

Introduction

It is recognised that there is need for an index of house prices which controls changes in the composition of the houses sold in each period. Account needs to be taken of the characteristics of the houses being sold, of the changes from period to period in the mix of these characteristics. For example, assuming that large houses are more expensive than smaller ones, an increase between periods in the average size of house sold would result in the average house price rising, even if the price of each size house remained totally unchanged.

The largest available data set which provides sufficient information to enable consistent account to be taken of changing characteristics is that possessed by **permanent tsb**.

permanent tsb accounts for approximately 20 per cent of residential mortgage loans in Ireland, and has a complete geographical spread throughout the country. Moreover, since the beginning of 1996, information on each transaction has been collected and collated in a consistent manner which enables a large number of housing characteristics to be taken into account in constructing an index. The data set is also large enough to enable a number of two-way disaggregations to be undertaken on a monthly basis, thus permitting the constructing of a range of separate indices as well as an overall national index.

The Type of Index

The most familiar type of price index, such as the consumer price index, is constructed by taking a representative “basket” of goods and services and pricing this exact basket in succeeding periods. This approach is not feasible for a house price index. In the first place no two houses are absolutely identical, as are two boxes, of, say Cornflakes. In the second place, the number of transactions is relatively low. Thus, even if a representative sample of houses could be assembled, very few of them would actually be sold in any period. Indices based on a constant sample must therefore rely on repeated valuations, with an inevitable subjective element, rather than on actual market transactions to obtain the relevant prices.

This problem is not unique to Ireland, and in other countries alternative approaches to constructing a meaningful house price index have been adopted and successfully applied for many years. The best known of these is probably the Halifax House Price in the UK, which has operated since 1984 and has become recognised as the authoritative measure of UK house prices. It is an hedonic index, which is the best available form of mix-adjusted index, giving results quite similar to those which would be obtained from an index based on a constant representative sample, were such available. A number of variables known to influence house prices are selected, and in each period a multivariate linear regression is run, “explaining” the price in terms of these variables. By substituting

the base period weight in each variable for the actual weight in each succeeding period, that part of the apparent change in price which is simple due to a different mix of characteristics is removed, leaving an index of “true” price changes for a consistent set of characteristics.

The Variables

A very large number of factors can influence the price of a dwelling, and it would clearly be impractical to have every potential factor represented by a variable in the equations. By their very nature, many factors are subjective and are not amenable to inclusion in the data set. However, sufficient factors of an objective nature, capable of being used as explanatory variables, are available in the **permanent tsb** data to explain a high proportion of the price of an individual dwelling. These fall into four main groups: size, dwelling type, location and other characteristics.

1. Size

The **permanent tsb** data contain three measures of size: the area of the dwelling in square feet, the number of bedrooms and the number of rooms. There is, however, a discontinuity in the information concerning the number of rooms, with total rooms being recorded for 1996 and the number of principle rooms other than bedrooms from 1997 onwards. A major potential problem is that the area of the dwelling is likely to be highly correlated with the number of rooms (however measured) and with the number of bedrooms.

Initial tests showed that there was indeed a high degree of multi-collinearity between the variables, with consequent instability in coefficients of each in the regression equations. It was therefore decided that size should be represented by only one variable, with square footage being the obvious choice, both because it provides the greatest range of information and because it showed the best consistency and significance in trial equations.

2. Dwelling Type

The data distinguish between detached, semi-detached, terraced, bungalow and apartment as types of dwelling. Tests indicated that there was no significant difference between detached and bungalow in most periods, while the number of bungalows in the Dublin area was too small to provide consistent results between periods. Similarly, the number of apartments outside Dublin was too small for reliability. Thus for the non-Dublin and national indices the house types included as dummy variables in the equations were restricted to Detached, Semi-detached and Terrace, while for the Dublin index an Apartments dummy was also included.

3. Location

In many ways this is the most difficult factor to deal with. Conventional wisdom dictates that it is of paramount importance in determining the price of a dwelling, but it is clearly impossible to codify in a comprehensive manner. The locational information in the data-set is simply the address of the property, from which it must be accepted that such locational aspects as the quality of the view or the distance to schools or shops cannot realistically be deduced. However, while what could be called "micro-locational" characteristics are unavailable, "macro-locational" variables could be constructed.

It was immediately apparent from the data that, as would be expected, house prices were higher in Dublin than in the rest of the country, even when other factors were held constant. However, again as would be expected, there were significant locational variations within both Dublin and the rest of the country. Examination of the residuals from the early trial regressions indicated clear patterns to these variations. Accordingly it was decided to divide both Dublin and the rest of the country into three price bands: upper, middle and lower. Through an iterative process of running regressions and studying the pattern of the residuals, these bands were refined to the form used in the final regressions, where they proved both reasonably stable and highly significant. They also largely, but not entirely, conform to prior expectations.

