cost of the structure and is also published as an independent series. At the present time, the three variants of the NHPI are published for 21 metropolitan areas in Canada.

Housing market analysts, academics, and the public use the NHPI as a timely indicator of past and current housing market conditions. The NHPI is also used as an input in the compilation of other economic statistics. For instance, it is used for estimating certain shelter components of the Consumer Price Index. Moreover, the Canadian System of National Accounts uses the NHPI in estimating the constant price value of new residential construction. Due to the level of geographic detail provided and the sensitivity to changes in supply and demand, the NHPI series are of particular interest to the real estate industry for providing a proxy estimate of changes in the value of resale houses sold. The information provided by the NHPI is also of interest to building contractors, market analysts interested in housing policy, suppliers and manufacturers of building products, insurance companies, federal government agencies such as Canada Mortgage and Housing Corporation (CMHC), and provincial and municipal organizations that are responsible for housing and social policy.

The prices collected are asking prices by the builders and exclude the Goods and Services Tax and other tax related rebates. Missing prices as a result for example of the absence of a sale by a builder in a particular month, are imputed using the best estimate the builder can provide as if a house was to be sold. Not all types of housing are included in the NHPI. Condominiums are excluded from the sample, while single-family detached units as well as row and detached houses are included. Given that builders do not report the price of building lots uniformly, the land price indexes may be less accurate and precise than the overall NHPI. The same caveat applies to the derived residual values that are used for constructing the price indexes for the structure only. Large builders as well as smaller independent builders are represented in the sample used for the NHPI.

A problem with the Canadian NHPI is that it is not representative for resale houses in Canada (or for new houses built in the central areas of the cities surveyed). The houses surveyed for the index are generally in new tracts in suburbs of the survey cities where the price of land is significantly lower than in the city core areas. The movements over time in land prices in suburbs are generally different than the movements in the well established areas of Canadian cities. While the construction price index part of the NHPI is likely to be accurate, the land component probably understates residential land price inflation for the existing housing stock by a significant amount in recent years.[11]

Teranet–National Bank Composite House Price Index ™[12]

The Teranet-National Bank House Price Index™ (TNBHPI) is an independent estimate of the rate of change of home prices in six metropolitan areas, namely Ottawa, Toronto, Calgary, Vancouver, Montreal and Halifax. The price indexes for the six metropolitan areas are then aggregated into a composite national index. The indexes are estimated on a monthly basis using transaction prices for condominiums, row/town houses, and single-family detached homes within the six metropolitan areas.

The TNBHPI uses the repeat sales methodology. Estimating the indexes is therefore based on the premise that houses that are traded more than once in the sample periods are of a constant quality. The TNBHPI attempts to adjust for quality changes of the individual housing units by minimizing or eliminating the influence of any changes in the physical characteristics (e.g.,

[11] See Figure 11.1 for a comparison of the NHPI with other indexes for Canada. This figure provides support for the likely downward bias of the land component of the NHPI.

[12] ©Teranet and National Bank of Canada, all rights reserved.

renovations, additions, etc.). Insofar as (net) depreciation of the properties that are resold is neglected, the index is likely to exhibit a small downward bias.[13] Properties that are affected by endogenous factors are excluded from the calculation of the repeat sales index. These factors may include: non-arms-length sale; change of type of property (for example after renovations); data error, and high turnover frequency (biannual or higher).

The MLS® Average Resale House Price Indicator

The Canadian Real Estate Association (CREA) tracks, on a monthly basis, the number and prices of properties sold via the Multiple Listing Service® (MLS®) systems of real estate boards in Canada. The statistics are available by paid subscription to those who want to use them. Although the coverage of the indicator is limited to only houses that are sold through the MLS®, the system is quite active with about 70% of all marketed residential properties using it. The data are available for over 25 urban markets defined by CREA, as well for the provinces and two territories; a national aggregate is also published.

The indicators are simple arithmetic averages of all sales prices in the market of interest, regardless of housing type. In addition, no consideration is given to the issue of compositional shifts in the sample over time or for disparities in quality in the sample of units. So a change in the price indicator could reflect many factors other than the true price development. These factors range from quality differences that exist in the sample from period to period to the influence of outliers with extremely high or low prices due to special circumstances. In their monthly reports, CREA staff have recently published a weighted version of the national index (available back to 2006 only), with weights corresponding to the share of owned dwelling units by major markets derived from the 2006 Census. However, the price for each major market is still calculated as a simple average, and no attempt is made to track the potentially different trends among various housing types. The one major advantage of the MLS® price indexes over other indicators is their timeliness, since data are typically released two weeks after the reference month.

Bank of Canada - Royal LePage Survey of Canadian House Prices

Prices in the Royal LePage survey reflect the opinions of Royal LePage with regards to the "fair market value" for seven types of properties in a large number of geographical areas. The information obtained is based on local data and market knowledge provided by Royal LePage

[13] This downward bias does not seem to show up in Figure 11.1, since the TNBHPI is more or less in between its two competitor indexes that cover the resale market, but the latter indexes also do not make adjustments for net depreciation. Some housing economists argue that the repeat sales method may have an *upward* bias due to a sample selectivity problem; it may be that dwelling units that are sold more frequently than the average unit are being more intensively renovated and upgraded and hence the quality of a repeat sales unit has actually increased between the two sale dates (rather than decreased due to depreciation).

brokers. The geographical coverage is broad, just like the MLS data, and the classification of housing is more refined. For example, the survey includes prices on four types of singles or detached houses (detached bungalow, executive detached two-storey, standard two-storey, senior executive), two types of condominium apartment units (standard and luxury), and a townhouse. Royal LePage standardizes each type in terms of the square footage, the number of bedrooms, the number of bathrooms, the type of garage, lot characteristics, the status of the basement, and other criteria. In addition, the properties in the survey are considered to lie within average commuting distance to the city centre and are typical of other housing in the neighbourhood. As long as the broker filling in the survey sticks to these guidelines, this is one way of ensuring some degree of constant quality. A comparative disadvantage of the Royal LePage price data is its long publication lag.

This survey is a basis for one of the house price indicators used by the Bank of Canada for monitoring developments in housing markets in Canada[14]. Despite the wealth of price information on many other types of houses in the Royal LePage survey, the indicator developed at the Bank relates only to a subset of singles that were regarded as representative of the market when it was created in 1988[15]. For Canada and 11 local markets, the Bank's price indicator is calculated as a weighted sum of the price of detached bungalow (weight of 0.75) and the price of executive detached two-storey (weight of 0.25). The price of each type of housing is in turn a weighted sum of sub-regions, with weights set to be the sub-regional share of units sold as of a fixed date in the late 1980s. The "units" data were obtained from MLS®.

A Comparative Analysis

A comparative analysis of the four types of property price indexes available in Canada is given in Figure 11.1. The period of analysis covers February 1999 to March 2010. All four series show an upward trend in residential property prices over this period. However, the growth rates differ among the four series. The NHPI recorded the smallest increase at 55% over the entire period. By contrast, the MLS® showed an increase of 122%, more than double that of the NHPI. The Teranet-National Bank House Price Index™ and the Bank of Canada-Royal Lepage indicator increased by 100% and 92%, respectively.

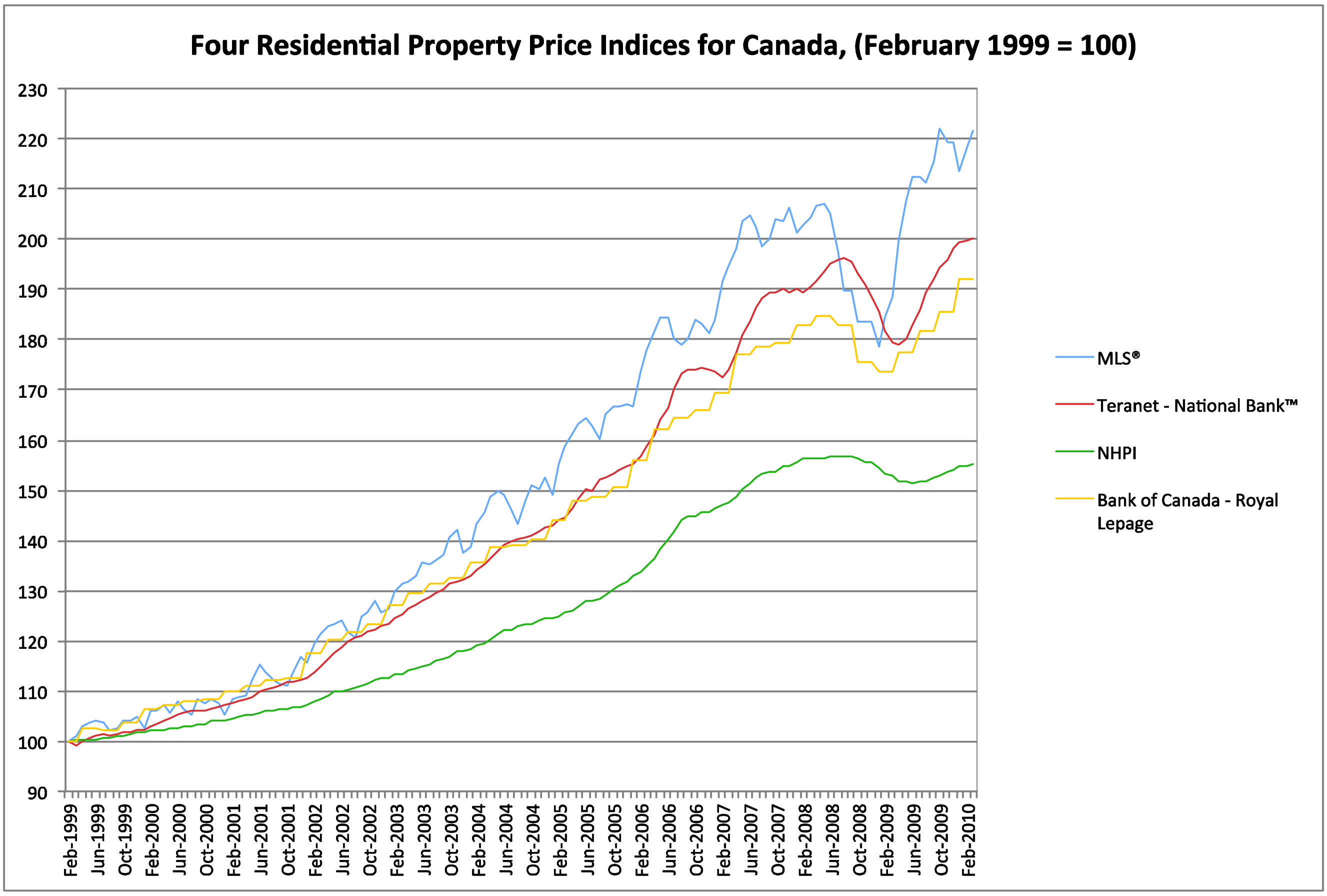
The higher growth rate of the MLS® price indicator may be partly explained by the average price methodology which is used for its calculation and which does not control for period-to-period compositional shifts. The NHPI's slower rate of increase is probably explained by the fact that the index, although it controls for house type over time, does not control for location.

[14] http://www.bankofcanada.ca/en/rates/indinf.html

[15] The Bank of Canada indicator is limited to detached bungalows and executive detached two storey houses.

New houses are constructed farther and farther away from the city centre where markets behave differently compared to properties sold in or near the city core.

Figure 11.1



All four indexes show the drop in house prices that occurred during the economic downturn which began late in 2008. But the MLS® index starts falling slightly sooner than the three others and its drop is deeper. Compared to the other three indexes, the fall in the NHPI starts slightly later and is not as acute. All four indexes start to show an upswing early in 2009 but the MLS® index starts to turn earlier while the turning point from the NHPI index occurs last.

In terms of volatility, the MLS® is the more volatile around its trend due to the compositional shifts in the sample of houses sold each month. The other three indexes, which to some extent adjust for quality changes, show less erratic behaviour over time.

11.3.2 Case study: Germany

Quarterly residential property price index series for Germany are available from 2000. Prior to that date the situation in Germany could be characterised as an uncoordinated set of different indicators provided by several private institutes. “These indicators mostly lacked a clear methodological foundation and had a restricted coverage. Moreover they gave – to some extent – contradictory signals.”[16]

The Federal Statistical Office of Germany (Destatis) took action to improve the situation building on available data sources. Germany had well-established construction price statistics and statistics on purchasing values of building land. In addition, at the local level, the nationwide institution of Expert Committees for Property Valuation, regulated by federal law, provided access to comprehensive databases which contained transaction prices of building land and dwellings and the corresponding property characteristics. The main barrier to the exploitation centrally of the available data had been the differences in the collection systems across the federal states and among the individual local committees.

The methods followed by Germany provide an interesting example of data integration i.e. the drawing on multiple data sources.

Residential Property Price Indexes

Different data sources and compilation methods are used to construct price indexes for different market segments. These are then combined to compute a residential property price index covering all types of properties and sub-indexes relating to existing and new dwellings respectively. The weights used in the compilation of a price index for existing dwellings are the transaction expenditures in the base-year broken down into houses and flats and by the federal state. For turn-key dwellings, the weights are derived from official building activity statistics and for self-builds construction weights are used. Indexes are published within 90 days of the end of the reporting.

Newly built turnkey-ready dwellings and existing dwellings

Data is taken from the information gathered by the local Expert Committees for Property Valuation. This data, that is collected at the time a contract is concluded, covers all sales (cash and mortgage) and consists of actual transaction price (both cash and mortgage) and a number of price-determining property characteristics – type of dwelling (single-family house, two-family house, freehold flat); type of house (free-standing, terraced, semi-detached); type of construction (conventionally built, prefabricated); year of construction; size of plot of land; size of living area; furnishing/luxury elements (kitchen, sauna/swimming-pool, attic storey); car parking facilities; characteristics of location (state, district, municipality; general rating of location: simple/medium/good); number of rooms/floors. In addition, a land valuation is provided.

[16] See Hoffmann and Lorenz (2006).

A combination of hedonic techniques and stratification (one stratum for single-family/two-family houses and one for flats in apartment blocks) is used to adjust for the effects of quality changes in the type of properties being sold. The hedonic regression method that has been adopted is the "double imputation" approach, which was described in Chapter 6, where prices are estimated both for the base period and for the comparison period. Outliers are excluded.

"Self-built" properties[17]

The compilation of a price index for "self-built" properties draws on information from official country-wide construction price statistics. Construction price indexes are available for various types of structure (e.g., residential/non-residential buildings, roads, road bridges) as well as for maintenance work. Prices are collected quarterly for about 190 construction operations (materials included). In total, about 30,000 prices are reported by about 5,000 enterprises at every collection date. The prices refer to the transaction prices relating to contracts concluded in the quarter, excluding value added tax (VAT), i.e., profits and changes in productivity are taken into account. For self-builds, the construction price index for "conventionally built single-family residential buildings" is used. The matched model approach is followed for the construction of the index.

Prefabricated dwellings

The price index uses official producer price statistics for industrial products, in particular the price index for prefabricated single-family houses without a basement with a specific set of characteristics.

Again a matched model approach is adopted for the computation of the index. It is a specific feature of prefabricated houses that the contracts usually provide for the purchase/sale and the construction of complete houses (type of house, e.g., single-family house without cellar), the characteristics of which do not change significantly over the short-term.

Building land

The price indexes for prefabricated dwellings and "self-builds" exclude the cost of the land. A separate price index for building land is compiled from official data on the transaction prices of building land, recorded at the time a contract is concluded. Each data set incorporates the following characteristics: location; characteristics of the municipality; sale date; size of plot; the details of the outline planning permission e.g. whether for a house or for flats and building size.

[17] Self builders include both future owners who do a major part of the building themselves as future owners who involve a building firm that is responsible for the main part of the building work (where the household finalizes the work).

The aggregate price index for developed building land is a weighted average, using the total sales value, of unit value indexes for sub-aggregates. These sub-aggregates are formed on the basis of regional differentiation, mainly a differentiation by districts, building area types and municipality size classes within federal states. The federal states are weighted by combining data on the total of prices paid for developed building land in residential building areas and in rural areas, turnover achieved through building activity and the number of building permits for residential buildings with one or two dwellings.

11.3.3 Case study: Japan

Information on Property Prices

In Japan, official property price indexes only relate to land prices. Information provided by the public sector includes the Public Notice of Land Prices (PNLP) conducted by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT), the Land Price Survey of each prefecture, the Land Value for Inheritance Tax of the National Tax Agency, and the Land Value for Fixed Asset Tax of each municipal government. All of these sources of information represent appraisal values estimated by licensed real estate appraisers.

Information on residential property price indexes (including structures) is collected by the private sector. The most representative property data set is called REINS, which stands for the Real Estate Information Network System. REINS is a data network that was developed using the multi-listing service (MLS) of the US and Canada as a model; the information is obtained via real estate brokers. The REINS data set contains both the asking price when the property is put on the market and the final transaction price at the time of the sale contract. A second, and quite unique, housing price data source is accumulated by housing advertisement vendors.

Both data sources have been used by the private sector to compute and publish housing price indexes. However, all of these indexes have shortcomings and do not fully meet the needs of users. MLIT has therefore begun a work programme which should lead to the construction of an improved index. This will be the first residential property price index to be published by the public sector.

An overview of the property price indexes in Japan is provided in Table 11.1. This includes indexes based on land appraisal values as well as indexes relating to property sales. It is the latter that generates the material for residential property price indexes.

