

The development of a land registry-based national house price index

David Wall
Office for National Statistics

E-mail: info@statistics.gov.uk

National Statistics customer enquiry line: +44 (0)845 601 3034

Summary

The Office for National Statistics and H.M. Land Registry are investigating the possibility of constructing a house price index for England and Wales based on all domestic property transactions as notified to the Land Registry. The purpose of this note is to summarise progress achieved so far and to outline further work still needed.

Background

The most widely-used indicators of house price movements in the UK are those published by the Department of the Environment, Transport and the Regions (DETR - formerly DoE) and those published by Halifax plc and the Nationwide Building Society. Whilst these indices are useful measures of house price movement, none of them is ideal - for a variety of reasons.

DETR Index

This is based on a relatively small (five per cent) sample of completions as recorded by members of the Council of Mortgage Lenders - which comprises almost all UK-based lenders. The index takes no account of property purchases for cash only, which are estimated to account for about 25 per cent of all transactions.

Halifax and Nationwide Indices

These indices:

- also exclude cash purchases;
- are limited to applications for mortgages through these lenders only;
- are based on *approved applications* only, which may not always go through to completion.

The H.M. Land Registry database

H.M. Land Registry is the only institution in England and Wales that maintains a database of **all** domestic property transactions including cash purchases. However, this database was not compiled until mid-1994 and therefore only from that date did a house price index based on Land Registry data become a practical possibility.

The prices in the DETR, Halifax, Nationwide and Land Registry datasets are measured at different points in the property transfer process, as follows:

Halifax and Nationwide	Price at time of approval of application
DETR:	Price at completion
Land Registry:	Also price at completion - but not available until transfer documents are lodged with Land Registry. This can be up to several weeks, even months, after completion.

On the above basis the Halifax and Nationwide indices are the most timely measures. However, some of the applications may not go through to completion; and, for those that do proceed to completion, the price of some transactions will be re-negotiated prior to completion. Consequently the Halifax and Nationwide indices may not be a true reflection of prices actually paid.

In terms of its total coverage, the Land Registry database is at a considerable advantage compared to other data sources. However, its usefulness is constrained because of:

- the time lag between completion and receipt of transfer documents at Land Registry;

- the range of particulars recorded for each property is extremely limited compared to the DETR, Halifax and Nationwide datasets.

Nevertheless, the comprehensive coverage of the Land Registry database justifies further investigation of its potential as a source for a house price index. Consequently, in 1996, the Office for National Statistics commissioned a feasibility study to assess the potential for constructing a house price index based on Land Registry data - see subsequent paragraph.

House price index methodology

Conceptually there are two different approaches to house price index construction:

- the "hedonic regression" approach (the basis for the Halifax and Nationwide indices);
- a simpler, matrix approach (the basis for the DETR index).

Both approaches incorporate the principle of **mix-adjustment**. This ensures that the relative importance of different property types in different regions remains fixed throughout the life of the index - rather like the way the Retail Prices Index measures the changing total cost of a fixed basket of goods and services. If mix adjustment were not used, the house price index would reflect not just changes in house prices but also changes in, say, the proportion of property types sold from one period to the next. For example, if the proportion of detached homes sold in period t was higher than in period t-1, without mix adjustment this could lead to a rise in the house price index even if the general level of house prices had remained unchanged.

Details of these two approaches are described more fully in the Annex.

1996 feasibility study of Land Registry data

To assess the potential for constructing a house price index based on Land Registry data, ONS commissioned a study by Professor M. Fleming and Professor J. Nellis⁴. Fleming & Nellis constructed a house price index based on all property transfers processed by the Land Registry from the second quarter of 1994 to the third quarter of 1995. For analysis purposes they also created comparable datasets and price indices based on DETR data and Halifax data.

To illustrate the relative sizes of the Land Registry, DETR and Halifax

databases - during 1994/5 the number of property transfers for England and Wales, recorded by Land Registry, was about **40 times** greater than the sample of properties included in the DETR house price index for the same period. It was also about **6 times** greater than the number of properties in the Halifax database.

Using the "hedonic regression" approach Fleming & Nellis calculated three different price indices - one for each dataset. Despite the limited range of information on house characteristics, the Land Registry index numbers gave reasonably robust results - in the sense that they compared well with those derived from the DETR and Halifax databases. On the other hand, in spite of the extremely large size of the Land Registry database, its explanatory power⁵ was lower than that provided by the DETR and Halifax databases.

