

Official house price statistics explained

Joni Karanka¹, Robert O'Neill², Natalie Weaden², Ria Sanderson², Christopher Jenkins¹ & Derek Bird¹

Summary

The housing market has a large impact on the economy and on society as a whole. However, measuring the overall change in property prices is inherently difficult. This article reviews the need for accurate house price statistics, the different methods used to calculate a house price index, and the indices and official statistics available in the UK. It also explains the calculation of the ONS House Price Index and its relation to other official house price statistics. The article describes how differences in data sources, scope and methods lead to differences in reported house price levels, yet house price movements are similar across official sources.

Chapter 1 introduces the UK housing market, and the difficulties faced when producing house price statistics. Chapter 2 provides an overview of the different methods used to estimate house prices and where these are used. Chapter 3 provides a detailed explanation of the ONS House Price Index. Chapter 4 compares the different official price statistics, discusses their methods and compares their reported house price levels and house price movements. Chapter 5 draws conclusions and suggests future directions for the ONS HPI.

1. Introduction

1.1 The housing market

The aim of a house price index is to provide information about house price inflation over time. They are an important set of statistics for understanding the housing market. In the UK the housing market is a large component of the economy, comprising 27.4 million properties (source: DCLG). According to the 2011 Census in England and Wales 63.5% of households own their property either outright or with an outstanding mortgage or loan (a UK estimate will be published later in 2013). This rate of home ownership is slightly below the EU average of 70.8% (Eurostat, 2010). The rest of households live in accommodation that is either rented from a private landlord or from a social landlord such as a local authority or housing association (Figure 1.1).

¹ Prices Division, Office for National Statistics

² Survey Methodology Division, Office for National Statistics

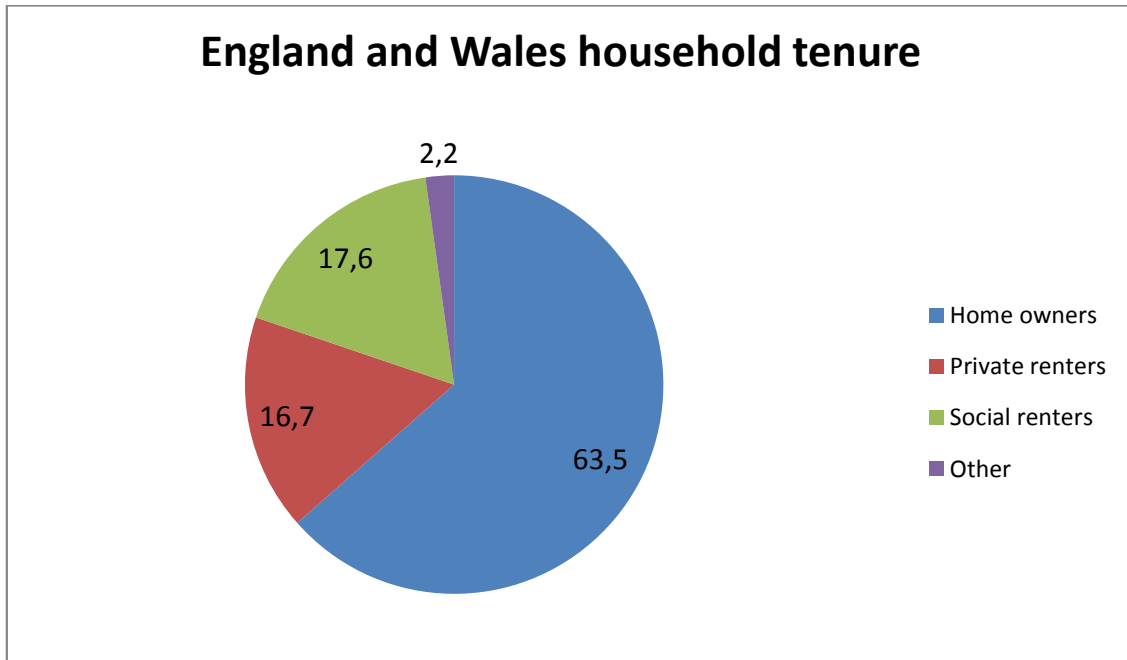


Figure 1.1. Household tenure for England and Wales. Source: Census 2011.

For those who choose to buy, a property is most often the largest investment made in a lifetime. Few buyers are able to buy property outright so most rely on large loans in the form of mortgages. The size of this investment is reflected in the fact that, in 2011, the average mortgage was 3 times the buyers' yearly joint income (see [table 30 of the House Price Index](#)). Demand for mortgages has built up a large market of lenders. As of the third quarter of 2012, the total value of outstanding mortgages is estimated at £1262 billion by the [ONS UK Economic Accounts](#). To put this number into perspective, the national debt of the UK at the end of 2012 amounted to £1111 billion (see [ONS UK Public Sector Accounts](#)).

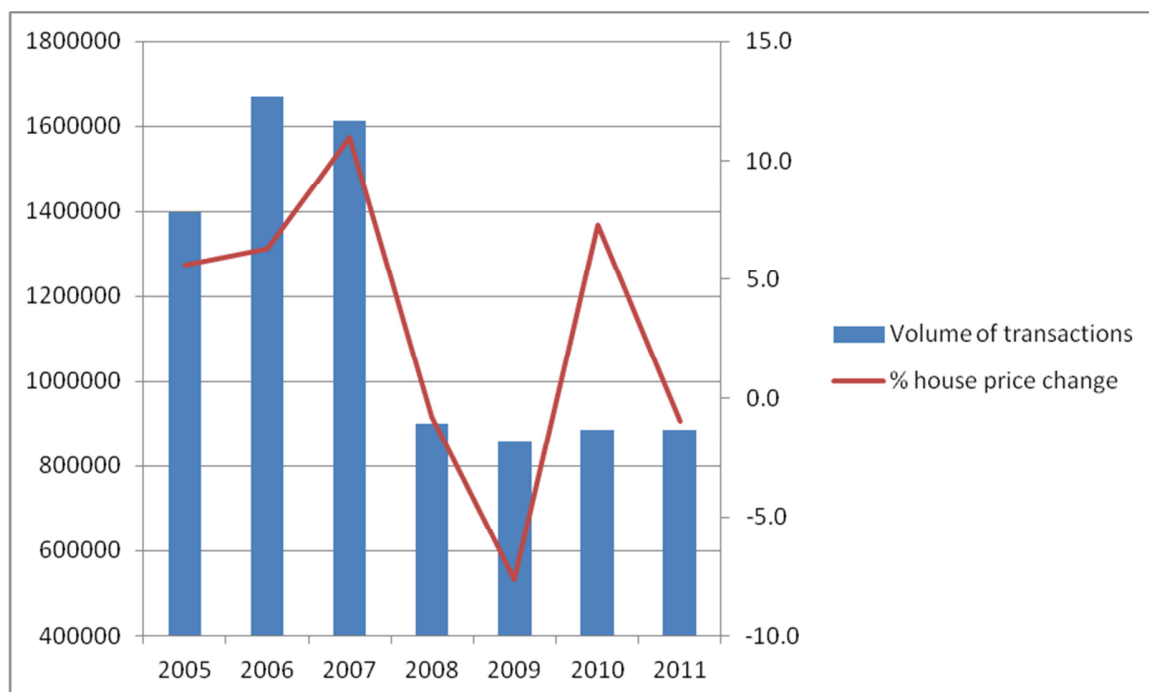


Figure 1.2. UK yearly volume of property transactions (source: HMRC) and percentage change in house prices (source: ONS).

As the housing market is linked to income, wealth and availability of lending, it is also very sensitive to the overall economic climate. For example, the 2008 credit crisis had significant repercussions for the housing market (figure 1.2). Large annual house price inflation of between 5.6% and 17.0% seen on every year between 1997 and 2007 reversed to a decline of 7.6% in 2009. For the last two years house price inflation has been relatively stable (source: ONS). The number of property sales in the UK has almost halved from the previous 'boom' years, from a peak of 1.67 million in 2006 to 0.86 million in 2009 (source: HMRC).

The importance of housing is reflected in a considerable number of government policies. Some examples of policies developed since 2009 include:

- **Stamp duty holiday.** Stamp Duty Land Tax (SDLT) is a tax payable on the purchase or transfer of land or property. The holiday was aimed at encouraging the first time buyer market by not charging SDLT for first time buyers when they purchased properties under £250,000.
- **Funding for Lending Scheme.** Allows lenders to borrow from the Bank of England at below standard rates as long as this money is lent to businesses and individuals. Its aim is to encourage lenders, including mortgage lenders, to lend more and at lower rates.
- **FirstBuy scheme for new properties (England only).** The Homes & Communities Agency and the house builder provide equity loan funding of up to 20% of the property price, with the buyer raising the remaining 80% with a mortgage and a deposit. It is aimed at first time buyers, providing them access to a mortgage when the deposit alone would not make one available.
- **Support for Mortgage Interest.** The government pays lenders the interest of up to £200,000 of a mortgage to people on certain income-related benefits. The main aim of this scheme is to reduce

repossessions.

1.2 House price statistics

A house price index (HPI) is a series that tracks the changes in the price of property relative to the price it had at a reference period in time. Changes in the series represent increases and decreases in house prices. The ONS HPI provides an estimate of the price of current property relative to its price in February 2002. For example the August 2011 UK HPI index was 176, which means that house prices were 76% higher than in February 2002. Usually house price indices are accompanied by other important information such as average property prices, and breakdowns by region or type of property.

House price indices seek to estimate the change in average property prices, even though there is no such thing as an 'average' house. For one to exist it would have to be part detached, part semi-detached, part terraced, it would be part house, part flat etc. Consequently, as with consumer price index 'baskets', the construct underpinning the HPI is there for ease of use and as a general guide. The more detailed a HPI is the closer to 'actual' houses it can get. For example, were a HPI able to produce estimates at the local area level for accommodation of different types, with varying numbers of bedrooms etc it would have a closer approximation to what was actually available. Producing estimates at such detailed levels is feasible with access to administrative information, remembering that housing transactions of these types are almost always captured by registration and/or planning requirements, but other issues remain. These relate both to methodological considerations, which are discussed below, and more prosaic issues, such as when one develops an estimate of price.

Capturing a house price sounds straightforward, but prices can change as the process of the sale unfolds. At the outset the house is offered at an advertised or 'asking' price. This is the first point at which one could begin to construct an index. However, asking prices often do not reflect the price for which a property actually sells and building an index on advertised prices runs the risk of including properties that do not sell.

Given the nature of the housing market, which is underpinned to a substantial degree by the mortgage market, a second approach would be to take the value of a property based on its valuation for mortgage purposes. In the UK a mortgage approval will not be made without a valuation of a property being produced. However, this too may not reflect the price a property is actually transacted for and so approximates a price in the real sense. The final point in the process is when vendor and buyer agree a price and it is this which is measured in the ONS HPI.