Asking Prices and Selling Prices

In Japan, the seller of a house usually sells it through a real estate broker. Individuals that contract with a broker have to sign one of two forms of a sales agent contract: the exclusive agency contract or the sole agency contract. The other option is to select a general agency contract. These contracts are regulated under Article 34-2 of the Building Lots and Buildings Transaction Business Law.

Table 11.1. Indexes of Property Prices – Published in Japan

Index	Sample	Method	Seasonally adjusted? And (frequency)	Weighing method	Stage of process
Land Price Cumulative Change Rate Index(MLIT)	Appraisal prices in Public Notice of Land Prices by MLIT	Preceding term index × Avg. Volatility	No (Annual)	No	Appraisal value in January 1st every year *(published in the end of March)*
Major City Land Transaction Price Basic Statistic (MLIT)	Sales prices	Average value of unit price per square metre, median value, standard deviation, quartile, etc.	No (Quarterly or annually)	No	Survey after sale registration *(sales price)*
Urban Land Price Index (Japan Real Estate Institute)	Appraisal prices in Public Notice of Land Prices by Japan Real Estate Institute	Preceding term index × Avg. change rate	No (Semi-annual)	No	Appraisal value in the end of March and September every year
Recruit Residential Price Index (Recruit Housing Institute)	Final asking prices in Magazine or Online prices in Magazine or Online	Overlapping Periods Hedonic Regression	Yes (Monthly)	Volume	Offer made *(final asking price)*
Residential Market Index (Japan Real Estate Institute, At Home Co., Ltd., Ken Corporation)	Asking prices or sales prices	Unit price per square metre (building age adjusted by hedonic regression)	No (Semi-annual)	No	Offer made? *(asking price or sales price)*
Tokyo Area Condominium Market Price Index (Japan Research Institute, Limited Real Estate Information Network for East Japan)	Sales prices registered at the Real Estate Information Network for East Japan	Hedonic regression	No (Monthly)	No	Completion of sales *(sales price)*
Newly-Built Condominium Price Change Index (Tokyo Kantei Co., Ltd.)	Asking prices	Moving average	No (Quarterly)	No	*(asking price)*

In the case of the exclusive agency contract, the seller can receive a report at least once a week from the real estate broker, but the seller loses the right to ask another broker to find a buyer and to look for a buyer himself. In the case of the sole agency contract, another broker cannot be asked to find a buyer, but the seller can look for a buyer on his own and the report from the broker will be at least bi-weekly. In the case of a general agency contract, the seller can look for a buyer on their own and ask multiple brokers to find a buyer. On the other hand, the seller does not receive reports from brokers.

In the case of the exclusive agency contract, the contracted broker must register the listing in REINS within five days of concluding the listing agreement and is required to widely look for buyers. In the case of the sole agency contract, the broker must register the listing in REINS within seven days and do the same. For registration in REINS, brokers are not only required to record the asking price at the moment of registration but also the final transaction price. Thus for some transactions made via brokers, both the asking price and the final transaction price are registered.

Public Data Gathering System of Transaction Prices

MLIT has compiled and published information on property transaction prices since 2005. Property transactions are registered by the Legal Affairs Bureau which then sends “Change in Register Information” to MLIT. Based on this information, MLIT sends a questionnaire to the buyer on vacant lots, land with buildings, buildings with compartmentalised ownership (such as office, retail, and apartments) asking for the transaction price. Next, information is added by real estate appraisers or their counterparts. This information includes building use, lot conditions (land form, etc.), road conditions (width of fronting road, etc.), distance to the nearest railway station and other information related to convenience, and legal regulations such as city planning. The resulting “Transaction Case Data” collected in this way is then made anonymous so that the actual property cannot be identified, and is then published as transaction price information on MLIT’s website.[18] Since neither the supply of information on transaction prices nor the supply of the information requested from real estate appraisers is mandatory, non-response and timeliness are issues. The information supplied – including the transaction price – cannot be independently verified.

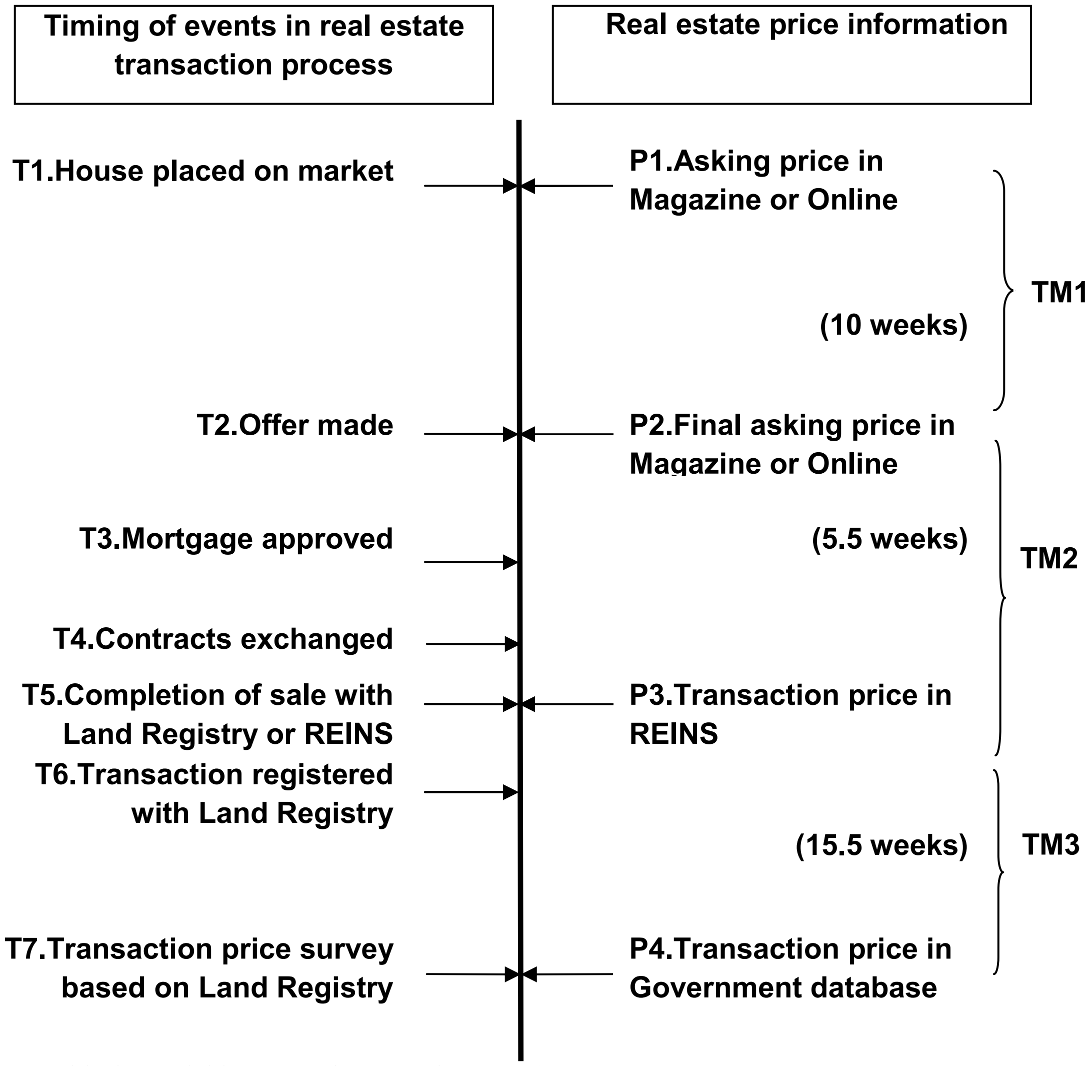
Time Line for Buying and Selling a House and Price Accuracy

The choice of data source is of importance when calculating a housing price index. There are various issues involved. One issue relates to the duration of the various stages in the process of buying a property, i.e. the moments at which price data is collected, the change in “price” (from the initial asking price to the final transaction price), and also how timely the price data is released. Figure 11.2 shows the real estate price information which is currently available in

[18] See www.land.mlit.go.jp/webland.

Japan on a time axis. On the right, four important stages are distinguished with prices P1 to P4. The corresponding time periods between those moments are: the "term" TM1 between the start of the selling process and the moment a buyer is found; the term TM2 from when a buyer is found until the sale contract is finalized; and the term TM3 between the final sale contract and the registration of the selling price in the government's database.

Figure 11.2 Property Information Flow



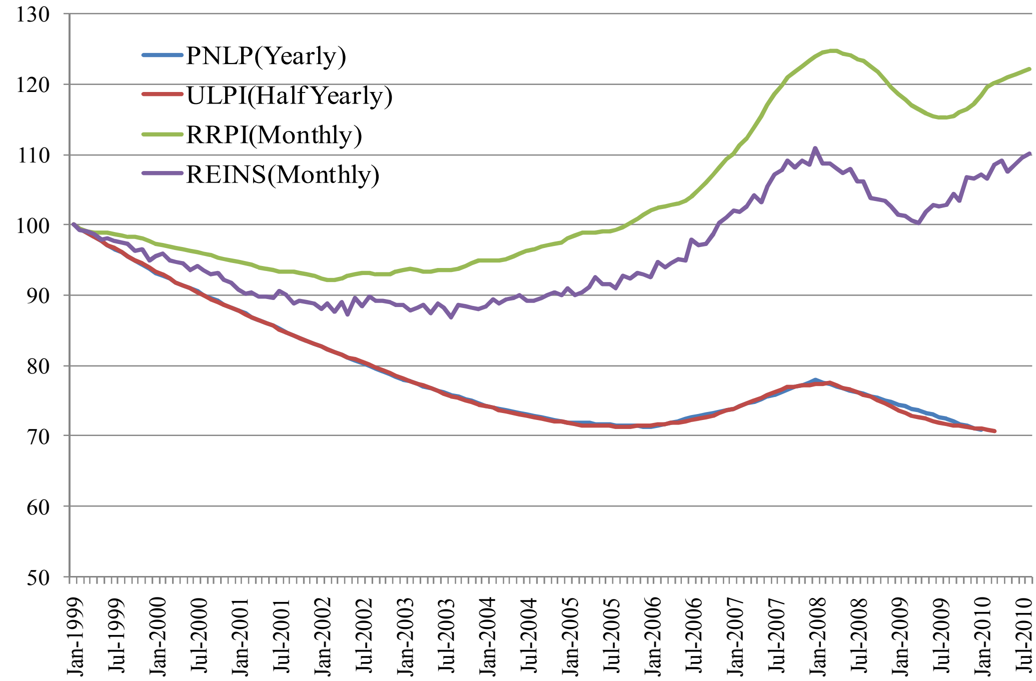
Source. Shimizu, Nishimura and Watanabe (2011)

The average duration of TM1 is 70 days. That is, on average a buyer is found 70 days after the seller enters into the selling process; the maximum duration was 3.72 years. The ratio of P2 to P1 is 0.976 on average, meaning that the price drops by 2.4% from the initial asking price to the last asking price. On average, TM2 is 39 days. The ratio of P3 to P2 is 0.956 on average, i.e. on average the transaction price is 4.4% lower than the final asking price. TM3 is on average 109 days. This means that (for surveyed transaction price data) there is a time lag of approximately 3 months until the selling price is registered in the government's database. The price differentials at different points in the selling process can, of course, vary over time depending on the state of the owner-occupier housing market.

Comparative Analysis of House Price Indices in Tokyo Metropolitan Area

Figure 11.3 compares four property price indices. The REINS data are used by the Real Estate Information Network for East Japan and the Japan Research Institute who jointly produce the Tokyo Used Condominium Price Index. This monthly index has been published since 1995 and is constructed using a hedonic regression method. The Recruit Residential Price Index (RRPI) is also a hedonic price index[19], based on the final offer price of properties in Recruit's magazine, and relates to re-sold single family homes and condominiums. This index is also monthly and has been published since January 1986[20], although only widely available in its current form since the beginning of 2000. Two land price indexes, thus excluding buildings, are shown in Figure 11.3, the bi-annually ULPI and the yearly PNLP. These are appraisal-based indexes.[21]

Figure 11.3. Four Residential Price Indexes for Japan (January 1999=100)



[19] The Recruit Residential Price Index uses the time dummy method and, in consequence, is subject to revision (see Chapter 6).

[20] See Shimizu, Takatsuji, Ono and Nishimura (2010) for details.

[21] Shimizu and Nishimura (2006) (2007) compare appraisal values and selling prices and point to the problems of valuation errors and smoothing in the appraisal-based indexes.

The property price indexes that include the structures clearly show a different trend than the land price indexes. Also, the former began to recover some years after the financial crisis in 2008 whereas the latter continued to decrease. Notice that the REINS index is much lower than the RRPI, in spite of the fact that both are hedonic indexes.

11.3.4 Case Study: United Kingdom

The UK probably has more house price indexes published on a regular basis than any other country. The range of residential property price indexes that are published in the UK mainly stems from the interrogation and exploitation by different organisations of the different data sets which are generated at different points in the process of buying and selling a house. The latter often takes place over a period of several months or more and the particular stage in this process at which the price is abstracted and entered into an index can impact on the measured rate of house price inflation. In the UK the exploitation of data on property prices occurs at the following stages:

- As soon as the property is on the market. *Asking price*. Data source: estate agents.[22] Publisher: estate agents, Financial Times and property websites.
- Mortgage approved. *Valuation by mortgage lender*. Data source: mortgage lenders. Publishers: various mortgage lenders.
- Mortgage completed. *Mortgage completion price*. Data source: mortgage lenders. Publishers: The Department for Communities and Local Government (DCLG)
- Transaction registered. *Transaction price*. Data source: Land Registry.

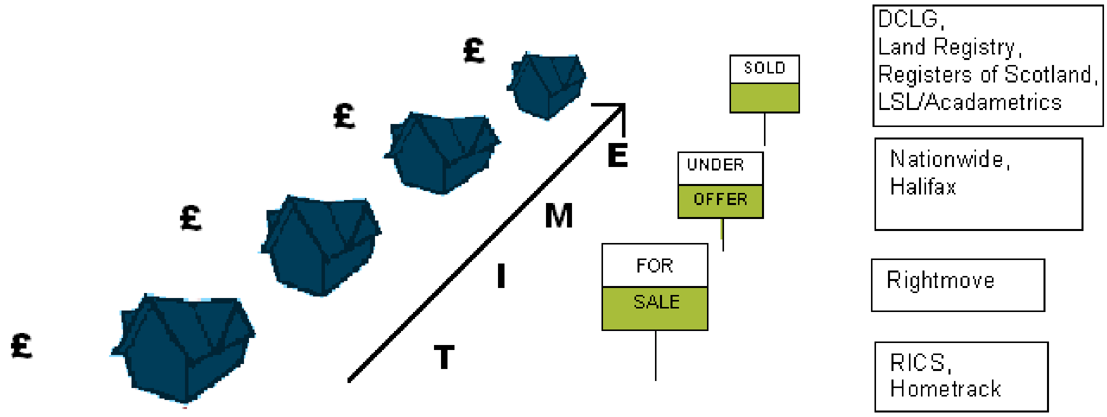
The time-line for buying and selling a house in the UK, including the different points at which information is collected and used to produce a house price index, is given in Figure 11.4.

The UK currently has two official house price indexes. One is published monthly by the Department of Communities and Local Government (DCLG) and is based on information provided by mortgage lenders, through the Council of Mortgage Lenders, on valuation price at the point when the sale is completed. It is published about six weeks after the reference date for the house sale – or, on average, about four-five months after a house is first put up for sale. It only covers purchases involving a mortgage. The other is published monthly by Land Registry based on sales of properties registered with them. It is published a month after the reference date; i.e., one month after the registration of the sale but suffers from a lack of

[22] Although not related to the issue of timing, a disadvantage of advertised prices and mortgage approvals is that not all of the prices included end in transactions, and in the former case, the price will tend to be higher than the final negotiated transaction price.

timeliness due to delays from homebuyers or their agents notifying the Land Registry of transfers of ownership.

Figure 11.4. House Purchase Time-line



Two mortgage lenders, Halifax and Nationwide, publish indexes based on their valuations of a property at the time that they grant a mortgage. These indexes are produced within a few weeks of the reference data for granting the mortgage and about three to four months after a property is put up for sale. They are a little more timely than the official DCLG index but have a much more restrictive coverage with no guarantee that the properties that they have granted mortgages on are representative of either all property transactions or all purchases involving a mortgage.

Another index is compiled by an organisation named Hometrack, a business service company which provides a range of market intelligence on the housing market to organisations across the residential sector including Developers, Housing Associations, Corporate Investors, Estate Agents, and Local and Central Government. Hometrack conduct a monthly survey of estate agents who are asked to gives their view on the achievable selling price for each of four standard property types. It is the most timely of all the published indexes, being published about three to four weeks after the reference period with in effect no other time-lags involved, but it is an opinion survey of the likely selling price of properties on the market.

A research based consultancy firm, Acadametrics, also publishes a house price index based on data provided by the Land Registry. The LSL/Acadametrics index is published a few weeks after the end of the reference period based on an "index of indexes" forecast method. The index for each time period is subsequently revised until all transactions have been included.

An index based on asking prices advertised on the Rightmove property website is also widely used in the UK.

Table 11.2 summarises the scope and definition plus the main aspects of compilation method for the seven indexes available in the UK shown in the time-line in Figure 11.4.