In Dublin, the upper price band includes the south inner city, the southern and south eastern suburbs from roughly Harold's Cross to the southern boundary of Dun Laoghaire-Rathdown, the coastal northern suburbs from Clontarf to Malahide, and a stretch of northern inner suburbs including Dromcondra and Castleknock. The lower price band consists of Ballyfermot, Clondalkin and Jobstown. The middle price band consists of the remainder of the city and county of Dublin. Obviously this three-way split is crude, as the dividing line between the bands is not clean-cut, especially with regard to the north-eastern suburbs, while there are clearly differences between areas in each of the three bands. However the regressions show high significance for the variables, while the quality of data would not justify more than a three-way split, at least on a monthly basis.

For the rest of the country, the upper price band includes the cities and suburbs of Cork and Galway, and the areas of surrounding counties contiguous

to Dublin. It would make little sense to include the whole of the counties Wicklow, Kildare and Meath in the upper band, as such towns as Arklow, Athy and Kells are well outside normal commuting range of Dublin, and thus share few characteristics with the essentially suburban fringes of those counties. Of course, the price gradient could be expected to decline fairly gradually with distance from Dublin, so that any cut-off line is bound to be somewhat arbitrary. In practice, we have adopted an irregular line lying roughly between fifteen and twenty miles from the centre of Dublin, running from Kilcoole on the Wicklow coast to Naas and Maynooth in Kildare and round Dunshaughlin in Meath. This upper band variable has proved consistently significant, and inspection of residuals has shown no substantial areas which should be either added to or subtracted from this band.

The lower price band for the rest of the country is less obvious, and does not confirm to pre-conceptions, as in the case of other bands. Purely on the basis of a study of residuals, the areas emerging as lower band are the South Midlands, comprising Counties Tipperary, Laois and Offaly, and the North West, comprising Counties Mayo, Roscommon, Leitrim and Sligo. This definition of the band tends to be confirmed by the consistency of the regression coefficients and their reasonably high significance. As with the other bands there are doubtless pockets within the areas which do not conform to its general characteristics, but there are insufficient observations to identify and re-classify such localities.

4. Other Characteristics

Information is available concerning a number of other factors which could be expected to influence the price of a dwelling. These are: whether the house is new or existing, whether the purchaser is

a first-time buyer, whether the house has a garage, or in the case of an apartment, a car parking space, the type of heating - gas, electricity, oil or solid fuel. Repeated tests showed that the only consistently significant variable here was between solid fuel and all other types of heating, probably because reliance on solid fuel usually implies the absence of central heating.

5. Undocumented Characteristics

Several other factors are likely to influence the price of a dwelling, although information on them is not currently available in the **permanent tsb** or any other data set. The problem of "micro-localational" factors has already been alluded to, and by its nature is likely to remain insoluble.

The condition of the dwelling is clearly of importance, but, at present, information on this is not collated in a usable form. It is hoped that in future such data will become available and can eventually be incorporated in the calculation of the index. Similarly, the size of the plot on which the dwelling stands can have a significant influence on the price, especially if it includes any development potential. Again information is not currently available but could become so in the future. No effort has been made to collect information on the age of the dwelling, apart from the simple division into new or existing. The relationship between age and price could be expected to be quite complex, while some of the effects of age would be subsumed within a "condition" variable, if such becomes available.

The absence of information on these various factors will obviously diminish the explanatory power of the equations underlying an hedonic index, so that it would be unrealistic to expect anything approaching a perfect fit in the equations. It is, however,

reasonable to assume that the mix of these undocumented characteristics will remain fairly stable over time, so that the absence of such variables will not unduly distort the index of price trends. Cases where it appears that the influence of one or more of these factors has an undue effect on the price of a particular transaction are dealt with by removing extreme observations from the equations, as described later.

Constructing the Indices

Having codified the data and selected the variables to be included, the next stage is to run multivariable regressions by ordinary least-squares for the year 1996 and for each month and quarter since the beginning of 1996.

Following established practice, the dependent variable on each equation is the natural log of the price of each dwelling. Of the independent variables, size in square feet is entered as an absolute value, while all others are entered as dummy variables, with a value of either one or zero. Where two or more variables are alternatives, as in the case of house type, one dummy has to be omitted from the equations for computational purposes. In effect, the equations calculate the difference in price of each transaction from that of a "standard" dwelling of average size. It does not matter much which collection of characteristics is adopted as "standard", but, for the record, the choice made was of an existing detached house (or bungalow), in a "medium" location (non-Dublin in the case of the national index), with a non-first time buyer, no garage, and gas, electricity or oil heating.