Their report concluded that the usefulness of the Land Registry database would be enhanced by incorporating more information about house characteristics. Previous research had shown that, of those variables missing from the LR database, the most important was an indicator of house size - measured either in terms of floor area or in terms of number of rooms (or bedrooms). The importance of house size as an explanatory variable has been consistently borne out in other studies of house prices that Fleming & Nellis have carried out for the Halifax and Nationwide and for DETR.

Fleming & Nellis concluded that consideration needed to be given to be given to extending the range of information collected by the Land Registry to include other relevant house characteristics - notably a proxy for property size.

Subsequent discussions with Land Registry have focused on how such information might be collected. The obvious solution was to modify the design of the transfer document. However, the range of information sought on the transfer forms is covered by legislation and an amendment to this legislation would be needed to make it compulsory for conveyancers to provide information on "number of bedrooms". This option has so far not been pursued.

As a compromise, Land Registry introduced the question (from April 1998) on a voluntary basis. To date, the level of response has been less than ten per cent. This in part may have been influenced by the fact that solicitors have been able to use the **unmodified** transfer document, which did not ask for number of bedrooms. There are also concerns that the sample of those

responding will be biased. This is because regions covered by conveyancers who provide this voluntary information will be very well represented in the sample but regions covered by conveyancers who fail to supply this information will be under-represented.

Whilst these early figures may not be indicative of response rates in the long term, it does suggest that there may be difficulties in constructing a representative index.

It is hoped that, in due course, the level of response will improve and Land Registry are currently discussing possible ways of achieving this.

The way ahead for the Land Registry option

Once a reliable datastream is available it will be necessary to accumulate a run of such data - covering at least five successive quarters. This will enable the ONS to carry out a further feasibility study to test whether the Land Registry-based price index, **including** information on property size, can be used as a reliable indicator of house price change. Such further study will also consider the relative merits of whether a Land Registry house price index should be based on a hedonic approach or a matrix approach.

Such are the imponderables that still have to be overcome it is too early to say whether a Land Registry-based house price index will ever become the "headline" house price index for England and Wales. Nevertheless the ONS is actively seeking solutions to these problems.

Parallel developments in Scotland and Northern Ireland will be

necessary if a Land Registry price index for England and Wales is eventually to be extended to become a UK house price index.

Other developments

In addition to the collaboration between the ONS and Land Registry, the Bank of England has undertaken a simple, mix-adjustment of the Land Registry dataset (using the matrix approach), in order to supplement the range of house price indicators that inform the Bank's Monetary Policy Committee.

An EU taskforce has been set up to consider the possible inclusion of owner-occupied housing costs in the new Harmonised Index of Consumer Prices (HICP). A report on the interim findings of the taskforce is due in the middle of 1999. If they recommend that acquisition costs of new dwellings should be included in the HICP, the implementation of this recommendation would be helped by the development of a national house price index based on Land Registry data.

To ensure a co-ordinated approach to house price index development and to avoid any duplication of effort, representatives of the ONS, H.M. Treasury, the Bank of England and the DETR now meet on a regular basis to share ideas and discuss alternative developments.

ANNEX

Method 1: The "hedonic regression" approach

A set of house prices, P_i ($i=1,2,\dots,n$), may be observed in any time period (t) in which each house (i) is sold. Given the supply and demand conditions in the housing market, such houses may be priced differently due to differences in qualitative characteristics (such as type of property, the availability of certain amenities, the location of the property etc) and to differences in quantitative characteristics (such as the age of the property and the number of bedrooms etc). Thus, for each house i we can write P_i as some function of these various characteristics, X_j , together with a group of unmeasured factors (assumed to be randomly distributed) which are specific to each house but for which data are not available, e_i . In general terms the relationship may be expressed as follows:

$$\ln(P_i) = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + \dots + b_j X_{ji} + e_i$$

where b_1, b_2, \dots, b_j are the regression coefficients corresponding to the qualitative and quantitative variables, X_j .

The first step in regression analysis is to determine an appropriate functional form for the estimating equation. Box and Cox (1964)¹ developed a statistical test for the functional form providing the "best fit" based on likelihood ratio tests and the procedure they proposed has been adopted in both the Halifax and Nationwide models. The results showed that the **semi-logarithmic** functional form (with the dependent variable P_i measured in natural logs) was to be preferred.

Given the nature of the data employed, qualitative characteristics can only be represented by "dummy variables" which take the value of one or zero depending upon the presence or absence of a particular attribute. For instance, in the Halifax model there are five dummy variables representing house type: detached; semi-detached; terraced; bungalow; flat. Each of these variables take the value 1 if the property is of that type, 0 otherwise. Obviously, for any particular property one of these variables must have the value 1 and the other four must all be zero.