Table 11.2. Indexes of Residential Property Prices – Published in the UK

Index	Sample	Method	Seasonally adjusted?	Weighing method	Stage of process
DCLG[23]	Sample of Mortgage Lenders	Mix-adjustment and hedonic regression	Yes	Expenditure	Mortgage completion *(transaction price on mortgage document)*
Land Registry (monthly)	Sales Registered in England and Wales with a previous sale since 1995.	Repeat Sales Regression	Yes	Expenditure	Sale registration *(transaction price)*
Halifax	Halifax loans approved for house purchase	Hedonic regression (quality adjustment)	Yes	Volume	Mortgage approval *(valuation price)*
Nationwide	Nationwide loans approved for house purchase	Hedonic regression (quality adjustment)	Yes	Volume	Mortgage Approval *(valuation price)*
Hometrack	Survey of estate agents (valuations)	Mix-adjustment	No?	Expenditure	Achievable selling price
Rightmove	Asking prices posted on website	Mix-adjustment	No	Expenditure	*(asking price)*
LSL/ Acadametrics	Sales Registered in England and Wales	Forecasting model, includes mix adjustment.	Yes	Volume	Sale registration *(transaction price)*

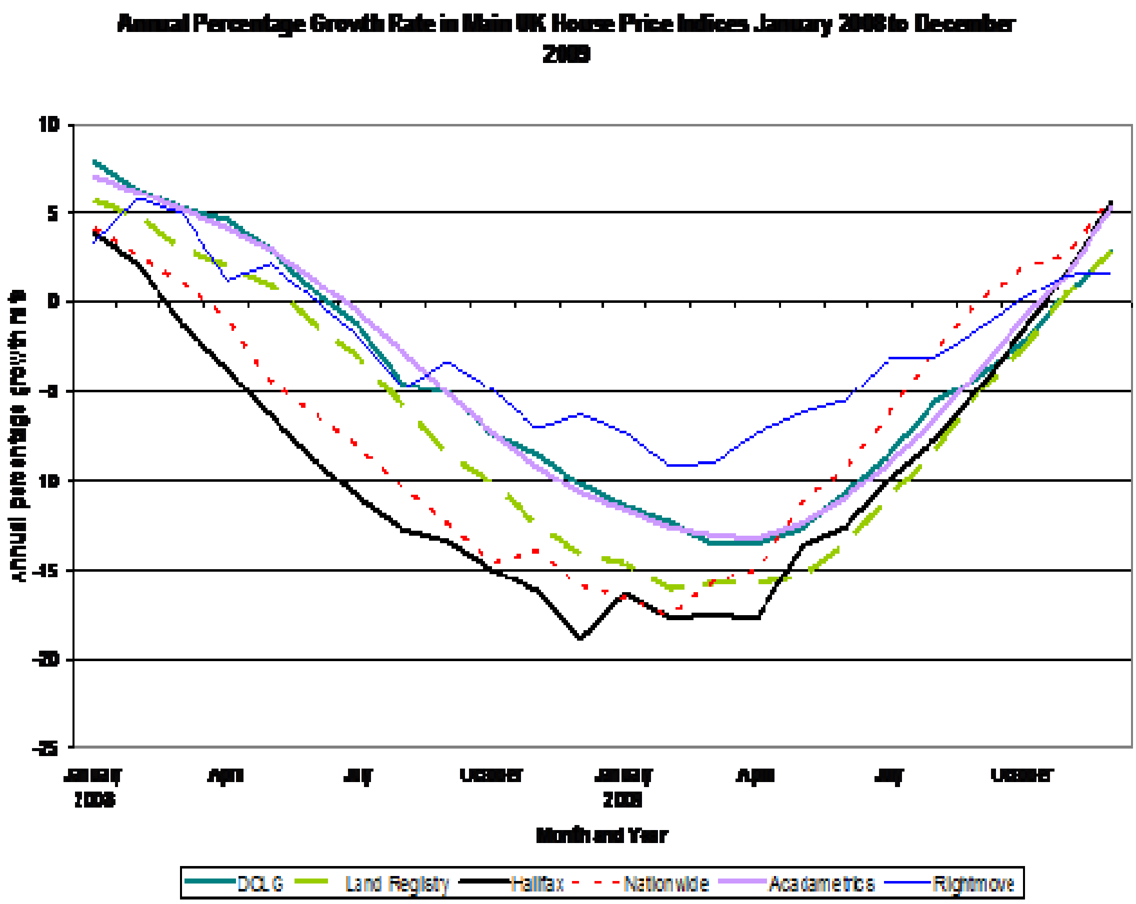
Figure 11.5 plots six of the seven indexes lists (Hometrack is excluded as it relates to a survey relating to achievable selling price). The Halifax, Nationwide and Rightmove indexes show the greatest volatility, which makes it more difficult for the early detection of turning points. The DCLG, Land Registry and Acadametrics indexes are smoother and show a close

[23] Department of Communities and Local Government.

correlation in terms of detecting a turning point although the Acadametrics index consistently shows larger decreases in prices.

A review by the UK National Statistician into house price indexes can be found on the national statisticians web pages.[24]

Figure 11.5



11.3.5 Case Study: India

Movement in prices of real estate, particularly residential housing, is of vital importance to the macro economy of India as well as to individual households. It is not surprising that there is a user demand for a relevant and reliable index for tracking house price movements. But a lack of transparency in the residential property market transactions and limited availability of price information pose important challenges for keeping track of real estate price dynamics.

[24] http://www.statisticsauthority.gov.uk/national-statistician/ns-guidance-and-reports/national-statistician-s-reports/index.html

Registration of the property price is a legal necessity for any property transaction in India. So in principle, the official authority of property registration has the details of all transactions during a reference period. In theory the data are available on a daily basis with a month lag from first reporting a change of ownership. However, it is well known that the registered prices of houses are grossly underestimated due to very high registration fees and stamp duty. The subsequent obligations for the payment of property tax acts as a further disincentive to individual purchasers (except corporate bodies) for revealing the exact sale price of a house. Further, the registration procedure and records maintenance are not computerized and the records are maintained in regional languages which necessitates further work with respect to bringing them into common format.

For these reasons, the administrative data relating to the registration of changes of ownership are not exploited and an alternative source of data has had to be found. This alternative data source relates to market data based on transaction prices collected by the National Council of Applied Economic Research (NCAER), a national level research organisation, from Resident Welfare Associations (RWAs), real estate agents and brokers. The valuation data of housing loans financed by Banks and Housing Finance Companies (HFCs) are collected to supplement the actual transaction price data collected through survey. These data are then used to compile the National Housing Bank's RESIDEX index.

The NHB RESIDEX Index

NHB RESIDEX is a pioneering attempt by the National Housing Bank (NBH), an apex bank for the housing sector owned by the Central Bank of India, to measure residential prices in India. As a pilot, five cities – Bangalore, Bhopal, Delhi, Kolkata and Mumbai – were studied. The process of data collection posed many challenges. Further, there were methodological issues relating to the analysis of data. In the event and after much work, the NHB launched its first RESIDEX for tracking prices of residential properties in India, in July 2007.

The index is based on actual transactions using the sale /purchase price plus supplementary data on valuations. Primary data on housing prices is collected from real estate agents by commissioning the services of a consultancy/research organization of national repute, who obtain transaction prices. In addition, data on housing prices are also collected from the housing finance companies and commercial banks. The latter relates to the valuation prices associated with the housing loans contracted by these institutions.

The salient features of NHB's RESIDEX are:

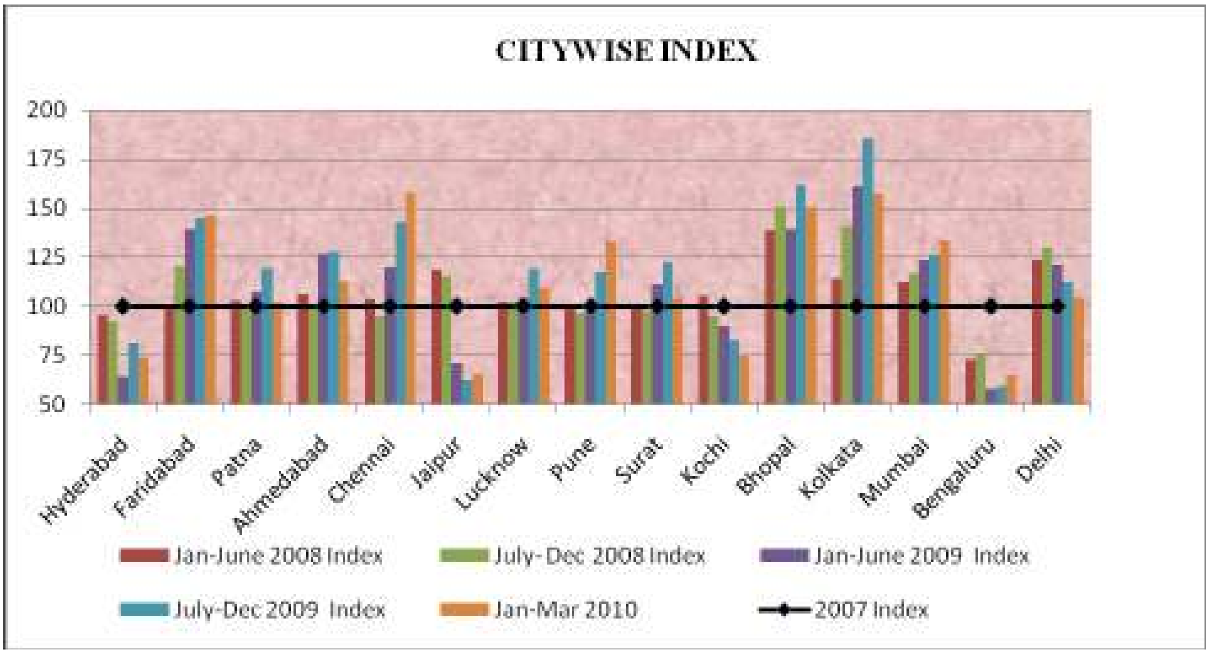
- It covers all types of residential properties in fifteen cities.[25]

[25] In due course, based on experience and depending upon the availability of data, it may be expanded to cover commercial properties, as well.

- With 2007 as base, NHB RESIDEX index is produced on a quarterly basis.[26]
- Alternative series are compiled based on transaction weights and stock weights.
- It covers cash purchases and purchases financed via a loan.
- It covers new and old constructions.
- The index is constructed "using weighted averages of price relatives".[27]
- No quality adjustment is currently made in terms of location, size etc.
- It is revisable to take account of late data.
- Information on the movement in prices of residential properties by location, zone and city, is also available, e.g., separate indices are available for each zone in each of the fifteen towns covered.

For a country the size of India the geographical dimension is important. For example, the city-wise price indexes, shown in Figure 11.6, help home buyers with their purchase decisions by enabling comparisons between localities and help builders and developers in making future investment decisions.

Figure 11.6. NHB RESIDEX Indexes, India



[26] 2001 was taken as the base year for the pilot index based on five cities to be comparable with the base year(s) of Wholesale Price Index and Consumer Price Index. Year on-year-price movements during the period 2001-2005 were captured, and subsequently updated for two more years i.e. up to 2007. The index was then expanded to cover ten more cities viz., Ahmedabad, Faridabad, Chennai, Kochi, Hyderabad, Jaipur, Patna, Lucknow, Pune and Surat, at which point the base year shifted from 2001 to 2007.

[27] It should be noted that this is a weighted Carli index and as such is likely to have an upward bias; see CPI Manual (2004), page 361.

Development of the NHB RESIDEX to increase its relevance to users continues:

- The index will be expanded in a phased manner to cover all 35 cities in India having a million plus population as per the 2001 Census.
- There is a proposal is to expand NHB RESIDEX to 63 cities which are covered under the Jawaharlal Nehru National Urban Renewal Mission, the flagship national mission of the Government of India, to make it a National Index.
- In due course, based on experience and depending upon the availability of data, it may be expanded to cover commercial properties.

11.3.6 Case Study: Columbia

A house price index for existing houses, the IPVU, is compiled by the Banco de la República (Central Bank of Colombia). There are some other indexes that relate to construction costs and the prices of new housing units, which are produced by DANE (the national statistics office of Colombia). No series is produced which amalgamates the information from the two series to produce an index covering sales of all residential property in Colombia.[28] In the past, consideration was given to the exploitation of administrative data but this was found not to be possible due to the complexities involved.

The IPVU

The project to construct a price index for existing houses in Colombia, the IPVU, started in 2003. In the past, the lack of access to basic information had been the principal barrier to the construction of such index. After consulting with several lending banks about the importance of having a measure of the value of existing houses, the project was launched with finance from the Central Bank of Colombia (Banco de la República). The Statistics Section of Banco de la República is in charge of the production and publication of the index.

The IPVU is restricted to the principal metropolitan areas of Colombia, covering the cities of Bogotá, Medellín, Cali and Soacha in Cundinamarca, and Bello, Envigado and Itaguí in Antioquia. The index is calculated using information from loan's appraisals reported by the mortgage lending banks Davivienda, BBVA, Av. Villas, Bancolombia, Colmena BCSC and Colpatria. In consequence, the index covers only properties purchased using a loan – cash purchases are excluded. The banks provide the Banco de la República with the commercial values and addresses of all approved mortgages. The prices which are entered into the index

[28] The integration of the two indices would raise the issues of a lack of consistency and incoherence. For example, the IPVU index is based on independent valuations when a mortgage is applied for and the DANE index is based on asking price.

are taken from independent valuations required by the mortgage lender. The valuation is close to the market price when the disbursement is made.

The index is published on the Bank's webpage, on a quarterly basis with a lag of a quarter and is revisable on a quarterly basis, reflecting the repeat sales methodology used (see below). In addition an index is published based on annual averages. Sub-indexes are produced for the principal metropolitan areas: Bogota; Medellin; and Cali.

Houses are classified according to whether they receive subsidies or not. These relate to the VIS and NOVIS indexes, respectively. The receipt of a subsidy depends on the value and location of the house. The term Low-Income Housing (LIH or VIS in Spanish) refers to residences which are developed to guarantee the right to a house for low-income households. On each development plan, the national government will establish the maximum price and type of residences meant for these households. They will take into account, amongst other aspects, households' access to credit markets, the amount of credit funding available from the financial sector, and available government funds aimed to target housing programs (subsidies).[29]

The methodology applied is similar to the Case-Shiller repeat sales methodology. There is a lack of detailed information on the characteristics of housing needed to address the constant "mix" requirements of the Case-Shiller method through the use of stratification. However, progress is being made with the expectation that the information provided by the mortgage lending banks will in the future include a wide array of data on house specific characteristics. The current lack of detailed characteristics is dealt with by data editing. If the property shows an "abnormal" price change, i.e. if it is deemed to be an outlier, the price information is discarded and does not enter the index. This is in order to prevent re-modelled or neglected houses from entering the index. The index is revisable, reflecting one of the characteristics of the repeat sales methodology.

A comparative Analysis

The detailed sub-indexes which are available provide the opportunity for a more-detailed analysis of the market in existing homes. An indication of the range of outputs available to the user is given by Figures 11.7-11.11.

The "indice nominal" uses the prices reported by the Banks, i.e., it is not deflated; the "indice real" is the IPVU deflated by the CPI average for the year. In the case of quarterly indexes the IPVU is deflated by the CPI quarterly average.

[29] For more information on this topic, see http://www.cijuf.org.co/codian03/junio/c31847.htm.