Because of data errors, or more probably because of an unusually strong influence from one of the undocumented factors, the price of a small number of dwellings in each period is liable to diverge strongly from the price predicted by the equations. To avoid these outliers or anomalies from having an undue influence on the equations, it has been decided to exclude any observation with a residual exceeding + 0.75 (roughly double or half the price expected for a dwelling with its characteristics), and to re-run the equation without these observations. In a typical monthly equation, between one and five transactions are thus excluded from the final equations, although in several months there are no exclusions at all.

In view of the number of undocumented factors, the fit of the equations obtained, ranging from an adjusted R-squared of about 0.55 to 0.75, with an average of about 0.63, can be regarded as satisfactory. However, such a degree of fit does leave the stability of the indices constructed from the equations heavily dependent on the assumption that the influence of the undocumented factors remains reasonably constant over time. As well as the general fit of the equations, it is reassuring that the signs and magnitude of the coefficients for individual variables are fairly consistent, and that most of them remain significant at (at least) 95 per cent level in all equations. The two variables displaying the highest significance throughout and contributing most to the predicted price are size and "upper" location.

The next stage in calculating the price indices is to replace the weights or frequencies in each monthly or quarterly regression with the weights of the base period, and recalculate the predicted price on this basis. This

shows what the predicted price in each period would have been had the average dwelling size and the mix of other characteristics been the same as in the base period, and thus corrects for changing characteristics over time. Finally, the predicted prices for each period are converted back from log form to actual values, and divided through by the base period value to produce an index with 1996 = 100. Separate indices have been produced for all dwellings nationally, Dublin and non-Dublin, and nationally for new and existing dwellings, and for first-time buyers and current owner-occupiers.

The methodology used is applied to each of the categories based on separate sets of data. This is the internationally recognised methodology for calculating House Price Growth (regression based as opposed to weighted average) and explains why the national average growth can be greater than the percentage change for each of the geographic areas.

The Results

Seven separate monthly indices have been calculated from January 1996 to date, all with the year 1996 as the base. The results are set out in quarterly bulletins. It is evident that each index shows a fairly steady trend over the period. Inevitably, there may be some slight apparent instability, especially on the disaggregated indices. For this reason trends are best assessed on the basis of three-monthly moving averages, which serve to smooth out monthly irregularities. Quarterly values of each index were also calculated directly, and it is hoped that in future the quarterly series will permit more disaggregation than is feasible on a monthly basis. For example, **permanent tsb** data shows that one in three first-time buyers purchase a 3 bed semi-detached home. Thus, a quarterly index for this type of home has been introduced. It is not yet clear

whether there is any significant seasonality in the series. When sufficient observations are available, standard seasonality tests will be applied, and a seasonally adjusted index introduced if necessary.

Rebasing the Index to 2003

Since its establishment the **permanent tsb** House Price Index has used annual weights from 1996. It is common practice for indices to be updated to use more current weights. As the time period of an index becomes longer the bias associated with weights from a fixed period increases as the fixed weights become more out of date. One reason for this is consumers shifting preferences in the face of changing prices. By their nature, fixed-weight indices do not take account of these changing patterns, sometimes called "substitution bias". For example, the Consumer Price Index is updated every five years to take account of new information from the CSO Household Budget Survey. It was decided to re-weight the index to take account of the merger of Irish Permanent plc and TSB Bank to form **permanent tsb** and to take account of any changes in buying patterns since 1996. Thus, it was decided to use 2003 as the new base period.

Analysis of the data shows that there have been some significant changes in the proportion of transactions accounted for by the different variables. The change in proportions is sufficient to justify a rebasing of the index. This is particularly evident in the proportion of first-time buyers and the proportion of properties with solid fuel heating. Regressions indicate that in 2003 the variables chosen to explain the price of a house in 1996 remain significant. Using the 2003 weights the index was rebased following the hedonic regression methodology that has been used to construct the index

to date. Despite the possibility of substitution bias in response to the rapid increase in house prices the rebasing had relatively little impact on the index itself. This is partly a reflection of the fact that the price of all different dwelling types has risen substantially since 1996, offsetting some of the substitution effect. To ensure consistency with previously published figures the two series have been linked and the index based to 2003 is now provided back to March 1996, the start of the index.

Conclusions

The hedonic index approach to measuring changes in house prices has proved its value in the UK for over a decade. It seems fair to claim that our results show that the approach can be successfully adapted for use with Irish data. The fit of the monthly regressions is reasonably high, and most of the coefficients for individual variables show a good degree of consistency. The indices themselves show only minor monthly fluctuations around their trends, and these can be largely overcome through the use of three-monthly moving averages. Thus the **permanent tsb** house price indices appear to provide reliable measures of underlying trends. They should prove valuable in monitoring the course of house prices, in understanding the relationships between different segments of the housing market, and in enabling future econometric studies of the causes of house price movements.

Author:

David Duffy, ESRI

Queries, contact **permanent tsb**
Marketing on 01 669 5646

**Produced
in
conjunction
with**