The technique of ordinary least squares regression allows us to estimate the coefficients b_j pertaining to each of the explanatory variables X_j for any set of houses. These coefficients indicate the relative importance of the variables in explaining the variation of house prices in a particular time period.

Of course, the average price for a set of houses within a particular period depends on the number of observations on each characteristic in that period. For instance a rise in price from one period to the next could come about simply because of an increase in the number of detached houses sold as a proportion of the total. This can be overcome by fixing the mix of characteristics between one period and another, by applying a standard set of weights corresponding to the numbers of each characteristic observed in a chosen period - usually the base period. The index numbers calculated then represent the movement in average prices for houses having the same proportion of characteristics as those bought in the base period. The index numbers themselves are computed by comparing the weighted (ie mix-adjusted) prices in each current period with the weighted average price in the base period.

The mix-adjusted house price is then given by:

$$\ln(P_i) = b_{0,i} + W_1 \cdot b_{1,i} + W_2 \cdot b_{2,i} + W_3 \cdot b_{3,i} + \dots$$

where the W_j are the weights - which are fixed over time. The weights are set as the proportion of houses with that characteristic during the base period.

The Halifax and Nationwide indices are effectively geometric means of different house prices, since the b_j can be interpreted as the log of the prices of different cells.

The above methodology is then used to produce base-weighted standardised house-price index numbers, whereby a weighted average of the estimated regression coefficients is calculated (each coefficient being regarded as an implicit characteristics-price). The steps involved are as follows:

- calculate the weights, W_j : the proportions of the qualitative variables and the means of the quantitative variables present in the chosen base period.
- with price recorded in natural log form, use the technique of ordinary least squares to estimate the regression coefficients b_j for the j explanatory variables in both the base period and for every subsequent time period.

- 1 calculate a base-weighted (Laspeyre's type) index for the current period (I_t) as follows:

$$I_t = \frac{\text{antilog } \hat{Y} \sum b_{jt} W_j}{\text{antilog } \hat{Y} \sum b_{jo} W_j} \times 100$$

Summation is over all variables included in each regression.

Both the Halifax and Nationwide house price indices are determined by carrying out a hedonic regression on the entire UK database.

Method 2: The matrix approach

Unlike the wide range of property characteristics available to the Halifax and Nationwide, the range of information used to calculate the DETR index is limited to:

- price;
- region;
- property type;
- age;
- number of habitable rooms;
- type of borrower (first-time purchaser or former owner-occupier).

A 314-cell matrix is constructed, each cell representing a unique combination of the five independent variables.

The weights used in the calculation of the index are revised every January and are based on the proportions of transactions monitored during the preceding three calendar years. So a 314-cell vector is set up with these "weights". A second vector provides the average property price in the base period (1993) for each of the 314 different sets of properties.

To determine the current quarterly house price index the complete set of transactions for that quarter are divided up into the 314 categories, and the average price for the quarter calculated for each category.

The index is then calculated as:

$$I_t = \frac{\hat{Y} \sum W_j P_{tj}}{\hat{Y} \sum W_j P_{oj}} \times 100$$

In other words, the national weighted-average price is calculated for the current period and divided by the national weighted-average price for the base period. Multiplication by 100 gives the current house price index. In the calculation above, the average house price for each cell is determined as an arithmetic mean. Similarly the national weighted-average house price is also computed as a weighted arithmetic mean.

References

- 1 Box, G.E.P. and Cox D.R. (1964), "An Analysis of Transformations", *Journal of the Royal Statistical Society*, 26, Series B, pp 211-243.
- 2 Fleming M, and Nellis J.G. (1984), *House Price Statistics for the United Kingdom: A Survey and Critical Review of Recent Developments*, Loughborough University Department of Economics, Occasional Research Paper No.80.
- 3 Fleming M, and Nellis J.G. (1984), *The Halifax House Price Index: Technical Details*, Halifax Building Society, Halifax.
- 4 Fleming, M. and Nellis J.G. (1996), *Feasibility Study of the Development of a Mix-Adjusted House Price Index using the Land Registry Database*, unpublished report for the Office for National Statistics.
- 5 An ideal model for constructing a house price index would be one where *all* price variation could be explained in terms of changes in the characteristics of the houses sold during the latest month or quarter. In practice, this is not possible because information on certain characteristics (such as quality of decor or environmental quality of the neighbourhood) are simply not available, even though they can influence the price of the house. The *explanatory power* (something less than 100 per cent) is therefore a measure of the extent to which the changes in house prices are attributable to differences in the characteristics of the properties being sold. For example, the explanatory power obtained in the official Halifax house price index regressions is usually above 70 per cent. In the 1994/5 analysis of Land Registry data by Fleming and Nellis, the explanatory power was less than 50 per cent.