Figure 11.7. Quarterly National House Price Index for Existing Units – Nominal and Real

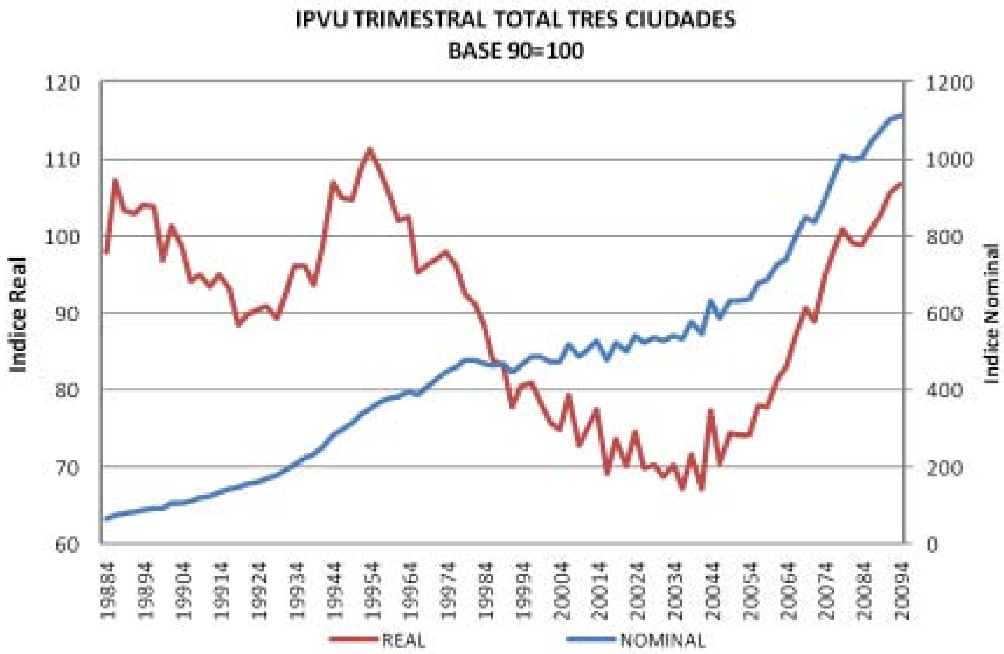


Figure 11.8. Quarterly National Real House Price Index for Existing Units – Annual Percentage Changes (Quarter on Quarter)

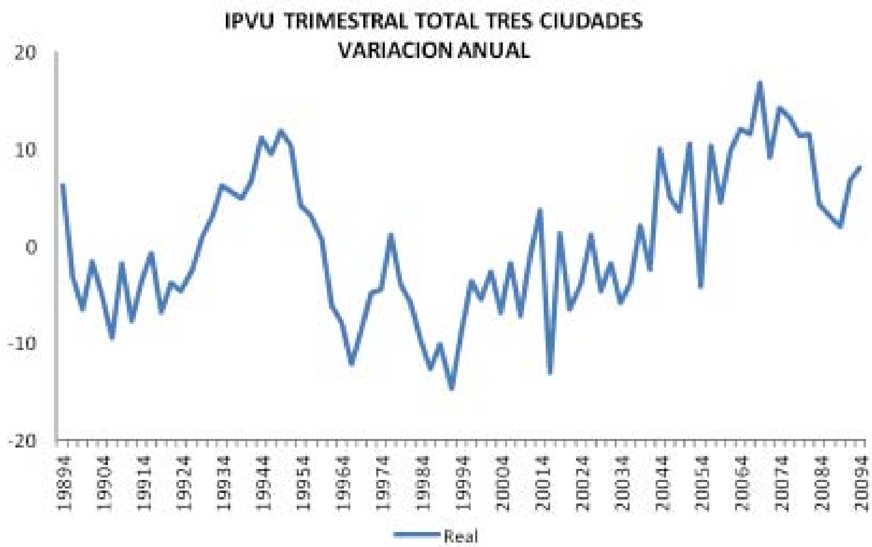


Figure 11.9. Annual National House Price Index for Existing Units – Year-on-Year Percentage Changes[30]

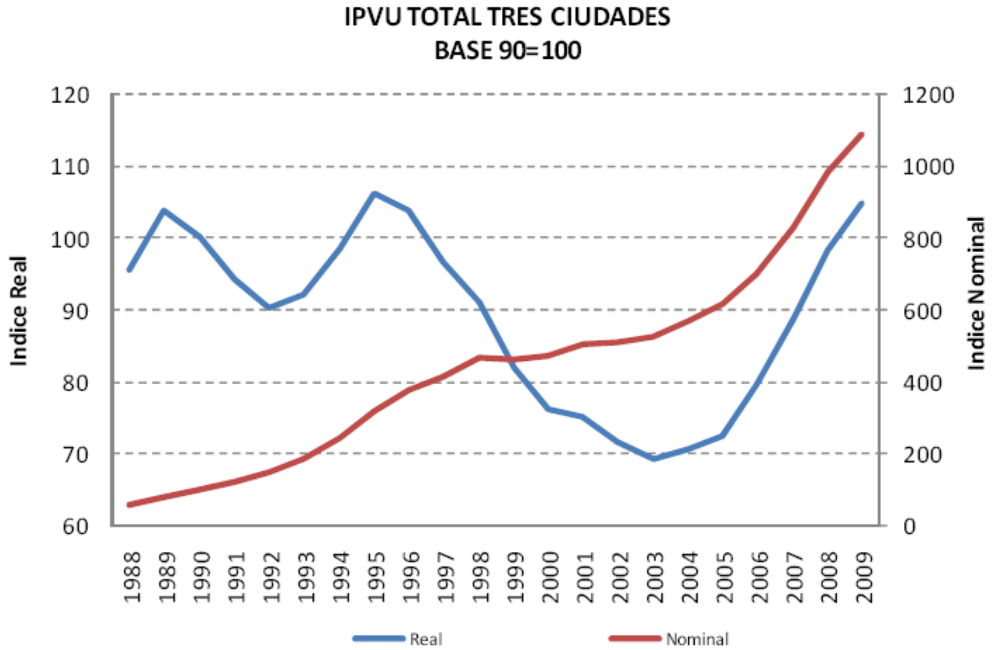
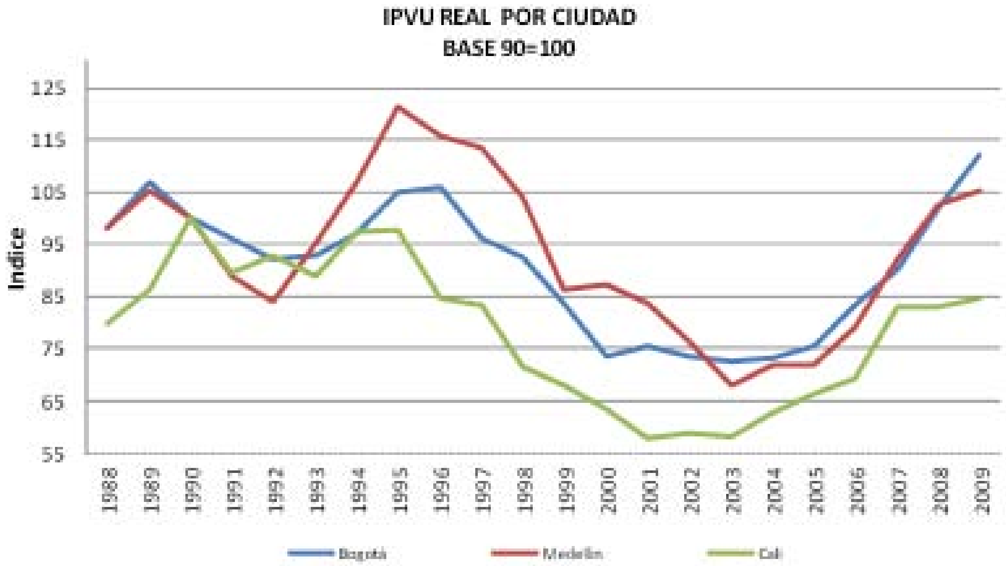


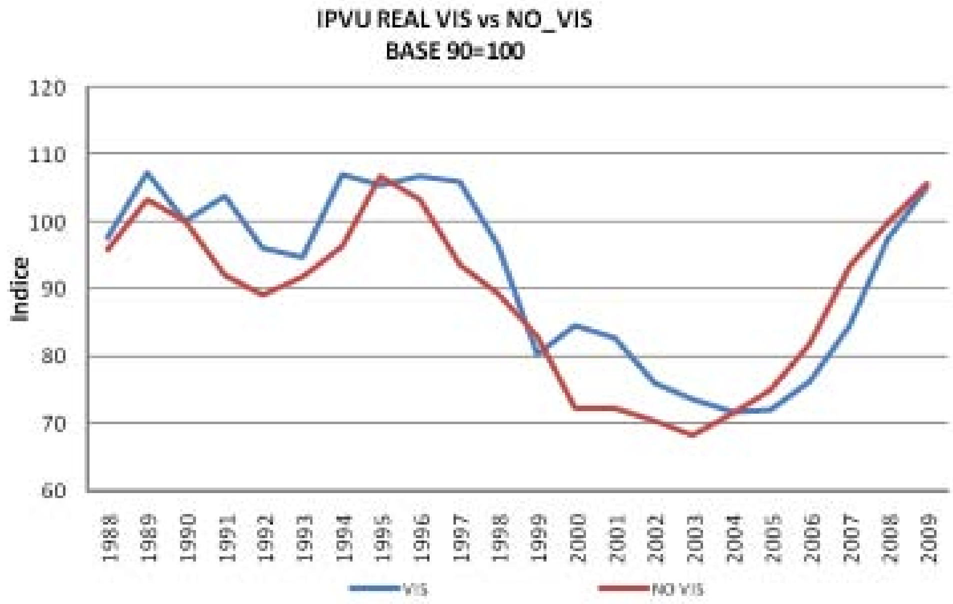
Figure 11.10. Annual Real House Price Index for Existing Units – Year-on-Year Percentage Changes

PRINCIPAL METROPOLITAN AREAS



[30] The annual publication of the IPVU takes the average index level over a period of twelve months and compares it with the average for the previous twelve months.

Figure 11.11. Annual Real House Price Index for Existing units: Houses with Subsidies (VIS) and Houses without (NOVIS) – Year-on-Year Percentage Changes



11.3.7 Case Study: South Africa

The following case study from South Africa provides a good illustration of the obstacles to the compilation of a residential property price index when a significant proportion of the housing stock relates to informal or traditional dwellings.

Introduction to the South African Housing Market

Diverse dwelling types characterise the South African housing stock; it consists of formal, informal, tribal, and other accommodation in backyard or shared property housing. Formal housing includes stand-alone houses (government subsidised and private houses), townhouses and flats, whereas informal housing includes shacks and traditional dwellings includes rondavels and huts. Backyard housing consists of dwellings that are situated in a backyard of a property with a main house, and shared property housing occurs when more than one dwelling is constructed on a single stand. The distribution of the South African housing market is as in Table 11.3.

According to the 2001 Population Census, the number of dwellings in the formal market has increased by 37.1% from 1996 to 2001; informal housing by 26.4% and traditional dwellings by 0.6%. In contrast, backyard or shared property has decreased by 14.5%.

Table 11.3. Tenure Status – All Housing in South Africa (According to Census 2001)

Housing type	Total	Owner-occupiers (%)	Renters (%)
Houses	6,238,454	66.1	45.6
*Subsidised housing**	*1,074,028*	*9.6*	-
Flats	589,109	2.9	16.3
Townhouses	319,868	3.3	4.1
Informal	1,836,230	10.3	18.4
Traditional	1,654,787	15.0	4.1
Backyard or shared property	532,986	2.4	11.5
Total	11,171,434	100.0	100.0

* National Treasury estimate.

In South Africa, builders and/or property developers construct all residential property, with the exception of tribal and informal housing. For the construction of formal housing, a monetary transaction takes place by financing the dwelling with the money of the buyer and/or a mortgage bond. The dwellings and their values are recorded at the local municipality and deeds office. For tribal and informal housing, very few monetary transactions take place. Where they do take place, the transactions will be small cash expenditures but the dwelling will generally not be recorded by a local municipality. However, due to the demand for basic services, government has begun to record the number of dwellings in informal settlements and rural areas, but the value of the dwelling is not recorded. The situation represents an exceptional challenge for compilers of residential property price indexes.

Residential Property Price Indexes in South Africa

There are various house indexes published in South Africa, but not by Statistics South Africa. Published house price indexes include the First National Bank (FNB) House Price Index, the ABSA House Price Index and the Standard Bank Median House Price Index.[31]

The FNB house price series is constructed using the average value of housing transactions financed by FNB. To eliminate outliers from the data sample, transaction values included in the sample must be above 70% of FNB Valuations Division's valuation of the property but below 130%, while purchase prices recorded as above R10-million are excluded. In order to reduce the impact on the index of rapid short-term changes in weightings of different property segments, due to relative shifts in transaction volumes, the weightings of the different market segments according to number of rooms are kept constant at their 5-year average weighting. A statistical smoothing function is applied to the data and the data may be revised. The FNB index is calculated monthly.

[31] ABSA, FNB and Standard Bank have the majority of the banking market share in South Africa

ABSA House Price Index (HPI) measures the nominal year on year house price movements of houses purchased through approved mortgage loans from ABSA. The ABSA HPI is based on the total purchase price of houses in the $80m^2$- $400m^2$ size category, priced at R3,1 million or less (including improvements). Prices were smoothed in an attempt to exclude the effect of seasonal factors and outliers in the data. The index is calculated monthly.

Standard Bank's index is based on the median house price of the full spectrum of houses, using a five-month moving average. National data from the Deeds Office are available only with a lag of up to nine months, so data from Standard Bank, which has a market share of about 27.7% and whose data are generally highly correlated with those of the Deeds Office, are considered a good proxy for the national market. The index is constructed on a monthly basis.

In the construction of these house price indexes only formal housing (i.e., houses, townhouses and flats) purchased by means of a loan are included – cash sales and "informal" housing are excluded.

Limitations to the Construction of a Residential Property Price Index

The difficulty in constructing a RPPI in South Africa is mainly due to the lack of acceptable estimates on housing stock and price information on informal and traditional dwellings. These dwellings make up 19.6% of all structures and therefore constitute a significant sector of the market in South Africa.

The sector also has its own distinct features. For example, what defines an informal dwelling?

- Residential areas where a group of housing units has been constructed on land to which the occupants have no legal claim, or which they occupy illegally;
- Unplanned settlements and areas where housing is not in compliance with current planning and building regulations;
- Informal dwellings are typically built out of corrugated steel plates for the walls and roof (shack);
- The households themselves mostly build these dwellings.

What is a traditional dwelling?

- This is a general term, which includes huts, rondavels[32], etc. Such dwellings can be found as single units or in clusters.
- The dwelling can be made of clay, mud, reeds or other locally available materials.

[32] A circular often thatched building with a conical roof.

Primary Concerns in the Construction of a Residential Property Price Index

As stated elsewhere in this handbook, two main problems in the construction of a residential property price index are the sporadic nature of transactions and a lack of matching due to the fact that houses have unique price determining characteristics. In the case of formal housing, these two factors apply, but for informal housing, the second factor is much less important. Informal dwellings have, exceptionally, standard attributes since most of them are made of corrugated steel and have one to four rooms. Similarly their location will tend to be in the same types of areas. In these circumstances the matching principle may not be difficult to apply. In addition, the fact that the owner of the shack does not own the land that the dwelling stands on, implies that a decomposition of the index into land and structures is not relevant. The census 2001 indicated that the distributions of rooms are as in Table 11.4.

For traditional dwellings, the decomposition into land and structures is not relevant either. In this case, the land is allocated to the person or household by the chief of the tribal area, and no cost or only a small fee is levied. However, to estimate the price of the dwelling may prove problematic if, unlike formal dwellings, mainly natural materials are used in the construction.

Table 11.4. Distribution of Number of Rooms in Informal Dwellings

Number of rooms	% of total informal dwellings
1	40.0
2	27.2
3	15.1
4	10.5
5 +	7.2

Weighting of Non-Formal Housing

Weighting of non-formal (informal and traditional) housing will be complex in nature as the owners construct most of the dwellings themselves and monetary transactions are limited. In addition, materials for the construction of an informal dwelling are mostly second-hand and for traditional dwellings, natural materials are used; cost estimates for these types of materials are difficult to obtain and, indeed, they may have been gathered rather than purchased.

Although most of the characteristics of the dwellings are known from the population census, the value of an informal or traditional dwelling is difficult to estimate because there are no organised markets and the values are not registered at a deeds or land registration office. Also, the movement of informal dwellings from one settlement to another may pose a problem in the estimation of the housing stock. The rate of new constructions and demolitions would be unknown, since it is uncertain whether all dwellings that were broken down were erected once more in the new area.

Pricing of Non-Formal Housing

Non-formal house prices do not depend on normal market price determinants. The plot area, location, age and renovations typically do not affect the price. The only aspects that influence the cost of the dwelling are the materials used and this is of course influenced by the size of the structure; see Table 11.5.

Table 11.5. Price Determinants

Price determinants	Traditional dwellings	Informal dwellings	Formal dwellings
Area of structure	No	No	Yes
Area of land	No	No	Yes
Location	No	No	Yes
Age	No	No	Yes
Renovations	No	No	Yes
Type of structure	No	No	Yes
Materials	Yes	Yes	Yes
Other price determining characteristics	No	No	Yes

Table 11.6. Percentage of Materials Used in the Construction of Informal and Traditional Dwellings in South Africa

Year	2002	2003	2004	2005	2006	2007	2008	2009
Materials used for roof								
Corrugated iron/zinc	72,1	72,1	71,6	78,2	79,5	78,6	78,6	83,6
Organic materials	23,2	24,2	23,8	16,8	16,2	17,1	15,8	13,3
Asbestos	1,9	1,6	1,4	1,7	1,8	1,2	2,1	0,5
Other	2,6	2,1	3,1	3,2	2,1	3,1	3,1	2,2
Total	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**
Materials used for walls								
Bricks	2,6	3,3	2,3	1,8	1,4	1,6	2,3	1,9
Cement block/concrete	2,9	2,2	2,8	1,9	2,5	2,3	2,4	1,4
Corrugated iron/zinc	35,1	36,1	33,9	40,0	43,6	43,9	41,4	42,2
Wood	9,8	9,4	8,9	9,6	10,5	10,8	10,1	8,6
Mud and cement mix	7,0	5,2	6,3	5,0	5,8	6,5	6,7	10,4
Wattle and daub	1,4	1,1	1,7	1,0	0,5	0,9	1,3	1,2
Mud	38,2	39,8	41,8	37,2	33,7	31,8	32,8	31,8
Other	2,6	2,9	2,3	2,6	1,8	2,2	2,9	2,5
Total	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**

Price collection for traditional and informal dwellings would be very difficult, since the owner constructs the dwelling him/herself in most cases and monetary transaction for the complete dwelling rarely takes place (the purchases of materials are normally in cash). The only way to obtain prices of newly constructed informal and traditional dwelling is to conduct a survey of newly constructed dwellings on a frequent basis, since most of these are not registered at the deeds office, and if registered, the value of the dwelling is not recorded. An alternative for these types of dwelling, yet to be explored, is to compile a "notional cost of construction" index based on the pricing of quantity information of the type that is shown in Table 11.6.[33]

Summary

It would be a very complex task to calculate an comprehensive residential property price index for South Africa, due to the diverse nature of housing in the country. Different methods will be required for the collection of prices for different housing types. In addition, weight estimation for each type of housing will be difficult, as different housing types have different cost determining characteristics. Furthermore, the limited data availability for each housing type exacerbates the problem.