Before moving on to consider the uses of the HPI there is one final issue that we need to consider at this point. That relates to the frequency with which properties are transacted. Given the approach adopted for the ONS HPI, we can only measure house prices for those properties that sell in a given period. Clearly, these will reflect only a small proportion of the housing stock

and so the approach creates potential for bias if the value of houses transacted is different to the stock of those not sold. There is some reason to expect that this bias will be real, in that sales of houses will happen more frequently in economically dynamic regions of an economy, for house types most in demand etc. These types of bias are difficult to assess, but it is important to bear them in mind when producing and using the indices.

The production of house price statistics is relevant for a large number of purposes. The importance of the indices is reflected in the development of a European regulation³, introduced in February 2013, which is the first stage in the introduction of owner occupiers' housing costs into the Harmonised Index of Consumer Prices. The regulation identifies house price indices as "important indicators in their own right" reflecting the consequences for an economy of unstable or unsustainable house price development. In the national context users include central Government which uses them as an economic indicator, mortgage lenders who use them to set their interest rates, and individual households which make decisions on buying and selling property. An analysis of the main users of these statistics can be found in the [National Statistician's Review of Official Housing Market Statistics](#).

Although the availability of accurate and timely house price statistics is desirable, there are a number of obstacles to their production. These are mainly due to the nature of housing compared with other goods:

- Properties are unique. Most price indices can collect the price of the same items over several periods. For example, the price of a pint of branded milk can be collected every month. However, this is not the case with properties. Each house is different, be it due to its size and features, or by its location. Furthermore, properties are traded less frequently than other items, making successive pricings very unlikely.
- Sales volumes, prices, and the composition of the sold stock vary between periods. House price indices should reflect the change in price of comparable properties. This is not straightforward when the types of properties and the volumes sold change between periods.
- Multiple methods are available for the estimation of average house prices. Average house prices can be estimated with a large number of different sources and methods. Each choice will produce slightly different results to the others. Four different methods are discussed in the next chapter of this article.
- Some house price statistics might not be relevant to the individual user. As noted above published average house prices represent the price of an 'average property', which might not reflect the exact features of a property that a prospective seller owns. Also, house price indices tend to be constructed to calculate the price of traded property. Therefore, they cannot be used to estimate the value of untraded property (such as the

³ Commission Regulation (EU) No 93/2013

entire housing stock of the UK) unless the indices are constructed specifically for this purpose.

The ONS House Price Index (HPI) was developed to produce a consistent series of traded house prices. Until its transfer to ONS in April 2012, the index was produced and published by the Department for Communities and Local Government (DCLG). One of its main features is the mix-adjustment of the monthly transactions to remove the effect of changing composition of sold stock. Since its conception the index relies on the use of mortgage transaction data, which provides good coverage of the housing market but excludes those properties that are cash sales. The mix-adjusted HPI series start in 1969, when the index was calculated from a 5% sample of the mortgage transactions of a number of building societies. From 1993 the coverage of transactions increased from building societies to cover all mortgage lenders. In 2002 the index started using hedonic regression to perform the mix-adjustment (hedonic regressions are explained in the next chapter). In 2003 the sample submitted by each mortgage lender was expanded from 5% to cover all their recorded mortgage transactions. These mortgage transactions are collected from the Regulated Mortgage Survey of the Council of Mortgage Lenders (CML).

In December 2010 the National Statistician's Review of House Price Statistics made a series of recommendations. These were mainly driven by the availability of numerous official house price statistics, which vary in their sample, methods and focus. The main recommendation was for the development of a definitive official statistic. Additionally, there is a need for clearer explanation of the methods used in the current statistics and how these relate to each other. The following chapters of this article deal with explaining the methodology used in house price indices.

2. Overview of methods

This chapter gives an overview of the different methods that are used to construct house price indices. First, it briefly discusses why specialised methods are needed for house prices that are not used in all other price indices. It then continues with a description of the mix-adjustment with stratification matrix, hedonic regression, repeated-sales measure and SPAR methods. The chapter finishes by summarising which methods are used as official statistics in different countries.

2.1 Peculiarities of house prices

There are some issues that make calculating a price index for properties very different from other price indices. Firstly, in most price indices, the same products are priced in each period. However, house price indices depend on prices collected from sales or valuations, and it is very unlikely that the same property is traded or valued in consecutive periods. Secondly, unlike many other products no two properties are identical. Two houses on the same street might differ in size, number of rooms or state of repair, all of which might impact the sale price agreed

between seller and buyer. On top of these two problems, it is unlikely that the properties with the same features are traded in the same locations in each period, resulting in changes in the composition of sales from period to period. As both the features of the property and its location can impact prices, using average sale prices or valuation estimates would result in a very unstable price index from one period to the next.

The following methods have been identified as solving these three issues. All of them produce a consistent price index in consecutive periods and adjust for the change in the characteristics of the properties being sold in each period. More detailed descriptions of the methods can be found in the [Eurostat manual for Residential Property Price Indices](#).

2.2 Mix-adjustment with a stratification matrix

A solution put forwards, amongst others, by the Australian Bureau of Statistics and previously used in the DCLG House Price Index (currently the ONS HPI) until 2002 is the mix adjustment with a stratification matrix. Stratifying involves dividing the sample of properties into subgroups called *strata*. These groups are expected to have similar numbers of transactions and they must be based on characteristics of the sample that are important when estimating property prices. Collected together, the stratified sample will form a matrix. An example of a matrix of house prices is shown in table 2.1.

	Region 1	Region 2	Region 3
1 bedroom	£85,000	£150,000	£95,000
2 bedrooms	£110,000	£205,000	£145,000
3 bedrooms	£135,000	£245,000	£190,000

Table 2.1. Average property prices three property sizes for three fictional regions.

A matrix might consist of a set of rows that represent different property sizes and a set of columns that represent different regions. In the example of table 2.1, there are three regions, and three property sizes defined by their number of bedrooms. The regions and property sizes are the strata mentioned above. The combination of each property size and region forms a cell for which an average price is collected. For example, in the first region the average price of a one bedroom property is £85,000, while in the second region it is £150,000. By creating averages for the same cells in each period we can remove the effect of many properties being sold in a certain region, or of a certain size, in a particular period. The averages for the same cells are calculated and the average prices for each cell are used to create sub-indices. These sub-indices can then be aggregated into an overall index using weighted averages. This method controls for changes in the composition of the sample like fluctuations in the number of sales by region.

The main advantages of this method are:

- Simple to set up if the variables that form the strata are available from the data sources.

- Easy to understand and assess.

The main disadvantages of this method are:

- The number of cells is limited. If too many cells are created the estimates will be less reliable as there will be fewer observations in each cell.
- This method requires enough data to fill each cell; therefore the cells need to be relatively similar in sample size.
- Good judgement is needed when choosing the variables that form the strata, as different choices will lead to different estimates of property price.

2.3 Hedonic regression

A more sophisticated form of mix adjustment can be done using standard statistical techniques, and in particular hedonic regression models. In the case of a hedonic regression the properties are defined in terms of a set of features or characteristics, each of which contributes to the value of the property. For example, the presence of a garden, the number of bedrooms, or the location of the property will all contribute to its value, but none of these features can be priced in isolation. A regression model is used to estimate the value of each of these features from the set of properties sold during a particular period. For example, the model might estimate the effect that every additional bedroom and each different location have in the sale price in a certain month. Then, the price of a particular property can be calculated by combining the values assigned to each of its features. This method allows us to estimate the prices of properties with every combination of features (such as number of bedrooms and regions), even if that particular combination did not trade in the period. A number of variants of this method are available, and they all control for changes in the composition of the sample. Some models use an 'average house' model, which has the most common features of a traded property and calculate its estimated price (e.g., a three bedroom semi-detached house in a medium sized town), while others estimate the prices of a large array of property features and aggregate these to produce the different average house prices (resembling the stratification method). There are many possible hedonic models, as each of them can have a different choice of features used to estimate prices, a different set of weights to aggregate the prices, or slightly different formulae used to estimate the contributions of the features.

Hedonic regression is widely used for estimating house prices. The Spanish statistics institute (INE) publishes [a quarterly house price index](#) based on notary records. In Spain all property transactions are recorded by a notary, and therefore the source has full coverage. A number of variables associated with each property are included, such as its size and location, and whether it is a new or old property. These are then used in a hedonic regression to estimate house prices in the different regions of the country. A similar approach is used by the German statistics

institute (DESTATIS), who use their Expert Committees for Property Valuation (an institution regulated by federal law) to access data on property transactions across federal Germany. As with the case of Spain, the data provides the statistics institute with a large number of relevant variables. DESTATIS [publishes separate indices for new and existing dwellings, prefabricated dwellings, non residential buildings and land](#).

The main advantages of this method are:

- Efficient use of data. All the data contributes to estimating property prices.
- Controls for property mix changes in the sample and, if sufficient property features are included, controls for quality changes.
- Price estimates can be produced for cells that have no recorded transactions in a given period.

The main disadvantages of this method are:

- Requires a large number of property features to be available from the data source.
- Hedonic regression models can be implemented in many ways. The choice of sources, property features that are used in the regression, the form of the regression and the way that data are aggregated will all lead to slightly different property price estimates.

2.4 Repeat sales method

Initially proposed by Bailey, Muth and Nourse (1963), this method makes use of the change in the price of a property between its latest and its previous sale. For example, if a particular property was sold in 2002 and then in 2012, the increase or decrease in price would be used to calculate the index. One of the main features of the repeat sales method is that the index can be constructed without needing as much information about individual properties as the stratification and hedonic regression methods. For example, this method will work even if the size of the property is not known. As far as the exact property can be accurately matched between the previous and current sale, the index can be calculated.

This method has been widely used in the United States, for example by both the Federal Housing Finance Agency (FHFA) and Standard and Poor's Case-Shiller Home Price Indexes. In the case of the FHFA the source of the data are the government-sponsored enterprises (GSE) Fannie Mae and Freddie Mac. These GSEs purchase the vast majority of mortgages from the mortgage lenders in the US and sell them in the financial markets as mortgage-backed securities⁴. This

⁴ More information on mortgage-backed securities and government-sponsored enterprises can be found at the US Securities and Exchange Commission: <http://www.sec.gov/answers/mortgagesecurities.htm>

ensures that mortgage lenders receive money to provide further lending for properties. By sourcing data from GSEs, the FHFA has a large coverage of US property data. The transaction data goes back as far as 1975, but the first repeat sales HPI was calculated in 1988 once several years of data had been acquired. Since 2011 data from the county recorder offices and FHFA itself have been added to this index.