Table 11.7. Evaluation of Barriers

Possible problems	Traditional dwellings	Informal dwellings	Formal dwellings
Organised market	No	No	Yes
Reliable price estimates exist about the cost of housing	No	No	Yes
Movements of dwelling from one settlement to another	No	Yes	No
Registration of property at deeds office	No	No	Yes
Monetary transaction at lending institution	No	No	Yes
Transfer of cash for building of structure	Sometimes	Sometimes	Yes
Dwelling constructed by property developer or builder	No	No	Yes
Price depends on typical price determining factors	No	No	Yes

The primary barriers to the construction of an inclusive residential property price index in South Africa are listed in Table 11.7 and include:

- The absence of an organised market for informal and traditional housing;

[33] See Blades (2009).

- The absence of reliable data estimates on the cost of informal and traditional housing;
- The nomadic life-style. If a survey is conducted, movements of informal settlements from one area to another pose a problem in terms of measuring the price development of this type of housing because prices are normally collected in specific areas;
- There is no registration of property at the Deeds Office;
- Monetary transactions do not always take place to obtain or build the dwelling;
- Prices do not depend on typical price determining factors such as the price of land and labour and material costs.

Chapter 13: Recommendations

This handbook provides detailed and comprehensive information on the compilation of residential property price indexes (RPPIs). It provides an overview of the conceptual and theoretical issues that arise, explains the different user needs for such indexes and gives advice on how to deal with the practical problems that statistical offices are confronted with in the construction of such indexes. Earlier chapters cover all relevant topics including: a description of the different practices currently in use; advice on the alternative methodologies available to the compiler; and the advantages and disadvantages of each alternative. The purpose of this chapter is to draw together all this information and make recommendations on best practice for compiling residential property price indexes, including how to improve international comparability. The recommendations necessarily take into account the different situations countries are confronted with in terms of data availability and therefore cannot be too prescriptive.

Users of RPPIs are also catered for. The handbook provides information not only on the different methods that are and can be deployed in compiling such indexes, but also on the statistical limitations of what is being measured. Users will want to bear the latter in mind so that the results of an index can be interpreted correctly. Any set of recommendations has to start with an understanding of the basic concept underlying the target index, in other words what a residential property price index is trying to measure. This will, of course, depend on user needs and the purpose of the index.

The recommendations in this chapter follow the same order as the earlier Chapters 4 to 9. Chapter 4 describes the main elements of a conceptual framework for residential property price indexes and Chapters 5 to 9 describe the main statistical methods used in constructing such indexes. The different methods essentially relate to alternative solutions to the problem of quality change, that is, how to adjust for both changes in the quality mix of the properties sold and for quality changes (the net effect of renovations, extensions and depreciation) of the individual dwellings.

Increasing international comparability can be achieved by the introduction and adoption of standards on conceptual, methodological and computational issues related to the construction of RPPIs. Against this background the following recommendations are made.

13.1 Conceptual Basis of the Index

In principle, the target index, in other words the type of index to be compiled, will depend on its purpose.

The *System of National Accounts (SNA) 1993* and its recent updating, the *System of National Accounts 2008*,[1] should be used as the conceptual framework for residential property price indexes.

13.2 Weighting

A price index which is required to measure the *wealth* associated with the ownership of residential property should be *stock-weighted*. A *stock-weighted* index is also appropriate for a financial stability indicator, most particularly for an index which is being used to identify property price bubbles.

A price index which is required for measuring the *real output* of the residential real estate industry should be *sales-weighted*. This is consistent in treatment to the acquisition or purchase of goods and services in a Consumer Price Index (CPI).

13.3 Index Scope

A price index which is required to measure the *wealth* associated with the ownership of residential property should cover all residential property, that is, both existing properties and properties which have been recently built.[2] This is also the case for an index being used as a financial stability indicator.

A price index which is required for measuring real investment in the residential real estate industry should cover sales of new property.[3] The construction part of new housing produced is part of *gross investment*. The cost of the land, apart from the value of any improvements made to this element, should be excluded for this purpose; see also section 13.5. However, as was explained in chapter 4, a price index for sales of both new and existing houses is required in order to construct real output measures for the activities of real estate agents in selling new and existing houses to purchasers. The scope of the index for this application should cover both the structure and land values of the residential property sales.

A price index restricted to new properties is also appropriate when a residential property price index is an input into a CPI for the measurement of owner-occupier housing costs on a net-acquisition cost basis, that is, where the CPI covers the cost of acquiring properties which are new to the owner-occupier housing market. This approach, one of a number of alternatives as

[1] See SNA (1993) and SNA (2009).

[2] This includes conversions of existing property, for example where a warehouse has been converted into flats or an existing property has been sub-divided.

[3] Renovations to existing dwelling units are also part of residential construction investment.

was explained in chapter 4, treats the purchase of a dwelling exactly like the purchases of any other consumption good.[4]

13.4 Constant Quality

Regardless of the different uses of the index, the purpose of a residential property price index is to compare the values of the sales or of the stock of residential property between two time periods after allowing for changes in the attributes of the properties. For this purpose it is necessary to decompose price changes into those associated with changes in attributes and the residual which relates to the underlying "pure price" change.

A constant quality price index is appropriate for both a stock and sales-weighted price index. There are a number of practical methodologies which can be used to construct such an index. Recommendations on which of the available methods should be used in which circumstances are given later in this chapter.

13.5 Decomposition between the Building and Land Components

A decomposition of the RPPI in structures and land components may be required, particularly if a country's balance sheet estimates of national wealth in the National Accounts make this distinction.

Such a decomposition may also be necessary when a residential property price index is an input into a CPI for the measurement of owner-occupier housing using the net-acquisition approach.

13.6 Statistical Methods for Compiling Constant Quality Indexes

The methods adopted by statistical agencies to construct constant quality RPPIs vary among countries and are dictated in large part by the availability of data generated by the processes involved in buying and selling a property.

The challenges of compiling constant-quality residential price indexes can be summarized by the following three factors:

- Residential properties are notoriously heterogeneous. No two properties are identical.

[4] The argument in favour of the net acquisition approach is that it is the closest to the "acquisition" approach which has traditionally been adopted for other parts of a CPI and is most appropriate for a CPI being used as a general indicator of current economic conditions. But the method can draw criticism from those who require a CPI as a compensation index, as neither the weight nor the price indicator properly reflect the shelter costs of owner-occupiers. For instance, a rise in interest rates would not be reflected in a net acquisition index. See CPI Manual (2004) and the *Practical Guide to Producing Consumer Price Indices* (United Nations, 2009).

- Prices are often negotiated. The (asking) price of a property is not fixed and can change throughout the transaction process until the price is finalised. This means that a property's market value can only be known with certainty after it has been sold.[5]
- Property sales are infrequent. In many countries, less than ten per cent of the housing stock changes hands every year, which means that a particular house is likely to be resold approximately once every ten years.

The different methods of index construction used by a statistical agency reflect the differing solutions used to meet the above challenges. Four methods have been studied in depth in this handbook: stratification or "mix-adjustment", hedonic regression methods, repeat sales, and appraisal-based methods (i.e., the SPAR method). Below, recommendations are made on each. Each method attempts to adjust for the change in the "quality mix" of the houses whose prices are observed and combined to construct the index. Some methods, however, are unable to adjust for quality changes of the individual houses, i.e. for the net effect of depreciation of the structures and renovations and extensions. Where data from the administrative processes for buying and selling a residential property are used in the construction of the index, the price will usually relate either to the final offer price or to the selling price – these can differ from one another.

The recommendations do not address the challenge of computing an RPPI in countries where a significant proportion of the housing stock relates to informal or traditional dwellings. An example of computing an RPPI under the latter circumstances is given in Chapter 11 and draws on the experience of South Africa. In such circumstances it is not possible to be very prescriptive in terms of recommendations since the situation will vary considerably among countries and there is no ideal solution that will deliver a residential property price index which is conceptually pure and does not generate practical difficulties. Rather, the compiler will need to draw on the best available sources of information and will no doubt have to make conceptual and methodological compromises in computing an index. In these circumstances it is particularly important that statistical agencies provide evaluations of the resulting price indexes and guide users on their uses.

13.6.1 Stratification or mix-adjustment

Stratification or mix-adjustment is the most straightforward way to control for changes in the composition or quality mix of the properties sold. It is also addresses any user need for sub-indexes relating to different housing market segments. The effectiveness of stratification will depend upon the stratification variables used because a mix-adjusted measure only controls

[5] In some cases even the selling prices may not reflect the "true" market values, for example when they relate to distressed sales arising from divorce etc.

for compositional change across the various groups – a mix-adjusted index does not account for changes in the mix of properties sold within each subgroup or stratum.

In theory, the more detailed the stratification, the more the index controls for changes in the characteristics of the properties covered by the index. However, an increasing of the number of strata reduces the average number of price observations per stratum and in fact can quickly lead to empty strata. Strata or cells which are empty then lead to a lack of matching when the average price and quantity data in each cell are compared across two time periods. A very detailed stratification might also raise the standard error of the overall index. In addition, it may not be easy to identify the most important price-determining characteristics in the way that a method using hedonic regression can do (see next section).

The main advantages of stratification/mix-adjustment are:

- Depending on the choice of stratification variables, the method adjusts for compositional change amongst the dwellings.
- The method is reproducible, conditional on an agreed list of stratification variables.
- It is not subject to revision.
- Price indexes can be constructed for different types and locations of housing.
- The method is relatively easy to employ and to explain to users.

The main disadvantages of stratification/mix-adjustment are:

- It cannot deal adequately with depreciation of the houses unless the age of the structure is a stratification variable. The latter can result in problems associated with cells with small numbers of price observations.
- The method cannot deal adequately with houses which have undergone major repairs or renovations (unless information on renovations is available).
- It requires information on housing characteristics that are included in the strata so that the sales can be allocated to the correct strata.
- If the stratification scheme is very coarse, compositional changes will affect the indexes.
- If the stratification scheme is very fine, the cells can be subject to considerable sampling variability due to small sample sizes or some cells may simply be empty for some periods causing index number difficulties.
- The value of land cannot be separated out using this method.

Stratification/mix-adjustment is an appropriate method where

- an appropriate level of detail is chosen for the cells and can be applied in practice;
- the age of the structure is one of the stratification variables;

- a decomposition of the index into structure and land components is not required.

Stratification/mix-adjustment is recommended where the volume of sales is large enough to support a detailed classification of properties. It is second best to hedonic regression if there are sufficient data available to use the latter method.

13.6.2 Hedonic regression

The application of hedonic techniques for the quality adjustment of prices and for computing price indexes has made a significant contribution to the methodological development of price indexes in recent years and is rapidly becoming a preferred method for compiling constant-quality residential property price indexes. There is no uniformity in the practical application of hedonic regression, but the idea underlying hedonics is rather simple. Hedonic regression is a statistical technique that measures the relationship between the observable characteristics of a good or service and its price or value. In the context of residential property price indexes, the "best" form of the hedonic function may be linear rather than log-linear to reflect the fact that the value of a property is generally equal to the sum of the price of the structure and the price of the land.

There are basically two alternative methods of application of hedonics to residential property:

- The time dummy variables method. This method generally uses a single regression, with time dummies and fixed characteristics coefficients, which covers all periods and which is re-run each time the price index is compiled. The (exponentials of the) time dummy coefficients are taken to represent the period by period price changes excluding quality (mix) changes. This method has the benefit of simplicity. One of the drawbacks is that it raises the issue of "revisability" of the index because the time dummy coefficients will be updated each time the regression is run.

  However, there is a variant of the time dummy method, called the rolling window time dummy method, which can work well in practice and solves the revisability problem. A hedonic regression is run using the data for the last N periods and the last time dummy is used as a chain link factor for updating the index for the previous period. For references to the literature on this method and an example, see section 9.7.

- The hedonic imputation method. A separate hedonic regression is performed in each time period and the "missing" current period prices for the properties sold in the base period are imputed using the predicted prices from the estimated hedonic equation. A symmetric approach is possible by also imputing the "missing" base period prices for the properties sold in the current period and then taking the geometric mean of both hedonic imputation indexes.

Both hedonic regression methods can potentially suffer from omitted variable bias if some important price determining characteristic is omitted from the equation. Multicollinearity can be a practical problem, particularly when a decomposition of the index into structures and land components is required. The time dummy variable method has been frequently used by academics, in part due to its simplicity, but the hedonic imputation method is more flexible – it allows characteristics prices to change independently over time whereas the time dummy method forces characteristics prices to move in a proportional manner – and is essentially similar to the traditional matched-model methodology to compute price indexes.

Hedonic regression methods can be used in conjunction with stratification to deal with any residual quality-mix change that remains within the strata. This has the added advantage of dealing with the fact that different model specifications may be needed for different segments of the housing market or that the "value" of some characteristics will vary across different market segments.

The main advantages of hedonics are:

- If the list of property characteristics is sufficiently detailed, the method adjusts for both sample mix changes and quality changes over time of the individual houses.
- Price indexes can be constructed for different types of dwellings and locations through stratification and the application of hedonics to each individual stratum.
- The method maximizes the use of the available data.
- It can in principle be used to decompose the overall price index into land and structures components, subject to the availability of data.

The main disadvantages of hedonic regression are:

- The method is generally regarded as being data intensive.[6]
- It may be difficult to control sufficiently for location if property prices and price trends differ across detailed regions.
- The method can be sensitive to the variables used in the regression and the functional form for the model.
- The method is not particularly easy to explain to users and from their perspective may lack transparency.

The hedonic regression method is generally the best technique for constructing a constant quality residential property price index. Subject to the required data being available hedonic

[6] As was seen in previous chapters, in some cases satisfactory results can be obtained with hedonic regression methods using only a few characteristics.

regression using the imputations approach is the recommended method. It is furthermore recommended that stratified hedonic indexes be computed.

13.6.3 Repeat Sales

The repeat sales method observes the price development of a specific house over a period of time by reference to the selling price each time it is sold. The price change of a selection of houses during overlapping time periods can then be observed to estimate, using a simple dummy variable regression model, the general trend in residential property prices. Measuring the average price changes in repeat sales on the same properties ensures a "like for like" comparison (ignoring the fact that depreciation and renovations on the structure between the periods of sale can change the property).

The main advantages of the repeat sales method are:

- In its basic form, it requires no information on characteristics of the dwelling units other than the addresses of the properties that are traded. Source data are often available from administrative records.
- It follows a matched-model methodology, under the assumption that depreciation and renovations have not changed the dwelling unit over the time period between subsequent sales.
- Standard repeat sales regressions are easy to run and the resulting price indexes are easy to construct.
- No imputations are involved. By construction, location is automatically controlled for.
- The results are, in principle, reproducible.

The main disadvantages of the repeat sales method are:

- The method does not use all of the available selling prices; it uses information only on those properties that have sold more than once during the sample period.
- The standard version of the method ignores (net) depreciation of the dwelling unit.
- Sample selection bias can arise from the restriction to properties that have been sold more than once during the sample period.
- The method cannot generate separate price indexes for structures and for land.
- The reliance on repeat sales means that there may not be enough data points to compute monthly residential property price indexes for smaller categories of property.

- The sample is updated as new transaction information becomes available. This means that the repeat sales property price index could be subject to backward revision over a long time period.[7]
- Since a house must be sold at least twice in a repeat sales index, newly built dwelling units are excluded from such an index.

Although a natural starting point for constructing an index, the repeat sales method is not preferred above the hedonic method for constructing a constant quality residential property price index. However, it can offer a solution where there is limited or no information on housing characteristics and there are a relatively large number of repeat transactions to provide enough data points for the required types of residences and where sample selection bias is not considered a problem. It is not recommended when a distinction needs to be made between the price of the structure and the price of the land.

13.6.4 Appraisal-Based Methods

Appraisal-based methods use "assessed" values, such as valuations for taxation purposes or valuations from specially commissioned surveys using estate agents, often done by reference to similar properties that have been sold, to overcome the two main problems associated with the repeat sales methodology – the relatively small number of price observations which are generated and the susceptibility to sample selection bias. Where the valuations all refer to a standard reference period, the matched model methodology which underlies appraisal-based methods also has the advantage that it can be applied in a straightforward way and with no necessity to use econometrics to adjust for compositional changes. However, like the repeat sales methodology, appraisal-based methods generally cannot deal adequately with quality changes to individual houses. Also, they generally rely on expert judgment on how much a property would sell for rather than on an actual transaction price. Thus, it can be argued, at the extreme, that appraisal-based methods are influenced by judgments or opinions, albeit authoritative and objective.

The Sale Price Appraisal Ratio (or SPAR) method uses appraisals with a common reference period as base period prices in a standard matched-model framework (though the results are normalized to obtain an index that equals 1 in the base period). The experiences of the few countries that have computed a SPAR index are generally positive although some researchers reported bias arising from frequent re-assessments and reduced precision over time arising from new appraisals. The main advantages of the SPAR method are:

---

[7] In practice, the link factor for the last two periods in the current repeat sales regression can be used to update the ongoing index.

- Being based on the standard matched model methodology, it is consistent with traditional index number theory.
- It is straightforward to compute.
- The method benefits from many more observations than the repeat sales method and is therefore less susceptible to problems arising from having a relatively small number of price observations.
- It is less susceptible to sample selection bias than the repeat sales method.
- It does not suffer from revisions to previously computed figures.
- It is reproducible.

The main disadvantages of the SPAR method are:

- It cannot deal adequately with quality changes (depreciation and renovations and) of the dwelling units.[8]
- Data on value assessments at the address level must be available for all properties.
- The method is dependent on the quality of the assessments.
- It cannot be used to decompose the overall property price index into land and structures components.[9]

The SPAR methodology addresses some of the weaknesses of the repeat sales methodology and is to be preferred to the latter methodology if assessment data of sufficient quality is available. The SPAR methodology does have its drawbacks but is a recommended when the use of hedonics is not possible, in particular in combination with stratification.

13.7 Seasonal Adjustment

If the initial house price series indicates that some seasonal fluctuations occur, then normal seasonal adjustment techniques can be used in order to seasonally adjust the initial series. However, if the hedonic imputation or the stratification method is used to construct the initial index, some more specific recommendations are made below.

---

[8] As with the repeat sales method, the price index generated by the SPAR method can in principle be adjusted by using exogenous information on the net depreciation of properties of the type being considered.

[9] Where official decompositions of the total assessed value of the property into land and structures components are available, these could be used to check the land and structures price indexes that are generated by hedonic regression methods.

If the stratification method is used to construct the initial index and it exhibits seasonality, then the rolling year method explained in Chapter 6 can be applied to seasonally adjust the series without relying on econometric methods.

If the hedonic imputation method is used to construct the initial price index and it exhibits seasonality, then in order to obtain a seasonally adjusted series, it may be useful to construct year over year monthly or quarterly series as an initial step. These initial series can then be aggregated using the rolling year method into a smooth seasonally adjusted series.

A Glossary of Main Terms

Acquisitions approach

An approach in which consumption is identified with the goods and services acquired by a household in some period (as distinct from those wholly or partially used up for purposes of consumption). See also *net acquisitions approach*.

Aggregate

A set of transactions relating to a specified flow of goods and/or services, such as the total purchases made by households on residential property in some period. The term "aggregate" is also used to mean the value of the designated set of transactions.

Aggregation

Combining, or adding, different sets of transactions to obtain larger sets of transactions. The larger set is described as having a higher *level* of aggregation than the (sub-)sets of which it is composed. The term "aggregation" is also used to mean the process of adding the values of the lower-level aggregates to obtain higher-level aggregates. In the case of price indexes, it means the process by which price indexes for lower-level aggregates are averaged to obtain price indexes for higher-level aggregates.

Asking price

The price at which the property is offered for sale. The asking price can be adjusted during the process of buying and selling a house until the final transaction price is reached.

Assessed value or appraisal

Valuation of the market value of a property. Valuations may be needed to obtain a mortgage loan. In some countries assessments are performed on the government's behalf for (property) tax purposes. Assessed property values are often referred to as appraisals. See also *Sale Price Appraisal Ratio method*.

Asset

Axiomatic, or test approach

The approach to index number theory that determines the choice of index number formula on the basis of its mathematical properties. A list of tests is drawn up, each test requiring an index to possess a certain property or satisfy a certain axiom. An index number may then be chosen on the basis of the number of tests satisfied. Not all tests may be considered to be equally important and the failure to satisfy one or two key tests may be considered sufficient grounds for rejecting an index.

Base period
The base period is usually understood to mean the period with which all the other periods are compared. The term may, however, have different meanings in different contexts. Three types of base period may be distinguished:

- the *price reference period* – the period that provides the prices to which the prices in other periods are compared. The prices of the price reference period appear in the denominators of the price relatives, or price ratios, used to calculate the index;
- the *weight reference period* – the period of which the expenditures serve as weights for the index. If the expenditures are hybrid (i.e., if the quantities of one period are valued at the prices of some other period), the weight reference period is the period to which the quantities refer;
- the *index reference period* – the period for which the value of the index is set equal to 100.

It should be noted that, in practice, the duration of the weight reference period for an RPPI is often a year, whereas the RPPI is typically calculated monthly or quarterly, the duration of the price reference period being a month or quarter. Thus, the weight and price reference period may not coincide in practice, at least when an RPPI is first calculated, although the price and index reference periods frequently coincide.

Bias
A systematic tendency for the calculated RPPI to diverge from some ideal or preferred index, resulting from the method of data collection or processing or the index formula used. See also *sample selection bias*.

Chain index
An index number series for a long sequence of periods obtained by linking together index numbers spanning shorter sequences of periods. A chain index, computed according to some index number formula (such as the Fisher), is the product of period-on-period indexes which are computed with the same formula. See also *Linking*.

Characteristics
The physical and economic attributes of a good or service that serve to identify it and enable it to be classified. For residential property these relate to both the structure (the building) and the location/land.

Characteristics prices approach

Component
A set of the goods and services that make up some defined aggregate. Also used in the context of decomposing the price property price (index) into land and structures components.

Consistency in aggregation
An index is said to be consistent in aggregation when the index for some aggregate has the same value whether it is calculated directly in a single operation, without distinguishing its components, or whether it is calculated in two or more steps by first calculating separate indexes, or sub-indexes, for its components, or sub-components, and then aggregating them, the same formula being used at each step.

Consumer price index (CPI)
A monthly or quarterly price index compiled and published by an official statistical agency that measures changes in the prices of consumption goods and services acquired or used by households. Its exact definition, including the treatment of *owner-occupied housing*, may vary from country to country.

Coverage
The set of properties of which the prices are actually included in a price index. For practical reasons, coverage may have to be less than the ideal scope of the index. That is, the types of property actually priced may not cover all of the types that are sold or that are available in the housing stock.

Current period, or comparison period
In principle, the current period refers to the most recent period for which the index has been compiled or is being compiled. The term is widely used, however, to mean the comparison period; that is, the period that is compared with the base period, usually the price reference or index reference period. It is also used to mean the later of the two periods being compared. The exact meaning is usually clear in the context.

Data clean(s)ing

Deflation
The division of the current value of some aggregate by a price index (described as a *deflator*) in order to revalue its quantities at the prices of the price reference period.

Depreciation

Domain
An alternative term for the scope of an index.

Drift
A chain index is said to drift if it does not return to unity when prices in the current period return to their levels in the base period. Chain indexes are liable to drift when prices fluctuate over the periods they cover.

Durable consumption good
A consumption good that can be used repeatedly or continuously for purposes of consumption over a long period of time, typically several years. A house is an extreme form of a durable consumption good due to its very long expected lifetime. This has led to different approaches to the treatment of *owner-occupied housing* in economic statistics like the national accounts and the CPI.

Economic approach
The economic approach to index number theory assumes that the quantities are functions of the prices, the observed data being generated as solutions to various economic optimization problems. While this approach is very relevant for the CPI as an approximation to a cost-of-living index, it is less relevant for a residential property price index. See also *axiomatic or test approach*.

Editing
The process of scrutinizing and checking the prices reported by price collectors. Some checks may be carried out by computers using statistical programs written for the purpose. See also *data cleaning*.

Elementary aggregate
Usually defined as the smallest aggregate for which expenditure data are available and used for index construction purposes. Elementary aggregates also serve as strata for the sampling of items to be priced. The values of the elementary aggregates are used to weight the price indexes for elementary aggregates to obtain higher-level indexes.

In the context of a sales-based residential property price index, the term elementary aggregate is not very appropriate. Since every property is basically unique, the quantities are equal to 1, so that weights are available at the most detailed level.

Existing dwellings

Fisher price index

The geometric average of the Laspeyres price index and the Paasche price index. The Fisher index is *symmetric* and *superlative*. Sales based residential property price indexes can always be computed using the Fisher formula because the quantities are equal to 1 (as each dwelling is essentially a unique good).

Fixed weight indexes
An abbreviated description for a series of weighted arithmetic averages of price relatives of price indexes where the weights are kept fixed over time. In a RPPI context, the weights can be sales (expenditure) weights or stock weights.

Geometric Laspeyres index
A weighted geometric average of the price relatives using the expenditure shares of the price reference period as weights.

Goods
Physical objects for which a demand exists, over which ownership rights can be established and for which ownership can be transferred between units by engaging in transactions on the market.

Hedonic regression
The estimation of a hedonic model, using regression techniques, that explains the price of the property as a function of its characteristics (relating to the structures as well as the location). See also *hedonic imputation approach* and *time dummy variable hedonic approach*.

Hedonic imputation approach
An approach to estimating a quality-adjusted residential property price index where “missing” prices are imputed using a hedonic regression model. The model parameters are re-estimated in each time period, which makes this approach more flexible than the time dummy variable hedonic approach.

Households
Households may be either individual persons living alone or groups of persons living together who make common provision for food or other essentials for living. Most countries choose to exclude groups of persons living in large institutional households (barracks, retirement homes, etc.) from the scope of their CPIs.

Housing stock

Hybrid models

Identity test
A test under the axiomatic approach that requires that, if the price of each item remains the same between the periods compared, the price index must equal unity.

Imputed price
The price assigned to an item (i.e., a property) for which the price is "missing" in a particular period. The term "imputed price" may also refer to the price assigned to a good or service item that is not sold on the market, such as a good or service produced for own consumption, including housing services produced by owner-occupiers. See also Rental equivalence.

Index reference period
The period for which the value of the index is set at 100 (or, alternatively, 1).

Informal market

Jevons price index
An elementary price index defined as the unweighted geometric average of the sample price relatives.

Laspeyres price index
A price index in which the quantities of the goods and services refer to the earlier of the two periods compared, the price reference period. The Laspeyres index can also be expressed as a weighted arithmetic average of the price relatives with the expenditure shares in the earlier period as weights. The earlier period serves as both the weight reference period and the price reference period.

Least squares regression

Linking
Splicing together two consecutive series of price observations, or price indexes, that overlap in one or more periods. If the two sequences overlap by a single period, the usual procedure is simply to rescale one or other sequence so that the value in the overlap period is the same in both sequences and the spliced sequences form one continuous series.

Lowe price index
A price index that measures the change between periods 0 and t in the total value of a set of goods and services at fixed quantities. The quantities do not necessarily have to consist of the actual quantities in some period. The class of indexes covered by this definition is very broad

and includes, by appropriate specification of the quantity terms, the Laspeyres and Paasche indexes.

Lower-level index
An sub-index as distinct from an aggregate index.

Matched models approach
The practice of pricing exactly the same product, or model, in two or more consecutive periods. It is designed to ensure that the observed price changes are not affected by quality change. The change in price between two perfectly matched products is described as a pure price change.

Market value
The value of a property at a certain point of time, or the price that would result if the property would be sold in a "free market".

Mean index
A price index that is calculated as the ratio of the sample means (*unit values*) of the properties sold in two periods.

Median index
A price index that tracks the change of the median property price over time. The median of a sample distribution is ….

Mix adjustment

Money outlays or payments approach
One of the three main approaches to including Owner Occupied Housing into a Consumer Price Index. The out of pocket expenses relating to home ownership are simply added up.

Monotinicity

Multicollinearity

Net acquisitions approach
One of the three main approaches to including Owner Occupied Housing into a Consumer Price Index. Dwellings added to the owner occupied housing stock (in general mainly newly-built dwellings) are part of the coverage of the index; existing dwellings are excluded. See also *Acquisitions approach*.

Nonlineair regression

Sample
A (random or non-random) selection of elements from a finite population. In the housing context, the properties sold in some time period can be viewed as a sample from the housing stock. This sampling view is particularly relevant for a stock based residential property price index.

Offer price
See Asking price.

Outlier
A term that is generally used to describe any extreme value in a set of survey data. In a RPPI context, it is used for an extremely high or low property price or price relative, which requires further investigation and should be deleted when deemed incorrect.

Owner-occupied housing
Dwellings owned by the households that live in them. The dwellings are fixed assets that their owners use to produce housing services for their own consumption, these services being usually included within the scope of a CPI. The rents may be imputed by the rents payable on the market for equivalent accommodation or by user costs. See also *rental equivalence* and *User cost*.

Paasche price index
A price index in which the quantities of the goods and services considered refers to the later of the two periods compared. The later period serves as the weight reference period and the earlier period as the price reference period. The Paasche index can also be expressed as a weighted harmonic average of the price relatives that uses the actual expenditure shares in the later period as weights.

Payments approach
See *money outlays approach*.

Price index

Price reference period
The period of which the prices appear in the denominators of the price relatives. See also *Base period*.

Price relative
The ratio of the price of an individual product in one period to the price of that same product in some other period.

Products
A generic term used to mean a good or a service. Individual sampled products selected for pricing are often described as items.

Pure price change
The change in the price of a property of which the characteristics are unchanged or the change in the property price after adjusting for any change in quality (due to renovations, extensions and depreciation).

Quality change

Quality adjustment
An adjustment to the change in the price of a property of which the characteristics change over time that is designed to remove the contribution of the change in the characteristics to the observed price change. In practice, the required adjustment can only be estimated. Different methods of estimation, including hedonic methods, may be used in different circumstances. These methods can also be used to control for compositional or quality mix changes over time in the samples of properties sold.

Quantity relative
The ratio of the quantity of a product in one period to the quantity of that same product in some other period.

Rental equivalence approach
One of the three main approaches to including Owner Occupied Housing into a Consumer Price Index. The imputed price for shelter costs should equal the price at which the dwelling could be rented.

Repeat sales method
This method to compile a residential property price index compares properties that were sold twice or more in the data set at hand. It is a regression-based approach that only includes time dummy variables.

Representative property

A property, or category of properties, that accounts for a significant proportion of the total expenditures within some aggregate, and/or for which the average price change is expected to be close to the average for all properties within the aggregate.

Residential property

Revisions

Reweighting
Replacing the weights used in an index by a new set of weights.

Rolling window approach

Sample selection bias
Bias in an index that can result when the sample is not representative of the population. In the housing context, the sample of properties may either not be representative of all sales (which is particularly relevant for a sales based index) or not be representative of the housing stock (which is relevant for a stock based index). In all sales are observed, there will be no sample selection bias in a sales based property price index.

Sampling frame
A list of the units in the universe from which a sample of units can be selected. The list may contain information about the units, which may be used for sampling purposes. Such lists may not cover all the units in the designated universe and may also include units that do not form part of that universe.

Scope
The set of products for which the index is intended to measure the price changes. The coverage of an index denotes the actual set of products included, as distinct from the intended scope of the index.

Seasonal goods
Seasonal goods are goods that either are not available on the market during certain seasons or periods of the year, or are available throughout the year but with regular fluctuations in their quantities and prices that are linked to the season or time of the year.

Selling (or transaction) price
The final transaction price of a property.

Specification
A description or list of the characteristics that can be used to identify an individual dwelling unit to be priced.

SPAR method

Stratification method
This method to compile a residential property price index subdivides the sample of properties sold into a number of relatively homogeneous strata or cells, according to a (limited) number of price determining characteristics. Average prices (unit values) or median prices will then be used to compute price indexes for each stratum. In the second stage, these straum indexes are aggregated up using sales weights or stock weights. Also known as *mix adjustment*.

Stratification can also be used in conjunction with other methods to control for quality (mix) changes, for example with hedonic regression, repeat sales or SPAR methods.

Superlative index
Superlative indexes are generally symmetric and have good properties from an index number theoretic point of view. Examples are the (arithmetic) Fisher and the (geometric) Törnqvist index.

Symmetric index
An index that treats both periods symmetrically by attaching equal importance to the price and expenditure data in both periods. The price and expenditure data for both periods enter into the index formula in a symmetric way.

System of National Accounts (SNA)
A coherent, consistent and integrated set of macroeconomic accounts, balance sheets and tables based on internationally agreed concepts, definitions, classifications and accounting rules. Household income and consumption expenditure accounts form part of the SNA.

Time dummy variable (hedonic) approach

Unit value or average value
The unit value of a set of homogeneous products is the total value of the purchases/sales divided by the sum of the quantities. It is therefore a quantity-weighted average of the different prices at which the product is purchased/sold. Unit values may change over time as a result of a change in the mix of the products sold at different prices, even if the prices do not change.

User cost

The cost incurred over a period of time by the owner of a fixed asset or consumer durable as a consequence of using it to provide a flow of capital or consumption services. User cost consists mainly of the depreciation of the asset or durable (measured at current prices and not at historic cost) plus the capital, or interest, cost.

Uses approach

An approach to CPIs in which the consumption in some period is identified with the consumption goods and services actually used up by a household to satisfy their needs and wants (as distinct from the consumption goods and services acquired). In this approach, the consumption of consumer durables in a given period is measured by the values of the flows of services provided by the stocks of durables owned by households. These values may be estimated by the user costs.

Value

Price times quantity. The value of the expenditures on a set of homogeneous products can be factored uniquely into its price, or unit value, and quantity components. Similarly, the change over time in the value of a set of homogeneous products can be decomposed uniquely into the change in the unit value and the change in the total quantities. There are, however, many ways of factoring the change over time in the value of a set of *heterogeneous products* into its price and quantity components.

In a housing context, value may also refer to a single property. The "price" of a property is actually a value as it is made up of the price of the structures and the price of the land that the structure is built on.

Wealth

Weight reference period

The period of which the expenditure shares serve as the weights or of which the quantities make up the set of properties for a Lowe index. There may be no weight reference period when the expenditure shares for the two periods are averaged, as in the Törnqvist index, or when the quantities are averaged, as in the Walsh index. See also *base period*.

Weights

A set of numbers summing to unity that are used to calculate averages. In a RPPI context, the weights are generally expenditure (sales) or stock value shares that sum to unity by definition. They are used to average price relatives for individual properties

Bibliography

Abraham, J.M. and W.S. Schauman (1991), "New Evidence on Home Prices from Freddie Mac Repeat Sales", *American Real Estate and Urban Economics Association Journal* 19, 333-352.

Acadametrics (2009), House Price Indices – Fact or Fiction, www.acadametrics.co.uk.

Alterman, W., W.E. Diewert and R.C. Feenstra (1999), "*International Trade Price Indexes and Seasonal Commodities*", Department of Labor, Bureau of Labor Statistics, Washington, DC: US Government Printing Office.

Australian Bureau of Statistics (ABS) (2005), "Renovating the Established House Price Index", Information Paper, Cat. no. 6417.0.

Australian Bureau of Statistics (ABS) (2006), "A Guide to House Price Indexes", Information Paper, Cat. no. 6464.