The main advantages of this method are:

- The data source does not need many features of the property, mainly a reliable way to identify the properties being transacted.
- The method does not rely on characteristics between properties matching, but uses the exact property match to identify the change.
- It automatically controls for location and traded property mix.
- There is an extensive literature for this method. This includes established frameworks to test house price growth and its standard errors.

The main disadvantages of this method are:

- The method can be data inefficient as a proportion of the sample is not used for the calculation of the index. Of the traded properties in a particular period, only some have been traded previously. However, the efficiency of the method will increase over time as more sales are recorded in the database.
- Changes in the quality of the dwelling (refurbishments, depreciation due to age, and extensions and conversions) are not automatically taken into account in the model. Properties that are traded after a long period can have drastic changes in quality which impact the estimated prices.
- Estimates might not be available for very fine property categories if sales have for these have not been recorded in a period.
- It is difficult to weight the method so that it takes account of different types of stock. By default it will weight more heavily those properties traded most often which can lead to biases.
- New properties are not taken into account as they haven't been previously traded.

2.5 Sale Price Appraisal Ratio

In some countries properties are appraised regularly. Appraisals are typically carried out for the purpose of estimating taxes linked to the property, and therefore tend to cover all the properties in the country. The Sale Price Appraisal Ratio (SPAR) method makes use of these

appraisal values in order to produce a house price index. In general terms, SPAR methods use the appraisal of all the properties as a reference period. After this period, as properties are sold they are matched with their appraised value in order to create an estimate of the change in value. The advantage of SPAR over a repeat sales method is that the change in price can be calculated using all sold properties. The condition to this is that the properties had to exist when the appraisal took place so that a comparison between their appraised and sold price can take place. When a new appraisal is carried out, the index is rebased to this new reference period. SPAR methods have been used in New Zealand since the 1960s and have also been adopted in Denmark, Sweden, Iceland and the Netherlands.

In the case of the Netherlands, two sets of data are used to produce a SPAR index. One of the sets comprises the appraisals of properties carried out by local authorities. These cover all the properties in the Netherlands and are produced every four years for taxation purposes. The appraised data has a high correlation with sale price data, and this correlation has been increasing with each appraisal (de Haan, van der Wal & de Vries, 2008). The other set of data used to match to the appraised prices is data collected by the Dutch Land Registry, which covers all property sales. The data is stratified by province and property type, and for each of these a SPAR index is produced and then aggregated using weights.

The main advantages of this method are:

- It is efficient as it makes use of all the sale data, apart from those sales that cannot be matched with an appraisal or those of properties built after the appraisal period.
- It is easy to implement from administrative sources if these are available.
- If property valuations are done frequently and these take into account substantial improvements to the properties, this method can effectively deal with quality changes.
- No information on the characteristics of the properties is needed.

The main disadvantages are:

- Appropriate sources with good coverage are required to implement this method.
- This method is very dependent on the quality of the appraisals. In many cases not all the regions or all the appraisers use the same methods, which can lead to some data being less comparable.
- If appraisals are infrequent this method cannot deal with quality changes.

2.6 International use of methods

Table 2.3 shows a sample of the different methods used by official statistics around the world.

	Stratified cell matrix	Hedonic regression	Repeated measures	Sale price appraisal ratio	Other
UK		X (ONS, NISRA)	X (Land Registry)		X (RoS)
Australia	X				
Austria		X			
Brazil	X				
Canada			X (Teranet-NBC)		X (StatCan new housing index)
Denmark				X	
Finland		X			
France		X			
Germany		X			
India					X (weighted price index)
Italy		X			
Netherlands				X	
Norway		X			
Spain		X			
Sweden				X	
United States			X		

Table 2.2. A summary of the methods used by international house price indices. The UK has more than one official index (discussed in Chapter 4), and Canada have a repeated sales index for existing dwellings and a separate index for new built dwellings.

3. ONS HPI methodology

This chapter explains in detail the methodology used by ONS to produce the house price index and its accompanying statistics. The methodology is provided in four main sections with the first giving an overview of the methods applied and why. Section 3.2 details precisely what data are used and Section 3.3 shows how the average price of a house is measured. Section 3.4

concludes with the final calculations for the HPI, the chain linked series and 12 month percentage change.

3.1 Aim of the index

The aim of the ONS House Price Index (HPI) is to measure the change in the average house price for owner-occupied properties in the UK, its component countries and regions. The index is constructed from transacted property prices and is calculated on a monthly basis. The index is chain linked to produce an index series that allows comparisons to be made across years.

All dwellings have different characteristics that affect their price, such as location and size. As an “average house” based on such characteristics does not exist in practice, the average house price attempts to find the average price across all houses with differing characteristics. Each characteristic has its own effect on the price of a house and this needs to be taken into account when calculating the average house price. A suitable model to capture the effects of these characteristics is a hedonic regression model, such as described in Chapter 2. The main motivation for using this technique is that due to the large number of combinations of housing characteristics and the limited amount of data available on a monthly basis, it is widely acknowledged as the best method to produce house price estimates. For example, there may be a combination of characteristics for which no properties have been sold. A hedonic regression model developed using the available data allows the average house price to be estimated for each possible combination of characteristics (see section 3 of this Chapter for more details) in the absence of such data.

3.2 Data used in the estimation of the average house price

The data used to estimate the average house price come from a sample of mortgage completions collected by the Council of Mortgage Lenders (CML) via the Regulated Mortgage Survey. Although this survey covers all UK mortgage lenders, lenders covering 75%-80% of the mortgage market submit data that is used to calculate the HPI. Only mortgage transactions for owner occupiers are included in the index; cash purchases are not included in the CML data set but they account for a lower proportion of all houses purchased by owner occupiers. Buy to let properties and remortgages are also excluded. The use of mortgage completions reflects the prices currently being paid for dwellings purchased with mortgages. This then means that any mortgage approvals that never advance to completion are not included. The use of data from actual transactions provides a more reliable indication of house prices.

The size of the sample used is dependent on the number of mortgage completions in the month of interest. Across the months of 2012 the average sample size was 27,000 transactions. However, the housing market activity varies within the year; the summer months are typically more active than the winter months, and the sample size fluctuates accordingly.

Unreliable data such as an extremely high price for a low price location can affect price estimates. For this reason the data is first cleaned by investigating records that are highlighted in a number of validation checks. Additionally, outliers (atypical values) are removed to reduce

the chance of dwellings with very low or very high prices within each unique combination of characteristics distorting the index. All observations not identified as outliers are used to estimate average house prices. In 2012, on average 0.1% of the typical sample was excluded every month, either during the data cleaning or the outlier identification stage. For more detail on how the outlier identification is performed see Annex B.

As noted above, cash sales are currently excluded from the ONS HPI but this takes out a reasonably large proportion of all house sales. The following information is based on initial work to evaluate a proof of concept only – the results are experimental, for information only and should not be quoted as definitive estimates until confirmed by the authors.

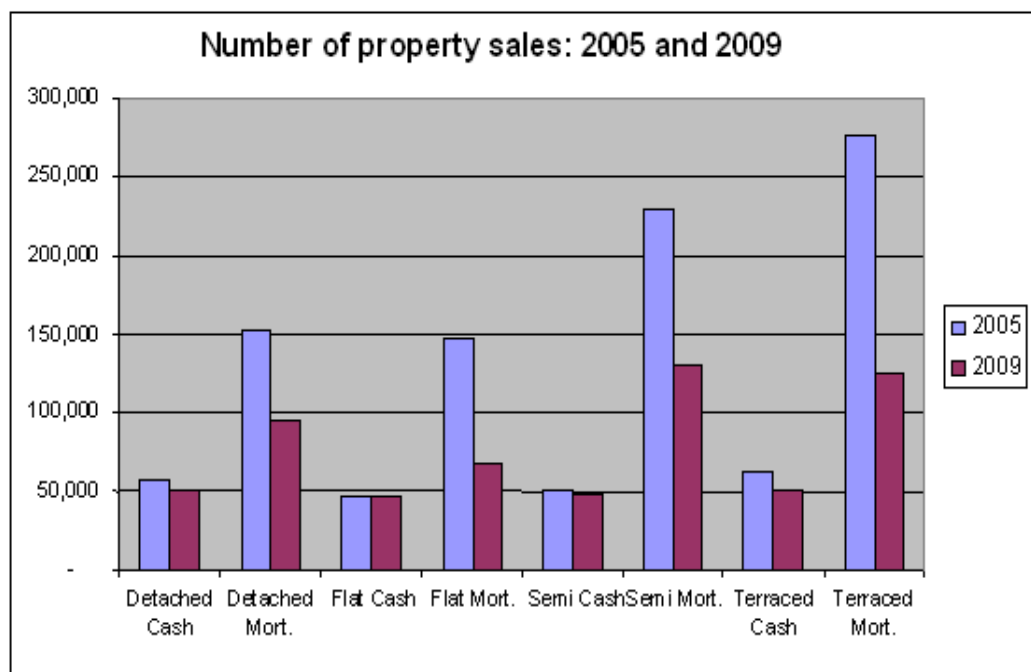


Figure 3.1. Number of cash and mortgage property sales by type of property in 2005 and 2009.

Figure 3.1 shows the number of property sales between 2005 and 2009, years before and after the economic crisis began. It shows that the number of transactions decreased significantly, with the total number of sales between the periods falling from just over one million to 620,000 (not on chart)

Interestingly cash sales were not affected in the same way as mortgages, falling by just 10% compared with a fall of nearly 50% for mortgages. The picture is similar across most property types and results in cash purchases increasing as a proportion of all sales, from 21% of the total in 2005 to 32% of the total in 2009 (not on chart)

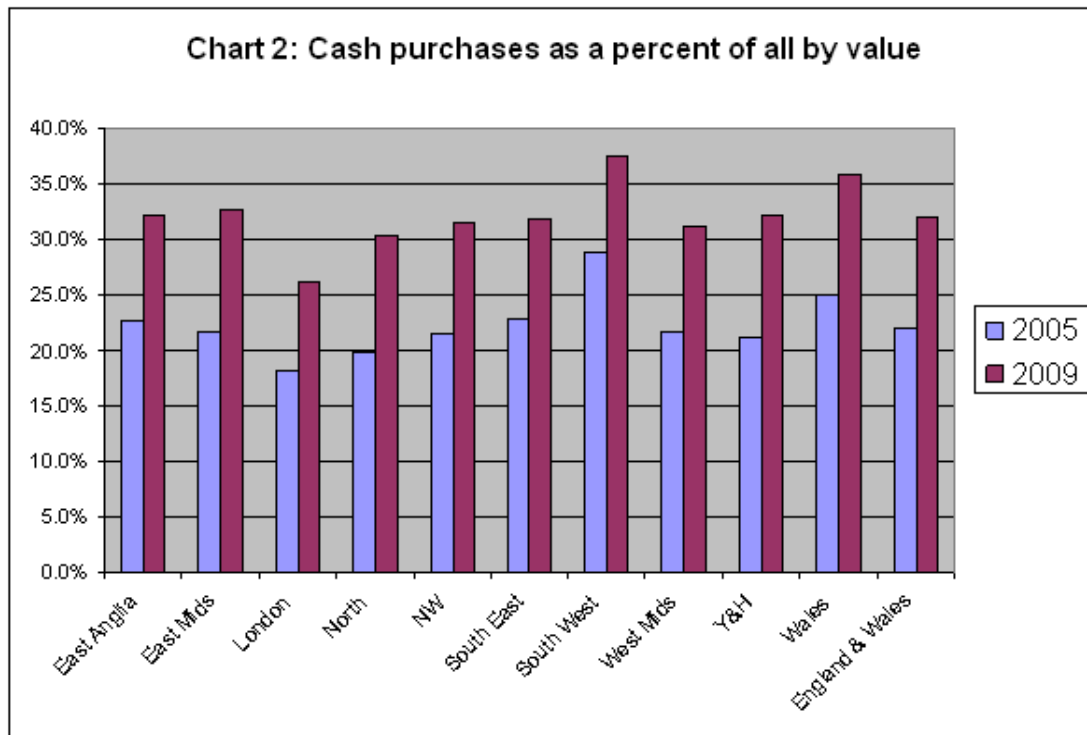


Figure 3.2. Cash purchases by region as a percentage of all purchases by value in 2005 and 2009.