Bailey, M.J., R.F. Muth and H.O. Nourse (1963), "A Regression Method for Real Estate Price Construction", *Journal of the American Statistical Association* 58, 933-942.

Baldwin, A., A. Nakamura and M. Prud'homme (2010), "An Empirical Analysis of the Different Concepts for Owned Accommodation in the Canadian CPI: The Case of Ottawa, 1996-2005", paper presented at the joint UNECE and ILO Meeting on Consumer Price Indices, Geneva, May 10-12.

Balk, B.M. (1998), "On the Use of Unit Value Indices as Consumer Price Subindices", in: Lane, W. (ed.) *Proceedings of the Fourth Meeting of the International Working Group on Price Indices*, Washington, DC: Bureau of Labor Statistics.

Balk, B.M. (2008), *Price and Quantity Index Numbers; Models for Measuring Aggregate Change and Difference*, New York: Cambridge University Press.

Belski, E. and J. Prakken (2004), "Housing Wealth Effects: Housing's Impact on Wealth Accumulation, Wealth Distribution and Consumer Spending", Working paper W04-13, Harvard University, Joint Center for Housing Studies.

Berndt, E.R. (1991), *The Practice of Econometrics: Classic and Contemporary*, Reading, Massachusetts: Addison-Wesley Publishing Company.

Blades, D. (2009), "Housing in ICP 2011: Issues to be Resolved", paper presented at the Technical Advisory Group Meeting, October 1-2, 2009, World Bank, Washington, DC.

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

Bover, O. and M. Izquierdo (2003), "Quality-adjusted Prices: Hedonic Methods and Implications for National Accounts", *Investigaciones Económicas* 27, 199–238.

Bourassa S. C., M. Hoesli and J. Sun (2006), "A Simple Alternative House Price Index Method", *Journal of Housing Economics* 15(1), 80-97.

Butler, J.S., Y. Chang and A. Crews Cutts (2005), "Revision Bias in Repeat-Sales Hone Price Indices", Freddie Mac Working Paper no. 05-03.

Calhoun, C.A. (1996), "OFHEO House Price Indexes: HPI Technical Description", Office of Federal Housing Enterprise Oversight, Washington, DC.

Campbell, J.Y. and J.F. Cocco (2007), "How Do House Prices Affect Consumption? Evidence from Micro Data", *Journal of Monetary Economics* 54(3), 591-621.

Carare, A. and M. R. Stone (2003), "Inflation Targeting Regimes", IMF Working Paper WP 03/9.

Case, B., H.O. Pollakowski and W.M. Wachter (1991), "On Choosing Between House Price Index Methodologies", *American Real Estate and Urban Economics Association Journal* 19, 286-307.

Case, B., H. Pollakowski, and S. Wachter (1997), "Frequency of Transaction and House Price Modeling", *Journal of Real Estate Finance and Economics* 14(1-2), 173-187.

Case, B. and J. M. Quigley (1991), "The Dynamics of Real Estate Prices", *Review of Economics and Statistics* 22, 50-58.

Case, K.E., Quigley, J.M. and Shiller, R.J. (2001), "Comparing Wealth Effects: The Stock Market Versus the Housing Market." National Bureau of Economic Research Working Paper Series, No. 8606.

Case, K.E. and R.J. Shiller (1987), "Prices of Single-Family Homes Since 1970: New Indexes for Four Cities", *New England Economic Review*, September-October, 45–56.

Case, K.E. and R. J. Shiller (1989), "The Efficiency of the Market for Single-family Homes", *The American Economic Review*, 79 (1), 125-137.

Case, K.E. and R.J. Shiller (1990), "Forecasting Price and Excess Returns in the Housing Market", *American Real Estate and Urban Economics Association Journal* 18, 253-273.

Case, B. and S. Wachter (2005), "Residential Real Estate Price Indices as Financial Soundness Indicators: Methodological Issues", pp. 197-211 in *Real Estate Indicators and Financial Stability*, BIS Papers No 21, Bank for International Settlements, Washington, DC.

Chinloy, P., M. Cho and I.F. Megbolugbe (1997), "Appraisals, Transactions Incentives and Smoothing", *Journal of Real Estate Finance and Economics* 14(1), 45-55.

Chiodo, A.J., Hernandez-Murillo, R., and M.T. Owyang (2010), "Non-linear Effects of School Quality on House Prices", *Federal Reserve Bank of St. Louis Review*, May/June, 185-204.

Cho, M. (1996), "House Price Dynamics: A Survey of Theoretical and Empirical Issues", *Journal of Housing Research* 7(2), 145–172.

Cleassens, S.M, A. Kose and M.E. Terrones (2008), "What Happens During Recessions, Crunches and Busts?", IMF Working paper 08/274, International Monetary Fund, Washington, DC.

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. (1980), "The Elasticity of Substitution for Land: The Effects of Measurement Errors", *Journal of Urban Economics* 8, 255-263.

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-Sale 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 Indices: Here Today, Gone Tomorrow?", *Real Estate Economics* 27(1), 79-104.

Clapp, J.M., C. Giaccotto and D. Tirtiroglu (1991), "Housing Price Indices Based on All Transactions Compared to Repeat Subsamples", *AREUEA Journal* 19(3), 270-285.

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.

Colwell, P.F. and G. Dilmore (1999), "Who was First? An Examination of an Early Hedonic Study", *Land Economics* 75(4), 620-626.

Congressional Budget Office (2007), "Housing Wealth and Consumer Spending", Background paper, January. Available at http://www.cbo.gov.

Court, A.T. (1939), "Hedonic Price Indexes with Automotive Examples", in *The Dynamics of Automobile Demand*, General Motors Corporation, New York, 99–117.

Crone, T.M. and R.P. Voith (1992), "Estimating House Price Appreciation: A Comparison of Methods", *Journal of Housing Economics* 2(4), 324-338.

Crone, T.M., L.I. Nakamura and R.P. Voith (2009), "Hedonic Estimates of the Cost of Housing Services: Rental and Owner Occupied Units", pp. 67-84 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press and www.indexmeasures.com.

Dale-Johnson, D, (1982), "An Alternative Approach to Housing Market Segmentation Using Hedonic Price Data", *Journal of Urban Economics* 11(3), 311–332.

Davidson, R. and J.G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

Davis, M.A. and J. Heathcote (2007), "The Price and Quantity of Residential Land in the United States", *Journal of Monetary Economics* 54, 2595-2620.

Davis, M.A. and M.G. Palumbo (2008), "The Price of Residential Land in Large US Cities", *Journal of Urban Economics* 63, 352-384.

Destatis (2009), *Handbook on the Application of Quality Adjustment Methods in the Harmonised Index of Consumer Prices*, Statistics and Science, vol. 13, Wiesbaden: Federal Statistical Office of Germany.

Diewert, W.E. (1974), "Intertemporal Consumer Theory and the Demand for Durables," *Econometrica* 42, 497-516.

Diewert, W.E. (1980), "Aggregation Problems in the Measurement of Capital", pp. 433-528 in *The Measurement of Capital*, D. Usher (ed.), Chicago: The University of Chicago Press.

Diewert, W.E. (1983), "The Treatment of Seasonality in a Cost of Living Index", pp. 1019-1045 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada.

Diewert, W.E. (1998), "High Inflation, Seasonal Commodities and Annual Index Numbers", *Macroeconomic Dynamics* 2, 456-471.

Diewert, W.E. (1999), "Index Number Approaches to Seasonal Adjustment", *Macroeconomic Dynamics* 3, 1-21.

Diewert, W.E. (2002), "Harmonized Indexes of Consumer Prices: Their Conceptual Foundations", *Swiss Journal of Economics and Statistics* 138, 547-637.

Diewert, W.E. (2003a), "Hedonic Regressions: A Consumer Theory Approach", pp. 317-348 in *Scanner Data and Price Indexes*, Studies in Income and Wealth, Volume 64, R.C. Feenstra and M.D. Shapiro (eds.), NBER and University of Chicago Press.

Diewert, W.E. (2003b), "Hedonic Regressions: A Review of Some Unresolved Issues", paper presented at the 7th Meeting of the Ottawa Group, Paris, May 27-29.

Diewert, W.E. (2005), "Issues in the Measurement of Capital Services, Depreciation, Asset Price Changes and Interest Rates", pp. 479-542 in *Measuring Capital in the New Economy*, C. Corrado, J. Haltiwanger and D. Sichel (eds.), Chicago: University of Chicago Press.

Diewert, W.E. (2009a), "Durables and Owner Occupied Housing in a Consumer Price Index", pp. 445-500 in *Price Index Concepts and Measurement*, W.E. Diewert, J. Greenlees and C. Hulten (eds.), Studies in Income and Wealth, Volume 70, Chicago: The University of Chicago Press.

Diewert, W.E. (2009b), "The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions", pp. 87-116 in *Price and Productivity Measurement: Volume 1—Housing*, Trafford Press and www.indexmeasures.com.

Diewert, W.E. (2010), “Alternative Approaches to Measuring House Price Inflation”, Discussion Paper 10-10, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1, December.

Diewert, W.E., Y. Finkel and Y. Artsev (2009), “Empirical Evidence on the Treatment of Seasonal Products: The Israeli Experience”, pp. 53-78 in *Price and Productivity Measurement: Volume 2; Seasonality*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press.

Diewert, W.E., J. de Haan and R. Hendriks (2010), “The Decomposition of a House Price index into Land and Structures Components: A Hedonic Regression Approach”, Discussion Paper 10-01, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1.

Diewert, W.E., J. de Haan and R. Hendriks (2011), “Hedonic Regressions and the Decomposition of a House Price index into Land and Structure Components”, Discussion Paper 11-01, Department of Economics, The University of British Columbia, Vancouver, Canada, V6T 1Z1.

Diewert, W.E., S. Heravi and M. Silver (2009), “Hedonic Imputation Versus Time Dummy Hedonic Indexes”, pp. 161-196 in *Price Index Concepts and Measurement*, W.E. Diewert, J. Greenlees and C. Hulten (eds.), NBER Studies in Income and Wealth, Chicago: University of Chicago Press.

Diewert, W.E. and D.A. Lawrence (2000), “Progress in Measuring the Price and Quantity of Capital”, pp. 273-326 in *Econometrics Volume 2: Econometrics and the cost of Capital: Essays in Honor of Dale W. Jorgenson*, L. J. Lau (ed.), Cambridge, MA: The MIT Press.

Diewert, W.E. and P. von der Lippe (2010), “Notes on Unit Value Index Bias”, *Journal of Economics and Statistics* 230(6),690-708.

Diewert, W.E. and A.O. Nakamura (2009), “Accounting for Housing in a CPI”, pp.7-32 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press and www.indexmeasures.com.

Diewert, W.E., A.O. Nakamura and L.I. Nakamura (2009), “The Housing Bubble and a New Approach to Accounting for Housing in A CPI”, *Journal of Housing Economics* 18, 156-171.

Dodge, D. (2003), “It All Starts with the Data,” Remarks by David Dodge, Governor of the Bank of Canada to the Conference of European Statisticians, Geneva. Available at www.bankofcanada.ca/en/speeches/2003.

Dreiman, M.H. and A. Pennington-Cross (2004), “Alternative Methods of Increasing the Precision of Weighted Repeat Sales House Prices Indices”, *Journal of Real Estate Finance and Economics* 28(4), 299-317.

Duffy, D. (2009), "Measuring House Price Change", Economic and Social Research Institute, Working paper 291.

ECB (2009), "Residential Property Price Statistics for the Euro Area and the European Union", in *Housing Market Challenges in Europe and the United States*, P. Arestis, P. Mooslechner and K. Wagner (eds.), Basingstoke: Palgrave Macmillan.

Edelstein, R.H. and D.C. Quan (2006), "How Does Appraisal Smoothing Bias Real Estate Returns Measurement?", *Journal of Real Estate Finance and Economics* 32(1), 41-60.

Englund, P., J.M. Quigley and C.L. Redfearn (1998), "Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices", *Journal of Urban Economics* 44(2), 171-196.

Eurostat (2001), Handbook on Price and Volume Measures in the National Accounts. Luxembourg: Office for Official Publications of the European Communities.

Eurostat (2005), "Final Report of the Joint NA-PPP Task Force on Dwelling Services", paper presented at the Meeting of the Eurostat Working Parties on National Accounts and PPP, November 23, Luxembourg.

Eurostat (2010), *Technical Manual on Owner-Occupied Housing for Harmonised Index of Consumer Prices*, Draft, February 2010.

Eurostat, International Monetary Fund, OECD, United Nations, World Bank (1993), *System of National Accounts 1993*, Luxembourg: Eurostat. Cited as SNA (1993).

Eurostat, International Monetary Fund, OECD, United Nations, World Bank (2009), *System of National Accounts 2008*, Luxembourg: Eurostat. Cited as SNA (2008).

Fenwick, D. (2005), "Statistics on Real Estate Prices: The Need for a Strategic Approach", pp. 368-372 in *Real Estate Indicators and Financial Stability*, BIS Papers No 21, Bank for International Settlements, Washington DC: The International Monetary Fund.

Fenwick, D. (2006), "Real Estate Prices: The Need for a Strategic Approach to the Development of Statistics to Meet User Needs", paper presented at the OECD-IMF Workshop on Real Estate Price Indexes held in Paris, November 6-7, 2006.

Finocchiaro, D. and V. Queijo von Heideken (2007), "Do Central Banks React to House Prices?", Working Paper Series 217, Sveriges Riksbank (Central Bank of Sweden), (revised 01 Aug 2009).

Francke, M.K. (2008), "The Hierarchical Trend Model", pp. 164-180 in *Mass Appraisal Methods: An International Perspective for Property Valuers*, T. Kauko and M. Damato (eds.), Oxford: Wiley-Blackwell.

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

Francke, M.K. and G.A. Vos (2004), "The Hierarchical Trend Model for Property Valuation and Local Price Indices", *Journal of Real Estate Finance and Economics* 28(2/3), 179-208.

Garner, T.I. and K. Short (2009), "Accounting for Owner-Occupied Dwelling Services: Aggregates and Distributions", *Journal of Housing Economics* 18, 273-279.

Garner, T.I. and R. Verbrugge (2009a), "The Puzzling Divergence of Rents and User Costs, 1980-2004: Summary and Extensions", pp. 125-146 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press and www.indexmeasures.com.

Garner, T.I. and R. Verbrugge (2009b), "Reconciling User Costs and Rental Equivalence: Evidence from the US Consumer Expenditure Survey", *Journal of Housing Economics* 18, 172-192.

Gatzlaff, D.H. and D. Ling (1994), "Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies", *Journal of Urban Economics* 35(2), 221-244.

Gatzlaff, D.H. and D.R. Haurin (1997), "Sample Selection Bias and Repeat-Sales Index Estimates", *Journal of Real Estate Finance and Economics* 14(1), 33-50.

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.

Geltner, D. (1996), "The Repeated-Measures Regression-Based Index: A Better Way to Construct Appraisal-Based Indexes of Commercial Property Value", *Real Estate Finance* 12(4), 29-35.

Geltner, D. and Pollakowski, H. (2006), "A Set of Indexes for Trading Commercial Real Estate Based on the Real Capital Analytics Transaction Prices Database," Working paper, MIT Center for Real Estate: CREDL.

Goetzmann, W.N. (1992), "The Accuracy of Real Estate Indices: Repeat Sales Estimators", *Journal of Real Estate Finance and Economics* 5(1), 5–53.

Goetzmann, W. and L. Peng (2002), "The Bias of RSR Estimator and the Accuracy of Some Alternatives", *Real Estate Economics* 30(1), 13-39.

Goetzmann, W.N. and M. Spiegel (1997), "A Spatial Model of Housing Returns and Neighborhood Substitutability", *Journal of Real Estate Finance and Economics* 14(1–2), 11–31.

Goodhart, C. (2001), "What Weight Should Be Given to Asset Prices in Measurement of Inflation?", *Economic Journal* 111, 335–356.

Goodhart, C. and B. Hofmann (2007), "Financial Conditions Indices," in *House Prices and the Macroeconomy: Implications for Banking and Price Stability*, Charles Goodhart (ed.), Oxford: Oxford University Press.

Goodman, A.C. and T.G. Thibodeau (2003), "Housing Market Segmentation and Hedonic Prediction Accuracy", *Journal of Housing Economics* 12(3), 181–201.

Gouriéroux, C. and A. Laferrère (2009), "Managing Hedonic House Price Indexes: The French Experience", *Journal of Housing Economics* 18, 206-213.

Griliches Z. (1961, 1971a), "Hedonic Price Indexes for Automobiles: An Econometric Analysis for Quality Change", in Z. Griliches (eds.), *Price Indexes and Quality Change: Studies in New Methods of Measurement*, Cambridge MA: Harvard University Press.

Griliches, Z. (1971b), "Introduction: Hedonic Price Indexes Revisited", pp. 3-15 in *Price Indexes and Quality Change*, Z. Griliches (ed.), Cambridge MA: Harvard University Press.

Gudnason, R. (2004), "Simple User Costs and Rentals", Paper presented at the 8th Ottawa Group Meeting, Helsinki, August 23-25.

Gudnason, R. and G. Jónsdóttir (2006), "House Price Index, Market Prices and Flow of Services Methods", paper presented at the OECD-IMF Workshop on Real Estate Price Indexes held in Paris, November 6-7, 2006.

Gudnason, R. and G.R. Jónsdóttir (2009), "Owner Occupied Housing in the Icelandic CPI", pp. 85-88 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press and www.indexmeasures.com.