Figure 3.2 shows that the cash sales issue is not confined to one region of England and Wales (analysis for the UK is currently not available). Indeed, it shows that as a proportion of all sales, those transacted with cash increased between the two years in all regions. The exclusion of cash sales from the HPI might not be a serious one if the rates of house price inflation for cash and mortgage transactions are the same. However, Figure 3.3 illustrates that there were noticeable differences between the price movements for cash and mortgage purchases between 2005 and 2009 (caveat – these estimates are simple averages and have not been mix-adjusted at this point).

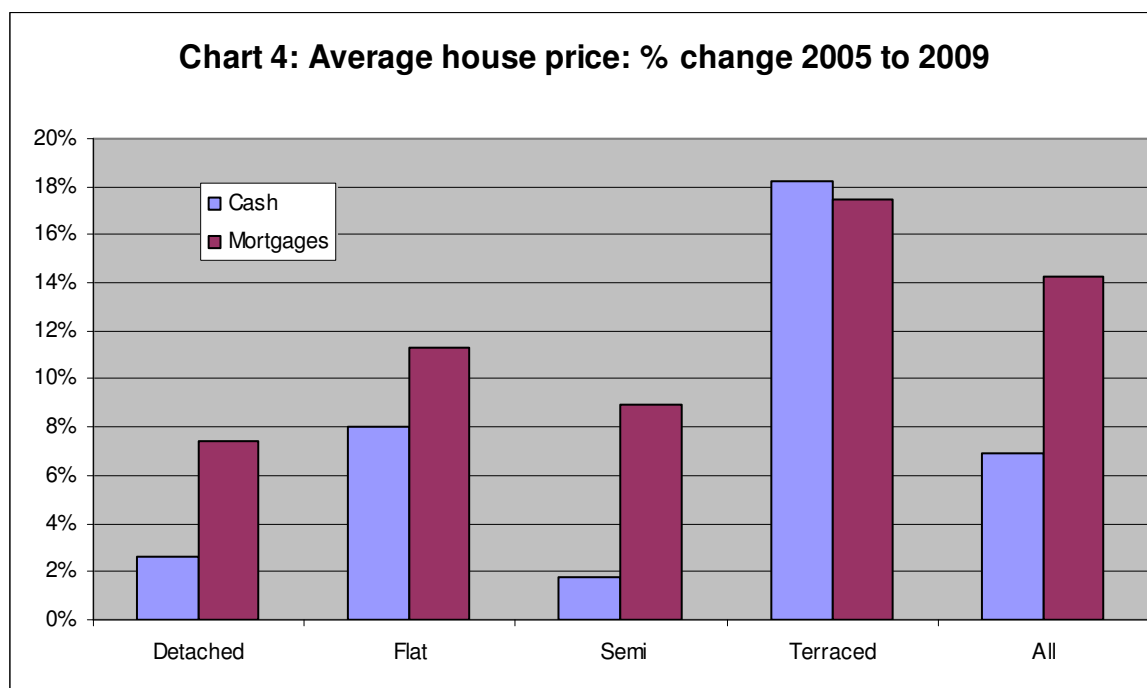


Figure 3.3. Average house price increases between 2005 and 2009 for cash and mortgage purchases by type of property.

Looking at all transactions, house prices for cash purchases increased by just under half those for mortgage purchases (figure 3.3). In summary, cash sales made up a significant and increasing share of all transactions between the two years considered; house price inflation appears not to be the same for the two types of transactions, which suggests it is important to represent cash sales in the ONS HPI.

3.3 Measuring the average house price

The calculation of the house price index is based on first determining the average house price for groups of dwellings with each particular set of characteristics, which are known as cells. The average house price for each cell is calculated and these averages are then weighted together to calculate the average house price for all dwellings. The method used to determine the average house price and a description of how the cells are defined are given in section 3.3.1. A description of how and why observation weights are applied to observations with missing data is provided in section 3.3.2.

3.3.1 Determining the average house price for a given set of characteristics

The ONS house price index uses a hedonic regression model to predict the average house price for a given set of characteristics (a “cell”, see above). This is a model that represents how the change in the combination of characteristics affects the price of a house. A summary of the statistical definition of the applied model is available in Annex A; whilst a full specification of the hedonic regression model used to calculate the index is available on the [ONS House Price Index Guidance and Methodology webpage](#).

The characteristics used in the ONS regression model are:

- Type of dwelling (bungalow, detached, semi-detached, terrace, flat)
- Whether the property is an old or new dwelling
- Number of habitable rooms/number of bedrooms
- Whether the buyer is a first time buyer or a former owner occupier
- County/London borough
- Type of neighbourhood (Acorn)⁵
- Local authority

In addition to the seven characteristics used to estimate house prices in the regression model, three interaction terms are also included.

- Dwelling type x old or new property
- Acorn group x Dwelling type
- Acorn group x first time buyer or former owner occupier

An interaction between two characteristics can be described as the effect of one of the characteristics on house price being different depending on the value of the other characteristic. To demonstrate this, consider the following example.

In general, old properties are more expensive than new properties and detached properties are more expensive than terraced properties. Additionally, the increase in price of a property being old instead of new might be more pronounced in the detached properties than in the terraced properties. This increase in price difference is called an interaction effect, and it shows that the effect that house type has on price varies between old and new properties. Interactions often

⁵ ACORN is a geo-demographic classification of all the neighbourhoods of the UK. It provides basic socio-economic information and is produced by CACI, More information can be found on <http://www.caci.co.uk/acorn-classification.aspx>

capture more subtle effects in the data, such as increased prices of detached properties in desirable urban locations.

The characteristics included in the ONS HPI model have all been found to be statistically significant in explaining the variation seen in house prices and are therefore used to define the cells. This means that each of the characteristics has a non-negligible influence on price that is unlikely to have occurred by chance. This level of influence is different for each characteristic. In the ONS HPI “Number of rooms” has more influence on house price than any other characteristic, explaining the largest amount of variation in house prices. The three interaction terms have also been found to be statistically significant. The characteristics used in the model are, of course, limited to the information collected about properties and contained in the mortgage transactions dataset.

To demonstrate the effect a characteristic may have on house price, figure 3.4 has been created using house price data which shows how house prices might change as the number of bedrooms increases.

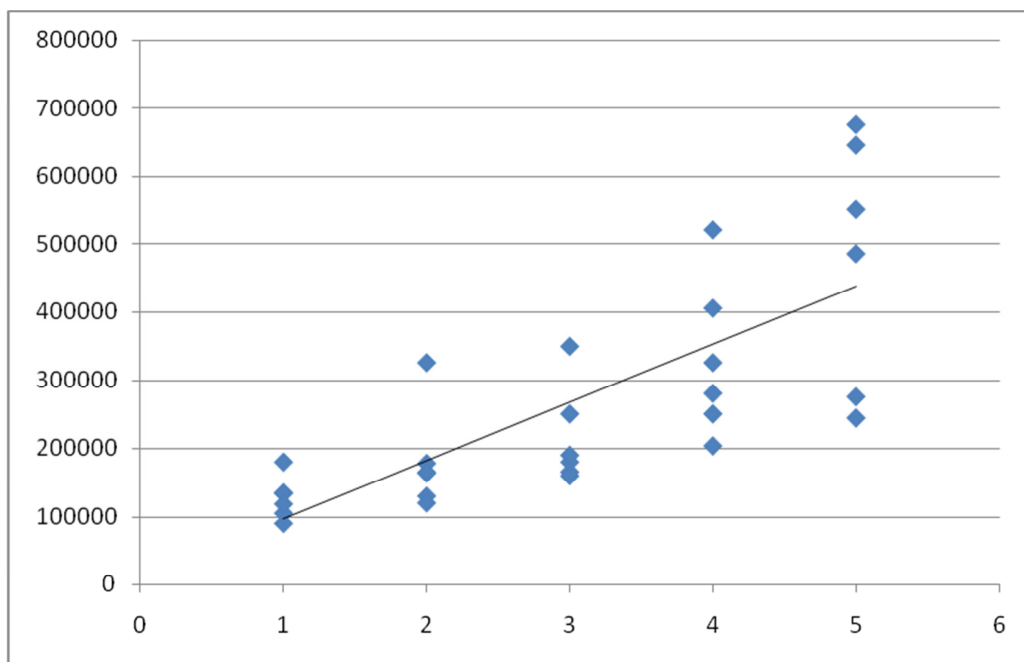


Figure 3.4: Scatterplot of number of bedrooms against house price with fitted regression line

Figure 3.4 shows that as the number of bedrooms increases, house prices tend to increase (note that in the actual ONS model the logarithm of the price is used, see Annex A). This general behaviour can be described by fitting a straight line through the data points using simple regression. The location of this line is determined so that it minimises the deviation of the data points from the line. The line on Figure 3.4 is therefore an example of a linear regression model which best represents the relationship between the variables “number of bedrooms” and

“house price”. For a given number of bedrooms, the regression model can then be used to predict the price of a house, and it is this prediction which is taken as the “average price”. Even though the data may not contain an observed price for a particular type of house, the price can still be estimated based on its characteristics. For example, if there were no prices recorded for properties with five bedrooms, the line provides a good estimate of the average price for this type of property.

This is an example of simple linear regression which has only a single predictor variable (number of bedrooms) for describing the price of a house. However, number of bedrooms is only one of the seven characteristics used to predict house prices in the model and hence is only one of the many effects these characteristics can have on house prices. The basic principle of regression is demonstrated above, but the example below considers how multiple characteristics may affect house prices.

Let’s consider how the change in two characteristics can affect the price of a house. For example, take the two properties below with the same characteristics apart from location.

House 1: £200,000

- Semi detached
- Old build
- 3 bedrooms
- First time buyer
- Location A

House 2: £188,000

- Semi detached
- Old build
- 3 bedrooms
- First time buyer
- Location B

In location A the house sells for £200,000; a house with the same characteristics in location B sells for £188,000. With all other characteristics being equal, the location has had an effect on price. Now, let’s say a third house has 4 bedrooms and is situated in location B.

House 3: £195,000

- Semi detached
- Old build
- 4 bedrooms
- First time buyer

- Location B

This third house sells for £195,000. A change in two of the characteristics has had a different effect on the original price of house 1. The effects of all the available characteristics need to be combined in order to model house prices in an effective way. A hedonic regression model takes into account the different combinations of characteristics and how each combination affects the price of a house.

As there are seven different predictor variables and three interactions in the ONS HPI model, this generates a very large number of combinations of characteristics. There are approximately 100,000 different cells (a cell could, for example, be a new one bedroom flat, situated in a wealthy borough of London, purchased by a first-time buyer) and the hedonic model is used to calculate a mean house price for each cell. For some of these 100,000 cells there will not be any observed data in a particular month. The hedonic regression model allows the prices for these cells to be predicted so that each unique combination of characteristics has its own average house price. This gives coverage of all house types and not just the houses sold in one month, which may not be representative of houses traded over a longer period of time. The average house price for each cell is then weighted together using transaction weights to give an overall average house price (see section 4.1).

3.3.2 Observation weights

For most transactions the dataset contains all the values of the characteristics; however, in a few cases, there are missing data for one or more of the variables. Both complete and incomplete observations are used in the ONS regression model, but incomplete observations have less influence on the model. The influence of the characteristics on the price is implemented as a weight, with more important characteristics having higher weights. This makes the hedonic regression, in practice, a weighted regression model. The weights are referred to as observation weights, and the estimation of these weights is described in Annex C.

A different observation weight is assigned according to the combination of missing data for house characteristics and how important the missing data is in predicting house prices. If the data missing are considered to be very important, then this observation will be assigned a small weight. In the same way, missing data of low importance will mean that a higher weight is assigned to the observation. The 'importance' is decided by comparing model fits using only complete observations and regression models which use different combinations of the characteristics. Complete observations have an observation weight of one and observations with missing characteristics have an observation weight between zero and one.

New observation weights are calculated annually and introduced into the January HPI. The data used for observation weight calculation is the October and November data sets from the previous year. For further detail on weight calculation see Annex C.

3.4 Calculation of the House Price index

To calculate an average house price for all dwellings, the average house prices calculated for each of the cells need to be combined. The calculation of the average house price for all dwellings is described in section 3.4.1. Section 3.4.2 then describes the use of the average house price for all dwellings to calculate the house price index. Sections 3.4.3 and 3.4.4 go on to describe chain linking and twelve-month percentage change respectively.

3.4.1 The Average House Price

To ensure that the HPI reflects a pure change in the price of houses and not just the change in the composition of houses sold in that month, the average house price is mix adjusted. Mix adjustment refers to determining the average house price for a dwelling with a particular set of characteristics (a cell) and combining these average house prices using weights. The weights used in the ONS HPI are based on the number of transactions of every type of property in the past three years.

The way in which a change in the composition of houses sold can distort the index can be easily demonstrated. For example, if by chance in a month more expensive properties were sold than in the previous month, using a simple (unweighted) arithmetic mean would indicate that the average house price had increased even if the prices of houses had remained the same.

Consider the following table:

Month 1			Month 2		
House	Number Sold	Average Price	House	Number Sold	Average Price
A	20	£100,000	A	20	£100,000
B	30	£200,000	B	30	£200,000
C	10	£300,000	C	10	£300,000
D	1	£500,000	D	5	£500,000
Total	61	£11,500,000		65	£13,500,000
Average House price		£188,525	Average House price		£207,692

Table 3.1: Comparison of arithmetic means over a time period when house prices do not change

In month 1, the arithmetic average of the four types of houses sold is £188,525. A month later the same number of houses of types A-C are sold again for the same price with an additional four of house type D. House prices have not changed in this one month period. The average house price in month 2 is now, however, £207,692. The increase in average house price is due to a larger frequency of sales of the more the expensive property type D. This demonstrates how changes between periods in the number and type of transactions can affected the calculated average price. In the example house prices appear to have risen, when they actually satayed the same. This method is therefore misleading and an alternative method cshould be used whereby the numbers of the different types of transactions (the transaction weights) are held fixed

between successive periods. Using fixed weights means that the results are not dependent on which types of houses are sold in a particular month.

Transaction weights are applied to the average house price for each cell to give an arithmetically weighted average house price. They are calculated each January based on the previous three calendar years and are fixed for the remainder of that year. As the weights reflect a different mix of properties, price comparisons between years have to be carried by chain-linking (see 3.4.3). The HPI better reflects the current housing market trends by carrying out this weight update. A transaction weight will be high if that house type had a high frequency of sales. Similarly a house with a particular combination of characteristics that has a low frequency of sales will be given a low weight. This is demonstrated in table 3.2.

House	Number Sold Over Previous 3 calendar years	Transaction weight
A	200,000	0.32787
B	300,000	0.49180
C	100,000	0.16393
D	10,000	0.01639
Total	610,000	

Table 3.2: Calculation of transaction weights

The transaction weights calculated in table 3.2 for each of the houses A-D are then applied to the average price of these houses. The average house price is the sum of the prices multiplied by the corresponding transaction weights. For a technical definition see Annex D.

Month 1			Month 2		
House	Transaction Weight	Average Price	House	Transaction Weight	Average Price
A	0.32787	£100,000	A	0.32787	£100,000
B	0.49180	£200,000	B	0.49180	£200,000
C	0.16393	£300,000	C	0.16393	£300,000
D	0.01639	£500,000	D	0.01639	£500,000
Average House price		£188,525	Average House price		£188,525

Table 3.3: Using mix adjustment to calculate average house prices

In table 3.1, the composition of houses sold indicated an increase in house prices when the prices did not change. Table 3.3 shows that when the fixed transaction weights are applied, the average house price stays the same, consistent with the house prices staying the same.

3.4.2 The House Price Index

The House Price Index is calculated as the weighted average house price for all dwellings in the

current month divided by the weighted average house price for all dwellings in January of the same year, known as the base period. This result is multiplied by 100 so that indices below 100 show prices lower than those in the base period, and indices above 100 show prices higher than those in the base period. See Annex E for the mathematical formula defining the index.

3.4.3 Chain Linking

An index based on the most recent January is calculated each year, starting with January of the base period, which always equals 100, and running up to January of the following year. These January based indices are “chain linked” from one year to the next in order to create an index series that is continuous and on the same scale. Month is the month designed as base as this is the month that the weights are updated. This means each January will have a house price index of 100 before chain-linking is performed.

The index series in any year is scaled to the previous year by multiplying the chain-linked value of the index in the most recent January by the index in each month. For example using the following data set, the March house price index can be calculated before and after chain linking.

Period	Average House Price	Index (not linked)	Index (linked)
Jan-04 (2003 weights)	163,645	109.7	135.4
Jan-04 (2004 weights)	162,559	100.0	
Feb-04	160,937	99.0	134.1
Mar-04	161,306	99.2	134.4
Apr-04	168,600	103.7	140.5
May-04	170,719	105.0	142.2
Jun-04	173,756	106.9	144.7

Table 3.4: Linked and non-linked HPI (rounding has been applied to one decimal place)

The house price index prior to chain linking for March 2004 would be:

$$\frac{\text{average house price in March 2004}}{\text{average house price in January 2004}} \times 100 = \frac{161,306}{162,559} \times 100 = 99.2$$

After chain linking the house price index in March 2004 would be⁶:

$$\frac{\text{linked index in January 2004}}{\text{index in January 2004}} \times \text{index in March 2004} = \frac{135.41880}{100.0} \times 99.22971 = 134.37$$

For a more technical definition of chain linking see Annex F.

⁶ Figures used for this calculation have been rounded to five decimal places to avoid rounding error in the answer incurred when using figures rounded to one decimal place.

3.5 ONS Publication

The ONS HPI follows a monthly publication schedule. At the time of publishing only the data from the month previous to the last is available, for example, in June the April data will be available. Therefore, the publication of the ONS HPI is lagged by two months. Each month the above methodology is used to produce average house price estimates and price indices at a UK, country and regional level. The house price data are also disaggregated by type of buyer (first-time buyer and former owner-occupier) and type of dwelling (new build and pre-owned). Currently the publication date coincides with the publication of both consumer and producer price indices (CPI and PPI). Full details of the ONS house price index publication can be found on the ONS website.

4 - Differences between other official statistics and ONS HPI

This section reviews the methods used by official sources for house price indices in the UK. The different methods and the resulting price levels and index movements are compared with those of the ONS HPI. The official statistics covered are the property price indices produced by Land & Property Services (LPS) in conjunction with the Northern Ireland Statistics and Research Agency (NISRA), Her Majesty's Land Registry (LR), and Registers of Scotland (RoS).

4.1 Northern Ireland Residential Property Price Index (NI RPPI)

LPS/NISRA [Northern Ireland Residential Property Price Index](#) (NI RPPI) uses a similar approach to ONS in terms of modelling the effects of several variables on the price of houses. However, their hedonic regression model is more detailed for Northern Ireland than the ONS model. For example it has more specific regions and includes floor space in the characteristics that can affect house prices.

LPS also has access to a more complete data set regarding sales of properties than the ONS does for Northern Ireland, holding approximately three times the number of transactions. The LPS dataset is in part larger due to the inclusion of mortgage lenders whose activities are limited to Northern Ireland and in part due to the inclusion of cash sales, which are excluded from the ONS HPI. The availability of cash sales is important as it may be that there are significant differences in the way in which prices are set for these additional sales. However, LPS does not identify or account for cash sales in their model so their impact is difficult to assess. The inclusion of cash sales would have an effect on the price estimates of any given house. Generally, Northern Ireland prices are higher in the ONS dataset than in the LPS dataset.

Once the estimated prices have been produced by their respective hedonic regression models, LPS and ONS use different methods to aggregate together these estimated prices. ONS relies on an arithmetic average, while LPS uses a geometric average. Given the same values, geometric averages will be lower, leading to a lower estimate of average price compared to the ONS price. For their weights, LPS uses the proportions of each model characteristic from the transactions within the last year to build an 'average property' for which a price is estimated by applying the parameters from the hedonic regression model. ONS takes a slightly different approach by estimating prices for all the different property combinations and applying weights based on the transactions of the three previous years to aggregate them into average properties. Both approaches are in practice very similar, with the main differences being the period of the weights (one or three years) and the aggregation method (arithmetic or geometric). These differences in aggregation formula are known to produce a significant difference between average house prices, with the NI RPPI average house price being lower than the ONS average.