Gyourko, J. and A. Saiz (2004), "Reinvestment in the Housing Stock: The Role of Construction Costs and the Supply Side", *Journal of Urban Economics* 55, 238-256.

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

de Haan, J. (2009), "Comment on Hedonic Imputation versus Time Dummy Hedonic Indexes", pp. 196-200 in *Price Index Concepts and Measurement*, W.E. Diewert, J.S. Greenlees and C.R. Hulten (eds.), Studies in Income and Wealth 70, Chicago: University of Chicago Press.

de Haan, J. (2010a), "Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Other Approaches", *Journal of Economics and Statistics* 230(6), 772-791.

de Haan, J. (2010b), "An Appraisal-Based Generalized Regression Estimator of House Price Change", paper presented at the Economic Measurement Group Workshop, 1-3 December 2010, Sydney, Australia.

de Haan, J. and H. van der Grient (2011), "Eliminating Chain Drift in Price Indexes Based on Scanner Data", *Journal of Econometrics* 161, 36-46.

de Haan, J., E. van der Wal and P. de Vries (2009), "The Measurement of House Prices: A Review of the Sales Price Appraisal Method", *Journal of Economic and Social Measurement* 34(2-3), 51-86.

Halvorsen, R. and H.O. Pollakowski (1981), "Choice of Functional Form for Hedonic Price Equations", *Journal of Urban Economics* 10, 37-49.

Haurin, D.R. and P.H. Hendershott (1991), "House Price Indexes: Issues and Results", *American Real Estate and Urban Economics Association Journal* 19, 259-269.

Heston, A. and A.O. Nakamura (2009a), "Reported Prices and Rents of Housing: Reflections of Costs, Amenities or Both?", pp. 117-124 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press and www.indexmeasures.com.

Heston, A. and A.O. Nakamura (2009b), "Questions about the Equivalence of Market Rents and User Costs for Owner Occupied Housing", *Journal of Housing Economics* 18, 273-279.

Hicks, J.R. (1946), *Value and Capital*, Second Edition, Oxford: Clarendon Press.

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), "Superlative Index Numbers: Not All of Them Are Super", *Journal of Econometrics* 130(1), 25-43.

Hill, R.J (2011), "Hedonic Price Indexes for Housing", OECD Statistics Working paper 2011/1, OECD Publishing, Paris.

Hill, R.J. and D. Melser (2008), "Hedonic Imputation the Price Index Problem: An Application to Housing", *Economic Inquiry* 46(4), 593-609.

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, 193-205.

Hill, R.J., D. Melser and B. Reid (2010), "Hedonic Imputation with Geospatial Data: An Application of Splines to the Housing Market", Mimeo.

Hill, T.P. (1988), "Recent Developments in Index Number Theory and Practice", *OECD Economic Studies* 10, 123-148.

Hill, T.P. (1993), "Price and Volume Measures", pp. 379-406 in SNA (1993).

Hoffmann, J. and C. Kurz (2002), "Rent Indices for Housing in West Germany: 1985 to 1998", Discussion Paper 01/02, Economic Research Centre of the Deutsche Bundesbank, Frankfurt.

Hoffmann, J. and J.R. Kurz-Kim (2006), "Consumer Price Adjustment Under the Microscope: Germany in a Period of Low Inflation", Discussion Paper Series 1, Economic Studies, No 16/2006, Deutsche Bundesbank, Frankfurt am Main.

Hoffmann, J. and A. Lorenz (2006), "Real Estate Price Indices for Germany: Past, Present and Future", paper presented at the OECD-IMF Workshop on Real Estate Price Indexes, Paris, November 6-7.

Hwang, M. and J.M. Quigley (2004), "Selectivity, Quality Adjustment and Mean Reversion in the Measurement of House Values", *Journal of Real Estate and Finance and Economics* 28(2/3), 161-178.

ILO, IMF, OECD, Eurostat, United Nations, World Bank (2004), *Consumer Price Index Manual: Theory and Practice*, Geneva: ILO. Cited as CPI Manual (2004).

ILO, IMF, OECD, Eurostat, United Nations, World Bank (2004), *Producer Price Index Manual: Theory and Practice*, Washington DC: IMF. Cited as PPI Manual (2004).

IMF (2006), *Compilation Guide on Financial Soundness Indicators*, Washington DC: IMF.

IMF (2007), "Review of Exchange Arrangements, Restrictions, and Controls", Available at http://www.imf.org/external/np/pp/2007/eng.

IMF, ILO, OECD, UNECE and World Bank (2009), *Export and Import Price Index Manual*, Washington, DC: IMF. Cited as XMPI Manual (2009).

Ivancic, L., W.E. Diewert and K.J. Fox (2011), "Scanner Data, Time Aggregation and the Construction of Price Indexes", *Journal of Econometrics* 161, 24-35.

Jansen, S.J.T., P. de Vries, H.C.C.H. Coolen, C.J.M. Lamain and P. Boelhouwer (2008), "Developing a House Price Index for the Netherlands", *Journal of Real Estate and Finance and Economics* 37(2), 163-186.

Jarociński, M. and F.R. Smets (2008), "House Prices and the Stance of Monetary Policy", *Federal Reserve Bank of St. Louis Review* 90(4), 339-365.

Katz, A.J. (2009), "Estimating Dwelling Services in the Candidate Countries: Theoretical and Practical Considerations in Developing Methodologies Based on a User Cost of Capital Measure", pp. 33-50 in *Price and Productivity Measurement, Volume 1: Housing*, W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds.), Trafford Press and www.indexmeasures.com.

Knight, J.R., J. Dombrow and C.F. Sirmans (1995), "A Varying Parameters Approach to Constructing House Price Indexes", *Real Estate Economics* 23 (2), 187-205.

Koev, E. and J.M.C. Santos Silva (2008), "Hedonic Methods for Decomposing House Price Indices into Land and Structure Components", unpublished paper, Department of Economics, University of Essex, England, October.

Koop (2009)…

Lancaster, K.J. (1996), "A New Approach to Consumer Theory", *Journal of Political Economy* 74, 132-157.

Lebow, D.E. and J.B. Rudd (2003), "Measurement Error in the Consumer Price Index: Where do we Stand?", *Journal of Economic Literature* 41, 159-201.

Leishman, C. (2000), "Estimating Local Housing Market Price Indices Using Land Registry Data", *Housing Finance* 47(August), 55-60.

Leishman, C., C. Watkins and W.D. Fraser (2002), “The Estimation of House Price Indices Based on Repeat Sales Regression Using Land Registry Data”, Report to RICS Education Trust, London.

Leventis, A. (2006), “Removing Appraisal Bias from a Repeat Transactions House Price Index: A Basic Approach”, paper presented at the OECD-IMF Workshop on Real Estate Price Indexes held in Paris, November 6-7, 2006.

Li, W., M. Prud’homme and K. Yu (2006), “Studies in Hedonic Resale Housing Price Indexes”, paper presented at the OECD-IMF Workshop on Real Estate Price Indexes held in Paris, November 6-7, 2006.

Lim, S. and M. Pavlou (2007), “An Improved National House Price Index Using Land Registry Data”, RICS research paper series, vol. 7, no. 11.

Luüs, C. (2005), “The Absa Residential Property Market Database for South Africa: Key Data Trends and Implications”, pp. 149-170 in *Real Estate Indicators and Financial Stability*, BIS Papers No 21, Bank for International Settlements, Washington DC: IMF.

Makaronidis, A. and K. Hayes (2006), “Owner Occupied Housing for the HICP”, paper presented at the OECD-IMF Workshop on Real Estate Price Indexes, Paris, November 6-7, 2006.

Malpezzi, S. (2003), “Hedonic Pricing Models: A Selective and Applied Review”, in *Housing Economics and Public Policy*, T.O’Sullivan and K. Gibb (eds.), Malder, MA: Blackwell.

Malpezzi, S., L. Ozanne and T. Thibodeau (1987), “Microeconomic Estimates of Housing Depreciation”, *Land Economics* 63, 372-385.

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.

McDonald, J.F. (1981), “Capital-Land Substitution in Urban Housing: A Survey of Empirical Estimates”, *Journal of Urban Economics* 9, 190-211.

McDonald, C, and M. Smith, (2009), “Developing stratified housing price measures for New Zealand”, Discussion paper no. 2009/07, Reserve Bank of New Zealand, Wellington.

McMillen, D.P. (2003), “The Return of Centralization to Chicago: Using Repeat Sales to Identify Changes in House Price Distance Gradients”, *Regional Science and Urban Economics* 33, 287-304.

Meese, R. and N. 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, 308-332.

Mudgett, B.D. (1955), “The Measurement of Seasonal Movements in Price and Quantity Indexes”, *Journal of the American Statistical Association* 50, 93-98.

Muellbauer, J. (1974), "Household Production Theory, Quality and the 'Hedonic Technique'", *American Economic Review* 64, 977-994.

Muth, R.F. (1971), "The Derived Demand for Urban Residential Land", *Urban Studies* 8, 243-254.

Nabarro, R. and T. Key (2003), "Performance measurement and Real Estate Lending Risk", paper presented at the IMF/BIS Conference on Real Estate Indicators and Financial Stability, October 2003.

Ohnishi, T., T. Mizuno, C. Shimizu and T. Watanabe (2010),"On the Evolution of the House Price Distribution," Discussion paper no.61, Research Center for Price Dynamics.

Pakes (2003), "A Reconsideration of Hedonic Price Indexes with an Application to PCs", *American Economic Review* 93(5), 1576-93.

Palmquist, R.B. (1980), "Alternative Techniques for Developing Real Estate Price Indexes", *Review of Economics and Statistics* 62, 442-448.

Palmquist, R.B. (1982), "Measuring Environmental Effects on Property Values without Hedonic Regressions", *Journal of Urban Economics* 11, 333-347.

Pennington-Cross, A. (2005), "Aggregation Bias and the Repeat Sales Price Index", pp. 323-335 in *Real Estate Indicators and Financial Stability*, BIS Papers No 21, Bank for International Settlements, Washington DC: IMF.

Phang, S.Y. (2004), "House Prices and Aggregate Consumption: Do They Move Together? Evidence from Singapore", *Journal of Housing Economics* 13, 101–119.

Plosser, C.I. (2007), "House Prices and Monetary Policy", Speech for the European Economics and Financial Centre Distinguished Speakers Series, July 11, 2007, Monograph, Federal Reserve Bank of Philadelphia.

Prasad, N. and A. Richards (2006), "Measuring House Price Growth – Using Stratification to Improve Median-based Measures", Research discussion paper no. 2006-04, Reserve Bank of Australia.

Prasad, N. L. and A. Richards (2008), "Improving Median Housing Price Indexes Through Stratification," *Journal of Real Estate Research* 30(1), 45-71.

Quigley, J.M. (1995), "A Simple Hybrid Model for Estimating Real Estate Price Indexes", *Journal of Housing Economics* 4(1), 1–12.

Reinhart, C.M. and K.S. Rogoff (2009), *This Time is Different; Eight Centuries of Financial Folly*, Princeton and Oxford: Princeton University Press.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", *Journal of Political Economy* 82, 34-55.

Rossini, P. and P. Kershaw (2006), "Developing a Weekly Residential Price Index Using the Sales Price Appraisal Ratio, paper presented at the Twelfth Annual Pacific Rim Real Estate Society Conference, Auckland, 22-25 January.

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.

Samuelson, Paul A. and Swamy, S. (1974), “Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis”, *American Economic Review* 64, 566-93.

Scheffé, H. (1959), *The Analysis of Variance*, New York: John Wiley and Sons.

Shi, S., M. Young and B. Hargreaves (2009), “Issues in Measuring a Monthly House Price Index in New Zealand”, *Journal of Housing Economics* 18(4), 336-350.

Shiller. R.J. (1991), “Arithmetic Repeat Sales Price Estimators”, *Journal of Housing Economics* 1(1), 110–126.

Shiller, R.J. (1993a), “Measuring Asset Values for Cash Settlement in Derivative Markets: Hedonic Repeated Measures Indices and Perpetual Futures”, *Journal of Finance* 48(3), 911–931.

Shiller, R.J. (1993b), *Macro Markets*, Oxford: Oxford University Press.

Shiller, R.J. (2005), “Comments on Session ‘Aggregation Issues’”, pp. 336-339 in *Real Estate Indicators and Financial Stability*, BIS Papers No 21, Bank for International Settlements, Washington DC: The International Monetary Fund.

Shimizu, C. and K.G. Nishimura (2006), “Biases in Appraisal Land Price Information: The Case of Japan”, *Journal of Property Investment & Finance* 24 (2), 150-175.

Shimizu, C. and K.G. Nishimura (2007), “Pricing Structure in Tokyo Metropolitan Land Markets and Its Structural Changes: Pre-bubble, Bubble, and Post-bubble Periods,” *Journal of Real Estate Finance and Economics* 35(4), 495-496.

Shimizu, C., K.G. Nishimura and Y. Asami (2004), “Search and Vacancy Costs in the Tokyo Housing Market: Attempt to Measure Social Costs of Imperfect Information”, *Regional and Urban Development Studies* 16(3), 210-230.

Shimizu, C., K.G. Nishimura and T. Watanabe (2010), “Housing Prices in Tokyo: A Comparison of Hedonic and Repeat Sales Measures”, *Journal of Economics and Statistics* 230/6, 792-813.

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

Shimizu, C., K.G. Nishimura and T. Watanabe (2011), “House Prices from Realtors, Magazines, and Government: A Statistical Comparison”, Mimeo.

Silver, M. (2009a), “Do Unit Value Export, Import, and Terms of Trade Indices Represent or Misrepresent Price Indices?”, IMF Staff Papers 56, 297-322, IMF, Washington DC.

Silver, M. (2009b), “Unit Value Indices”, Chapter 2 in XMPI Manual (2009).

Silver, M. (2010), “The Wrongs and Rights of Unit Value Indices”, *Review of Income and Wealth*, Series 56, Special Issue 1, S206-S223.

Standard and Poor's (2009), *S&P/Case-Shiller Home Price Indices; Index Methodology*, New York: Standard and Poor's.

Statistics Portugal (Instituto Nacional de Estatistica) (2009), "Owner-Occupied Housing: Econometric Study and Model to Estimate Land Prices, Final Report", paper presented to the Eurostat Working Group on the Harmonization of Consumer Price Indices", March 26-27, Luxembourg: Eurostat.

Steele, M. and R. Goy (1997), "Short Holds, the Distributions 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. (1956), *Quantity and Price Indexes in National Accounts*, Paris: OECD.

Thibodeau, T.G. (1992), "Residential Real Estate Prices: 1974-1983", The Blackstone Company, Mount Pleasant (MI).

Thibodeau, T.G. (2003), "Marking Single-Family Property Values to Market", *Real Estate Economics* 31(1), 1–22.

Thorsnes, P. (1997), "Consistent Estimates of the Elasticity of Substitution between Land and Non-Land Inputs in the Production of Housing", *Journal of Urban Economics* 42, 98-108.

Triplett, J.E. (2006), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes; Special Application to Information and Technology Products*, Directorate for Science, Technology and Industry, Paris: OECD.

Triplett, J.E. and R.J. McDonald (1977), "Assessing the Quality Error in Output Measures: The Case of Refrigerators", *The Review of Income and Wealth* 23(2), 137-156.

Tse, R.Y.C. (2002), "Estimating Neighbourhood Effects in House Prices: Towards a New Hedonic Model Approach", Urban Studies 39(7), 1165–1180.

Turvey, R., (1989), *Consumer Price Indices: An ILO Manual*, Geneva: International Labour Office.

UK Department of the Environment (1982), "A New Index of Average House Prices", *Economic Trends* 348, 134–138.

United Nations (2009), *Practical Guide to Producing Consumer Price Indices*, New York and Geneva: United Nations.

van der Wal, E., D. ter Steege and B. Kroese (2006), "Two Ways to Construct a House Price Index for the Netherlands: The Repeat Sale and the Sale Price Appraisal Ratio", paper presented at the OECD-IMF Workshop on Real Estate Price Indexes, Paris, November 6-7, 2006.

Verbrugge, R. (2008), "The Puzzling Divergence of Aggregate Rents and User Costs, 1980-2004", *The Review of Income and Wealth* 54, 671-699.

de Vries, P., J. de Haan, E. van der Wal and G. Mariën (2009), "A House Price Index Based on the SPAR Method", *Journal of Housing Economics* 18(3), 214-223.

Wang, T. and P.M. Zorn (1997), "Estimating House Price Growth with Repeat Sales Data: What's the Aim of the Game?", *Journal of Housing Economics* 6(2), 93-118.

Wallace, N.E. and R.A. Meese (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(1–2), 51–73.

Wenzlick, R. (1952), "As I See the Fluctuations in the Selling Prices of Single Family Residences", *The Real Estate Analyst* 21 (December 24), 541-548.

Wolverton, M.L. and J. Senteza (2000), "Hedonic Estimates of Regional Constant Quality House Prices", *Journal of Real Estate Research* 19(3), 235-253.

Wood, R. (2005), "A Comparison of UK Residential House Price Indices", pp. 212-227 in *Real Estate Indicators and Financial Stability*, BIS Papers No 21, Bank for International Settlements, Washington DC: The International Monetary Fund.

Wyngarden, H. (1927), *An Index of Local Real Estate Prices*, Michigan Business Studies Volume 1, Number 2, Ann Arbor: University of Michigan.

Zhu, H. (2005), "The Importance of Property Markets for Monetary Policy and Financial Stability", pp. 9-29 in *Real Estate Indicators and Financial Stability*, Volume 21, Bank for International Settlements (ed.).