The ONS estimates a single model for all property types included in the mortgage data set, with type of property included as one of the factors influencing price. LPS estimates separate models for houses and apartments as in their dataset the floor space is measured differently for apartments and there are extra characteristics which have been shown to affect the price of apartments but not houses. The estimates for house and apartment prices are weighted together using proportions of sales in the period being examined to produce the NI RPPI headline figure (see NI RPPI methodology document for further information). Similar differences are allowed for by ONS by including interacted terms in the pricing model.

The differences between the LPS and ONS approaches can lead to significantly different estimates of house prices, for example with significant differences in the average price of a purchased property (see figure 4.1). However the movements in the indices have been consistent across measures (see figure 4.2.). The differences in average prices can be found to be a sum of differences in the source data, the characteristics included in the models, the weighting applied to the prices or parameters and the aggregation formulae. The use of arithmetic aggregation formulae in the ONS HPI and geometric aggregation in the NI RPPI would cause a persistent difference in average house prices, even if the same data sets and models were used. Currently there is no established view on which is better to use when compiling a house price index.

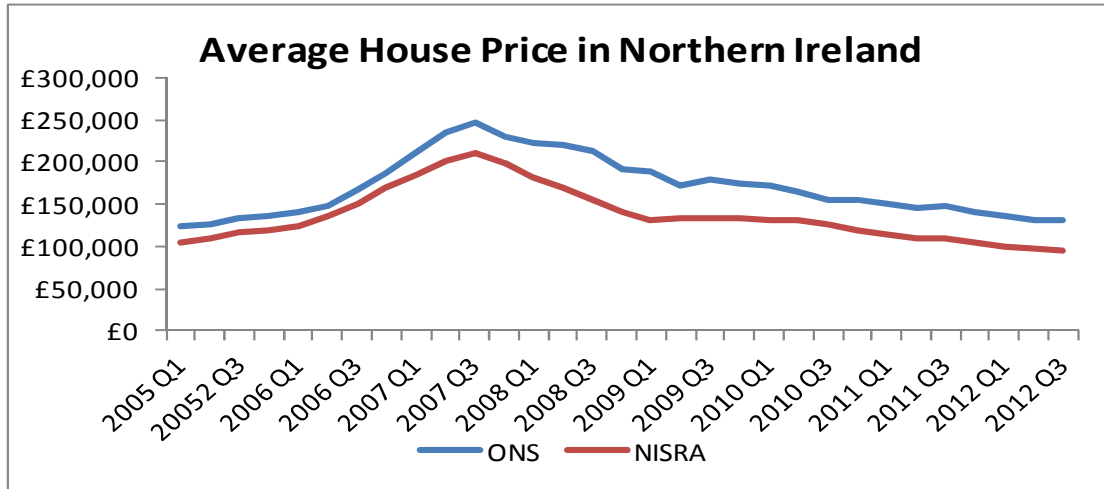


Figure 4.1. Average Northern Ireland house prices for the ONS HPI and NISRA HPI from the first quarter of 2005 to the third quarter of 2012.

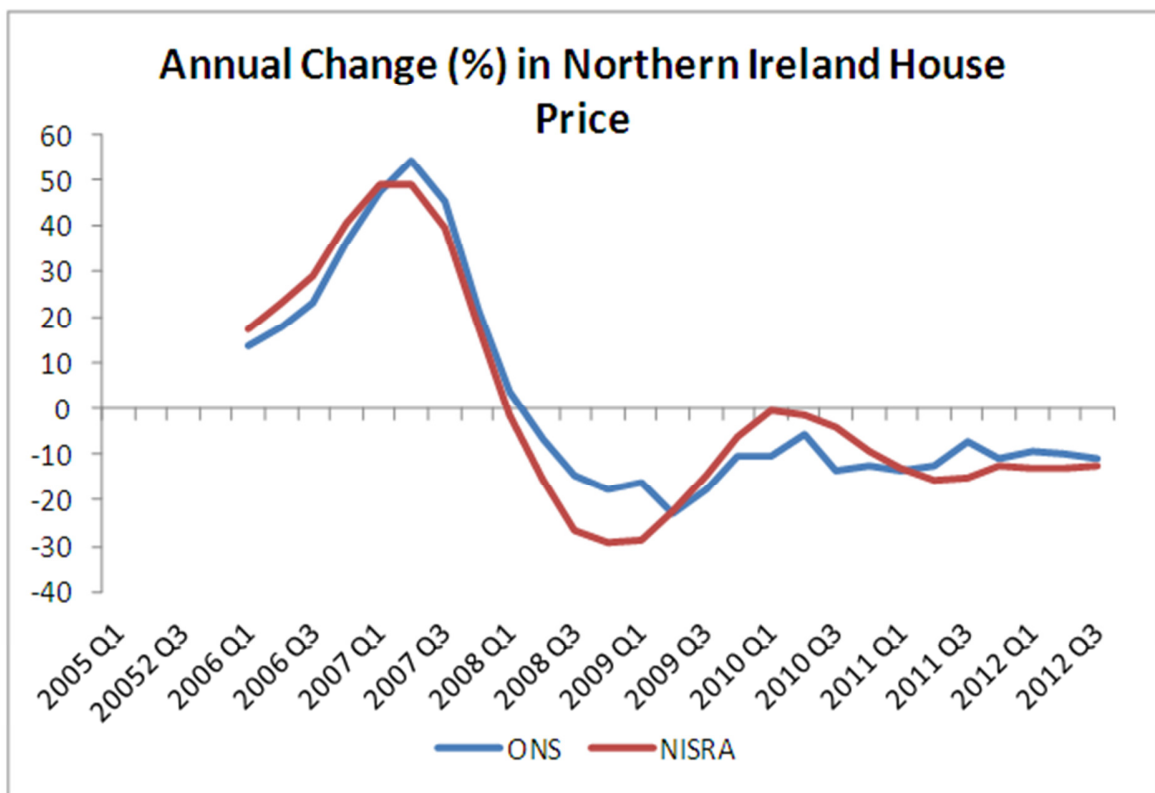


Figure 4.2. Annual percentage change in the ONS HPI and NISRA HPI from the first quarter of 2006 to the third quarter of 2012.

4.2 Her Majesty's Land Registry

[Land Registry](#) (LR) house price index has access to all residential property sales data in England and Wales. This includes cash sales, which are excluded from the ONS HPI. Their house price index employs [a repeat sales methodology](#) which looks at properties for which multiple sales have been observed. The basic idea of a repeated sales method is that once a property is sold for a second time, the difference between sale prices can be used to track the change of price over time. In addition to this, LR makes adjustments to compensate for changes in the quality of properties over time, such as renovations.

After each price has been adjusted for quality changes, the remainder of the change in price in the Land Registry model is assumed to be due to house price inflation, although the model allows for some part of the change in price being due to random variation. This estimation is performed using a simple ordinary least squares regression (an example can be found in Chapter 3) where the different time periods are used to estimate the growth rates. Some adjustments are made prior to this regression, such as converting the growths in prices into their natural logarithms and creating dummy variables to represent the time periods. The series is then transformed so that the reference period takes a value of 100. The resulting estimates are the index values for individual time periods. The growth rates in this index are then applied to the average sale price of April 2000 to produce an updated average house price for each period.

Chapter 2 of this article discusses the repeated sales method and some of its advantages and disadvantages. The advantages of the repeat sales regression model include the ease of identification of standard errors of growth rates and an established framework for testing various hypotheses about the data and the resulting estimates of the increase in house price. Additionally this method does not require detailed information on property characteristics that LR is not required to collect.

There are some disadvantages from the implementation of the repeat sales regression model. Firstly it is difficult to adjust accurately for changes in the quality of houses on a consistent basis, as properties are likely to be maintained to a number of different levels. Secondly the method has a potential bias towards houses which are purchased more often, and so may be biased for example towards the growth rate for properties purchased by first time buyers which are turned over more frequently. Thirdly the model assumes that the random element is normally distributed across the sample, which hasn't been tested.

The LR HPI data is derived from their own Price Paid Dataset, which at present contains details on over 16 million sales. Of these, over six million are identifiable matched pairs. This repeated sales method excludes new build properties, which are included in hedonic regression methods, from its price estimates.

Also, although the repeat sales method provides an estimate of the growth rate of purchased house prices it does not suggest a method for reporting the average price. The [LR website](#) tells us that the average price from the LR house price index is computed by taking the geometric mean price in April 2000 and applying estimated growth rates back to 1995 and forwards in

time. This method does not ensure that the growth rate estimated is consistent with the profile of properties observed in April 2000. As this price was initially constructed using the geometric mean, it is not surprising that the estimate of house price is lower, and carries on being lower, than that of ONS price estimate. The geometric mean is known to be lower than an arithmetic mean, although the LR average was an average of sale prices in a single month (April 2000) rather than a weighted average. ONS uses a three year period of sales by property type to construct weights and uses these to construct the average prices. As the index movements that are used to update this average are similar to those observed by the ONS it is not therefore surprising that this difference has persisted over the life of the indices.

Figure 4.3 shows the average England and Wales house price for the LR HPI and the average UK house price for ONS HPI. The ONS average price is the UK average price as opposed to an average price for England and Wales as published by LR. Figure 4.4 shows the annual index change for England and Wales for both the LR HPI and the ONS HPI. As in the case of the Northern Ireland prices above we see that although the house prices reported by the two levels are different the movements in their indices is very similar, which would indicate that either would provide a reliable measure of changes in the house price.

It is clear therefore that the difference between the index, and average prices, estimated by LR and ONS are due to a combination factors. These differences include: differences in estimation technique used for estimation of the growth rate, differences in coverage of the data used to estimate values and differences in the way that the heterogeneity of the differences are controlled for. Given this the two series produce different estimates, especially of average price where LR estimate the effect of inflation on the average house price as at April 2000 whilst ONS estimates an average price that is reflective of current trends in the housing market and reassess this average each year. However it is encouraging that the two measures agree regarding the size and direction of house price changes (figure 4.4).

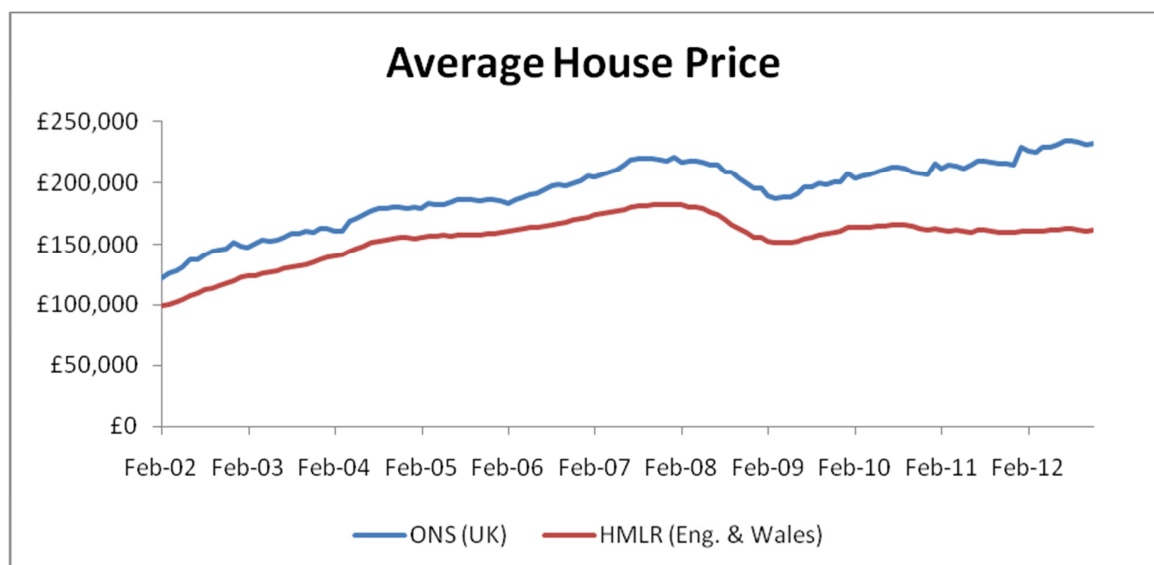


Figure 4.3. Average house prices for LR HPI and ONS HPI from February 2002 to November 2011. The LR reflects price for England and Wales and the ONS price for UK.

Index Numbers

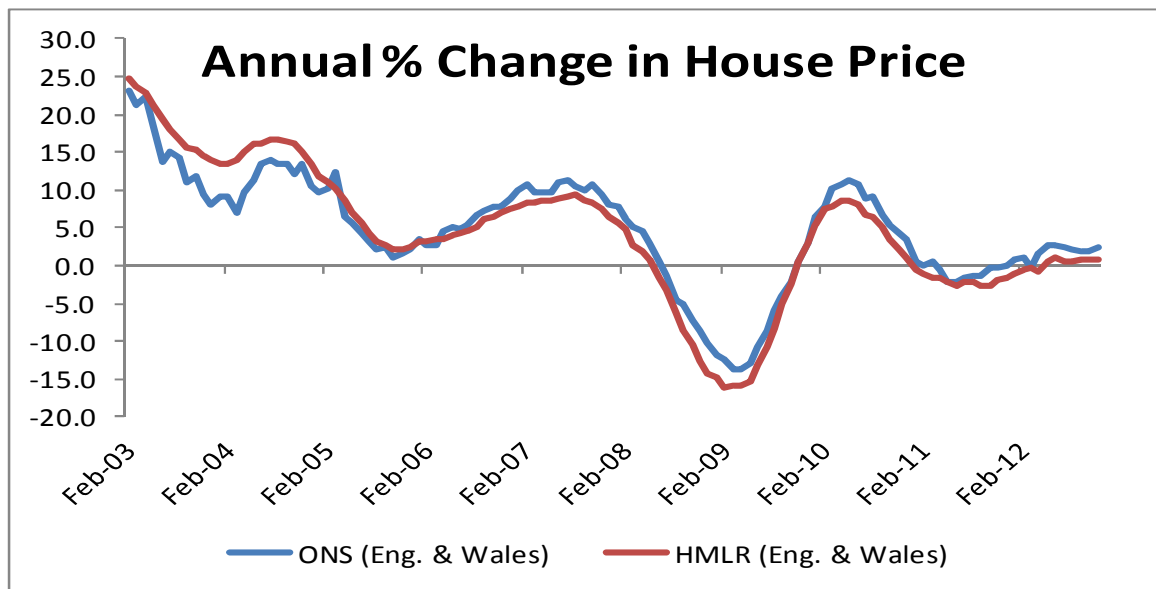


Figure 4.4. Annual percentage change in the LR and the ONS indices from February 2003 to November 2012.

4.3 Registers of Scotland

[Registers of Scotland](#) (RoS) report a number of summary statistics for sales of residential properties which are registered⁷ with RoS during a quarter and where the selling price of the property lies between £20,000 and £1,000,000⁸. The data used to produce these statistics undergoes a rigorous validation check to ensure that only appropriate data is included. The statistics produced include an arithmetic mean price, median price and the number of property transactions by property type for Scotland and for its local authority areas. They provide a summary of what has happened to the prices of purchased properties during a given period, and although they are not mix-adjusted they do include cash sales, which are excluded from the ONS HPI.

⁷ Registrations normally take place within a few weeks from the date of the sale.

⁸ Registrations with a value of over £1m are excluded to ensure that a single large value sale does not distort the average and Registrations with a value of less than £20k are excluded from the calculation to ensure that, as far as practical, transfers of part of the title to a residential property are excluded from the analysis

The statistics include general data on house types (detached, semi-detached, flats and terraces), although this data does not feed into any mix adjustment. The mix of properties included in the sample in any given quarter is liable to change, which can introduce a bias in the results due to particular properties being sold in a period. Some of the measures may also be affected by outliers that remain in the data set after removing transactions below £20,000 and above £1,000,000, which remains a risk. For example the mean of house prices may be sensitive to the small number of high value properties which may be observed on a quarter to quarter basis, making it difficult to compare the series through time, although this is to some extent mitigated by having removed the very high and very low sale prices. The method is much simpler than that used by ONS, LR or NISRA and this is likely to make it more readily understandable; however it may be that the added simplicity is achieved at the cost of less reliable measurement of prices through time.

In figure 4.5 we can see a graphical comparison of the ONS and RoS average house prices, and in figure 4.6 a comparison of their yearly movements. The RoS simple average seems reasonably close to the movements obtained from the ONS method. In recent periods the RoS average house price has increased less than the ONS estimate, which may be evidence that there has been a slight change in the mix of properties in the RoS dataset. However the difference in average price is less than that for either the NISRA or LR estimates considered above. Surprisingly the raw mean seems to be doing a good job of tracking the ONS index in this case, something we would expect to change if the mix of properties sold in Scotland changes rapidly at any point in the future.

Therefore, whilst both the average price and annual change in house price are similar for ONS and RoS, there are conceptual differences in the way each department computes these measures. Changes in the mix of the housing market in future periods could lead to larger differences.

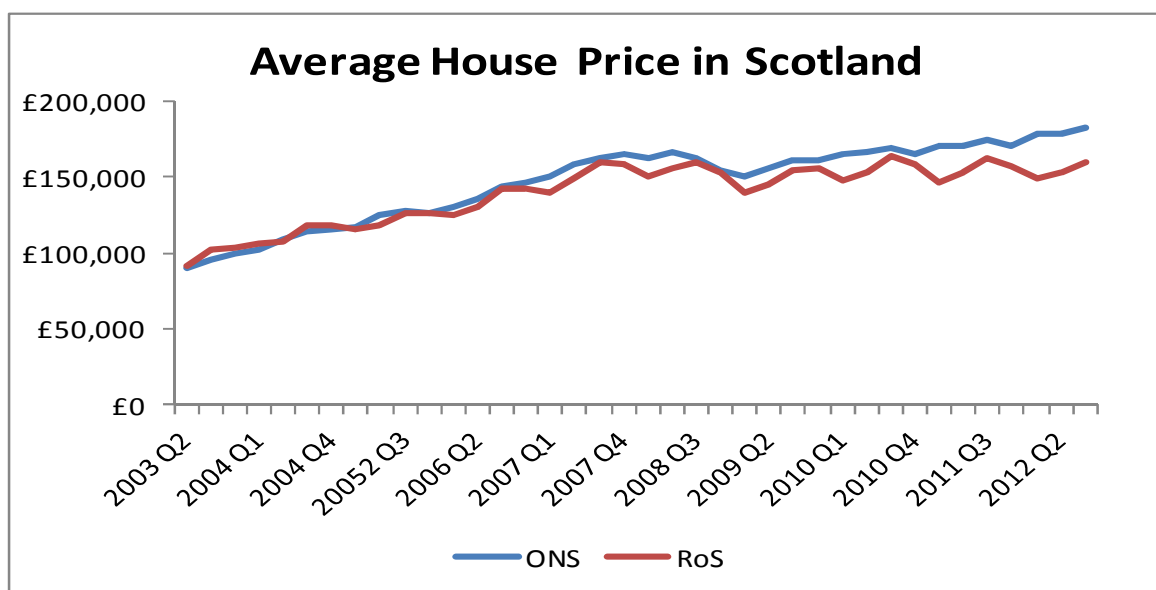


Figure 4.5. Average Scotland house price for the ONS HPI and the RoS HPI from the third quarter of 2003 to the third quarter of 2012.

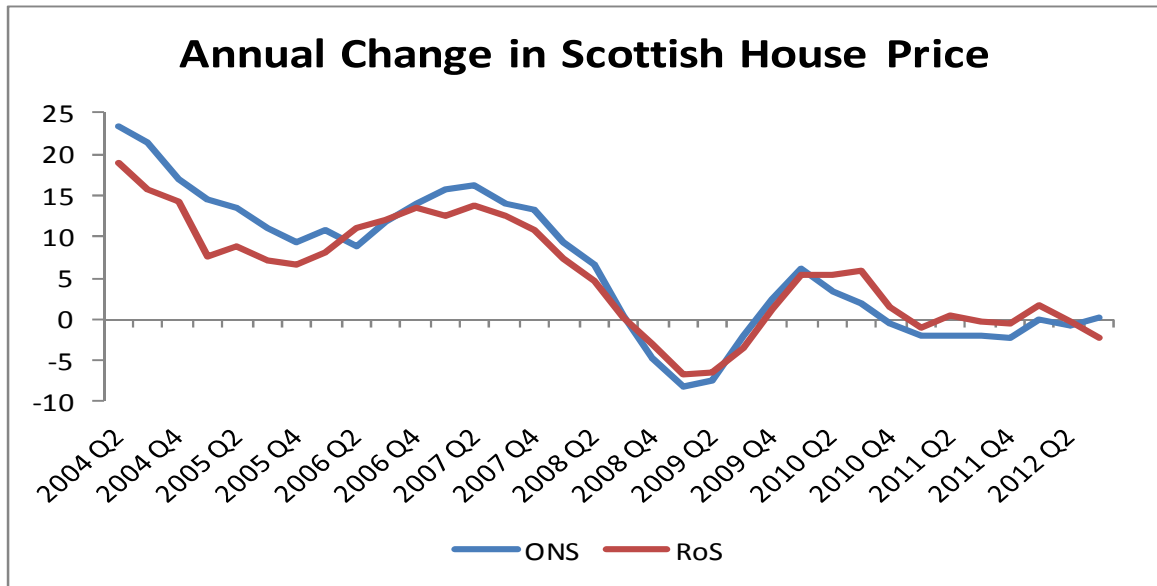


Figure 4.6. Annual percentage change in the ONS index and in the average Scotland RoS price from the second quarter of 2004 to the third quarter of 2012.

5. Conclusions

The purpose of this article has been to explain the methodology used to produce the ONS house price index and to explain why there are justifiable differences between the ONS house price index and other official sources of house price statistics. The measurement of house prices is a difficult process due to the nature of housing compared with other goods and services. In addition to this, there are also different sources of house price data and differing methodologies available for the construction of house price statistics and therefore it can be expected that different house price statistics may produce different results. In Chapter 2 we saw how different methods can be used to produce consistent series of house prices, and how they deal with mix-adjustment. Chapter 3 further explained the methodology used in the ONS HPI, among others, how the index weights the sales in each sample and how average house prices are calculated.

Given the importance of the housing market to the UK economy, it is essential that a rich source of data is available to those people who wish to understand the housing market and currently there are a number of official statistics available for this requirement. In Chapter 4 we saw how official house price statistics differ. A main conclusion was that in all cases the price movements

between the indices are similar, even if due to methodological differences the reported price levels are not comparable. These official indices are often the only source for local authority level house price statistics, which are not provided by the ONS HPI, and between them cover the whole of the UK. The Northern Ireland Residential Property Price Index was the most similar to the ONS HPI in terms of the statistical model used, although the characteristics used in the regressions are not identical but are tailored to the Northern Ireland market and data availability. However, the price levels it reports are not directly comparable to those of the ONS as different aggregation methods and data sets are used. The NI RPPI is based on all sales, including cash sales whilst the ONS methodology uses a very large sample of completed mortgages, thus excluding cash sales and some of the lenders who operate solely in Northern Ireland. In the case of the Land Registry (LR) HPI, the methods are different to those of ONS, making the price levels difficult to compare. LR provides local authority house prices for England and Wales, and although the price levels are not comparable to those of the ONS, the price movements are similar. Registers of Scotland (RoS) provides local authority level price statistics for Scotland, and although these do not constitute a price index, similar price movements to those of the ONS HPI can be implied from them. This confusion regarding official sources was well documented in the 2010 National Statistician's review of House Price Statistics (AD REF) which recommended that better commentary, explanation of methods and differences with other sources should be produced by the official statistics producer community to help the user community.

This article is a first step towards informing users of the methods used and differences in the production of official house price statistics. Discussions have been taking place between the main producers of official house price statistics to identify what work is required to take forward the recommendations of the 2010 review and the subsequent follow up review in 2012 which covered the wider housing market (ADD REF). In these discussions, it has been agreed that the ONS house price index will be developed to include cash sales (the index is currently based on mortgage transactions) to give full coverage of the UK housing market. The ONS house price index is the only index available at a UK level, which is a key requirement of a definitive house price index in the 2010 review. The ONS HPI and the NI RPPI will seek to National Statistic status. When a definitive UK level HPI is developed other official sources of house price data will continue to be published. While the ONS HPI will provide the user with a UK wide perspective, a more detailed local level picture of the housing market will be available from the official statistics produced by Land Registry, Registers of Scotland and LPS. The official producers will work together to provide fuller commentary that brings together each official measure and that helps the user understand each set of data. This suite of data should provide the user with sufficient house price information to make an accurate assessment of the UK housing market.

5.1 Future developments

There are also a number of other ONS house price index developments which are planned for the next 12 months.

- Alongside improving the commentary and explanation of statistics, ONS is looking to improve the monthly publication of its house price index. A user consultation is planned to take place in 2013 to understand what improvements users would like to see in the monthly ONS HPI.
- Work has been taking place to produce regular measures of statistical accuracy to sit alongside ONS house price statistics. A method for the regular production of accompanying standard errors has been developed and should be ready for publication in the second quarter of 2013. Following this initial publication, these standard errors will then be published regularly alongside the ONS HPI.
- Further work will take place to take forward the recommendations that have been made in the 2012 National Statisticians review of Housing Market Statistics. Work in the main recommendation, the production of an index that tracks the price of private rental housing prices, has started.

Annex A-Model Definition

The following is a summary of the model definition for which the full specification of the model can be found on the DCLG website archive⁹.

The distribution of house prices is highly positively skewed so a logarithmic transformation is applied to the observed house prices; the resulting distribution is then approximately normal. The observed characteristics of properties are used in the hedonic regression model to predict this log-transformed house price. The model takes the following form

$$\ln(P_{ymik}) = x'_{yi}\beta_{ym} + \varepsilon_{ymik}$$

Where P_{ymik} is the price of a dwelling k , in cell i , in month m of year y and ε_{ymik} are independent, identically distributed random variables with the following properties:

$$E[\varepsilon_{ymik}] = 0; \quad Var[\varepsilon_{ymik}] = \sigma_{ym}^2; \quad Cov[\varepsilon_{ymik}, \varepsilon_{znjl}] = 0 \quad \text{where } ymik \neq znjl$$

To estimate the average house price, a transformation back to the linear scale is applied. Under this transformation, the expectation of the exponential of the error terms is not zero. A

9

correction term is included in the equation for the price in each cell to account for this potential bias. The average house price in a cell can therefore be estimated as follows

$$P_{ymi} = E[P_{ymik}] = \exp\left(x'_{yi}\beta_{ym} + \frac{1}{2}\gamma_{ym}\sigma_{ym}^2\right)$$

where $\frac{1}{2}\gamma_{ym}\sigma_{ym}^2$ is the correction term and γ_{ym} is a factor that allows for non-normality of the error terms.

In practice the β values are estimated, so replacing the β values above with estimated values, $\hat{\beta}$, leads to the following equation for the average house price in a cell

$$\tilde{P}_{ymi} = \exp\left(x'_{yi}\hat{\beta}_{ym} + \frac{1}{2}\gamma_{ym}\sigma_{ym}^2\right)$$

The estimated values of $\hat{\beta}$ have an associated variance term $Var[x'_{yi}\hat{\beta}_{ym}]$, meaning that the expected value of \tilde{P}_{ymi} is defined as follows

$$E[\tilde{P}_{ymi}] = \exp\left(x'_{yi}\beta_{ym} + \frac{1}{2}Var[x'_{yi}\hat{\beta}_{ym}] + \frac{1}{2}\gamma_{ym}\sigma_{ym}^2\right)$$

An unbiased estimator for the average house price in a cell therefore needs to remove this variance term. The average house price can therefore be estimated as follows

$$\hat{P}_{ymi} = \exp\left(x'_{yi}\hat{\beta}_{ym} - \frac{1}{2}Var[x'_{yi}\hat{\beta}_{ym}] + \frac{1}{2}\gamma_{ym}\sigma_{ym}^2\right)$$

In practice in the House Price Index, a weighted regression model is used to predict the average house price, meaning that the above expression also includes the appropriate observation weights. The full details of the weighted regression model specification can be found in the detailed model documentation available on the DCLG website archive¹⁰.

Annex B – Outlier Identification

To identify outliers, the studentised residual for each observed house price is calculated. A residual, denoted, e_i , is the difference between the predicted house price by the model, \hat{X}_i , and the observed house price, X_i .

$$e_i = X_i - \hat{X}_i$$

The studentised residual, $rstudent_i$, is then calculated as the residual divided by its standard deviation.

10

<http://webarchive.nationalarchives.gov.uk/20120919132719/www.communities.gov.uk/documents/housing/pdf/141410.pdf>

$$rstudent_i = \frac{e_i}{STDEV(e_i)}$$

If the studentised residual is larger than a specified cut off, the observation is identified as an outlier and is excluded from the model. All other observations are used in the regression model.

Annex C – Observation Weights

Observations with missing and complete data for covariates are used in the model. This means that observations with missing values will have a greater variance than observations with values for all covariates and so the assumption of constant variance is no longer valid. In order to comply with this assumption of constant variance the observations have observation weights assigned.

Missing values are treated as a separate category for each variable in the model. For each combination of missing covariates the mean squared error (MSE) is calculated. This is then divided by the MSE of the observations that are complete from the fitted model using all covariates and this number will be the adjustment factor. The weight is then the reciprocal of the factor and hence will be between 0 and 1 (inclusive of 1). An observation weight of one therefore indicates a complete observation where all the variables have available data.

Annex D – Average House Price Calculation

The calculation of the average house price can be written as:

$$\sum T_i \hat{P}_{it}$$

Where \hat{P}_{it} is the predicted average house price for a dwelling with the combination of characteristics

i in period t

T_i is the transaction weight calculated as the proportion of expenditure on dwellings with

combination of characteristics, i , over the previous 3 calendar years.

Annex E – HPI Calculation

The house price index is calculated as the average house price in the current period divided by the average house price in the base period.

$$\frac{\sum T_i \hat{P}_{it}}{\sum T_i \hat{P}_{i0}}$$

Where T_i is the transaction weight for a house with combination of characteristics i

\hat{P}_{it} is the estimated price of cell i in the current period t

\hat{P}_{i0} is the estimated price of cell i in the base period 0

With some manipulation this can be re-written as the sum of the expenditure weighted price relatives:

$$\sum w_i r_i^{0,t}$$

Where w_i is the expenditure weight

$r_i^{0,t} = \frac{\hat{P}_{it}}{\hat{P}_{i0}}$ is the price relative of dwelling with combination of characteristics, i , between the current period and base period.

Annex F – Chain Linking

For example the house price index prior to chain linking for current period t would be:

$$I_{not\ linked}^{Jan_y, m_y} = \frac{\sum T_i \hat{P}_{i, Jan_y}}{\sum T_i \hat{P}_{i, m_y}} \times 100$$

Where T_i is the transaction weight for a house with combination of characteristics i

\hat{P}_{i, m_y} is the estimated price of cell i in month m of year y

\hat{P}_{i, Jan_y} is the estimated price of cell i in January of year y

After chain linking the house price index in March 2004 would be:

$$I_{linked}^{Jan_y, m_y} = \frac{I_{linked}^{Jan_{y-1}, Jan_y}}{I^{Jan_y}} \times I^{Jan_y, m_y}$$

Where m_y is the month m in year y

Jan_y is the January in year y

Jan_{y-1} is the January in the previous year $y - 1$

$I_{linked}^{Jan_y, m_y}$ is the chain linked house price index for m_y with Jan_y as the base period