INTRODUCTION

1. This report is being placed on the ODPM website to coincide with the launch, in September 2003, of ODPM's new, experimental, monthly house price index.

2. Section headings:
- Brief history of ODPM's house price index
- The House Price Working Group, origin and recommendations
- The construction of the new index

A BRIEF HISTORY OF THE ODPM HOUSE PRICE INDEX

3. ODPM (and its earlier incarnations DoE, DETR and DTLR) have published a quarterly house price index since 1968 based on data from the Survey of Mortgage Lenders (SML). The survey is conducted in partnership with the Council of Mortgage Lenders and has, for most of its thirty-five year history, involved a variety of mortgage lenders (both large and small) supplying, each month, a five per cent sample of their completions from the preceding month.

4. A five per cent sample of completions has always been too small to form the basis of a monthly index. However, advances in electronic data transfer has meant that many lenders now find it as easy to supply information about *all* their completions as for a five per cent sample. Consequently, over the past few years, the number of completions received each month has increased significantly – thereby making a monthly index a more realistic possibility.

5. An index based on SML data has its limitations because it is based solely on mortgage transactions - whereas nearly 25 percent of residential property transactions are *cash* purchases. However, the SML offers other benefits – namely that it also collects information about the buyer. e.g. whether a "first-time buyer", the age of the buyer(s), their total income and the size of advance. The only source of data on cash purchases is the Land Registry – and the reasons why the Land Registry dataset cannot yet be used as a basis for a timely, mix-adjusted house price index is discussed in paragraphs 13 and 14.

THE HOUSE PRICE WORKING GROUP

Origins

6. This was established in 2000 by the Office for National Statistics in response to concerns expressed by Treasury and the Bank of England that the existing monthly price indices (from the Halifax and Nationwide) often showed conflicting messages about what was happening to house prices. The Group's remit was to develop a reliable, official, monthly house price index by making better use of existing data sources. The steering group, which includes representatives from the Bank of England, Treasury, the ONS, Land Registry and ODPM, will continue to preside over further developments of the official house price index until a definitive house price index, including cash purchases, has been devised.

Main recommendations of the HPWG

7. These may be summarised as follows:

- There is a need for a reliable, monthly house price index that covers all housing and all transactions and is capable of providing the outputs and analysis required by users. At present none of the existing indices fully satisfy user needs, but there is scope to build on the current ODPM quarterly index.

- Current evidence supports the use of **completions data** in the construction of the index.

- The index should be **mix adjusted** to allow for the fact that different houses are sold in different periods.

- A **hedonic model** should be used to estimate arithmetic estimates of all cell prices.

- Weighting across the cells should be based on an **arithmetic approach with expenditure weights**, as used in the current ODPM index and in other price indices published by ONS such as the RPI.

- The **mix adjustment** should take into account the following **variables**:
  region
  location (within region)
  type of dwelling
  size of dwelling
  the distinction between new and old dwellings
  the difference between mortgage and cash transactions
  the difference between first time buyers and former owner occupiers

- The quality of the mix adjustment achieved by allowing for the recommended variables needs to be regularly assessed.

- Data on mortgage transactions alone may not provide a reliable indication of levels or movements for house prices in general; **cash transactions** may behave differently and need to be incorporated in any definitive house price index.

- The **weights** should be based on a rolling three year period updated annually.

- The index should be **chain-linked** and reset to 100 at the start of each year. The index should be **revised** as appropriate under clearly defined rules, consistent with other economic series.

- The index should be made available in **seasonally adjusted** form as soon as possible.

- **Buy-to-lets** may need to be included in the official house price index for some purposes. This will depend on whether their inclusion results in a significantly different series.

- The **Home Conditions Report** initiative (part of the proposed Home Information Pack) and the introduction of **electronic conveyancing** are significant developments which may impact on the source of the house price index in the future. These and other developments (e.g. a **National Property Database**) should be closely monitored to ensure that the needs of the house price index are met.

- Work on the Land Registry system should continue – on the need for a proxy **measure of house size** to improve the mix adjustment; on the **timeliness** of the data; and on transactions in **Scotland and Northern Ireland**.

CONSTRUCTION OF THE NEW HOUSE PRICE INDEX

8. The design of the new index has been largely driven by the above recommendations of the House Price Working Group. Ideally users would like to see an official index based on Land Registry data (and equivalent data sources for Scotland and N Ireland) alongside SML data (to provide information on borrowers). But this is not yet possible – as explained in detail below – so the new index being launched in September 2003 represents a pragmatic attempt to produce an index that best meets user needs within the current constraints of data availability and timeliness.

Choice of data source

9. Only two viable data sources were considered:

- **The SML dataset**
  with an expanded sample – thanks to some lenders switching from 5% to 100% of completions

- **The Land Registry dataset**
  containing all residential property purchases in England and Wales

10. The relative strengths and weaknesses of these two datasets compare as follows, with negative aspects of each dataset shown in bold and italics:

	SML	Land Registry
transactions	***sample of mortgages***	all purchases (incl cash)
coverage	UK	***England & Wales***
property type	yes	yes
dwelling size	yes	***no***
full postcode	yes	yes
timeliness	one month	***up to 3 months***

Transactions
11. In terms of data volumes the Land Registry dataset is the more comprehensive because it covers ALL completions (including cash purchases) whereas the SML dataset is based on a sample of mortgage completions only.

Geographical Coverage
12. Whilst the SML covers the whole of the UK, the Land Registry covers only England & Wales - and data from equivalent sources in Scotland and Northern Ireland is not yet available on a comparable basis.

Dwelling Size
13. The Land Registry does not collect information on the size of each dwelling sold* (e.g. number of rooms, number of bedrooms or square footage), yet this is a very important explanatory variable in determining reliable, mix-adjusted average house prices. *(* Note that whilst Land Registry requests information on dwelling size, conveyancers are not compelled to provide the data – so only five per cent of returns have this section completed)*

Timeliness
14. It can take up to three months for completions data to reach Land Registry.

Choice of dataset
15. Despite the obvious attraction of the Land Registry dataset, the expanded SML dataset was chosen because (a) it included "size of dwelling" and (b) the data can be collected and analysed more rapidly than the Land Registry dataset.

Exclusions from the index:

Cash purchases
16. The new index is based only on house purchases where the buyer(s) obtained a mortgage to help with the purchase. But nearly 25 per cent of all purchases are for cash – so the new index cannot be assumed to represent the changing level of all house prices. There are two reasons why an index based on cash purchases might move differently from one based on mortgage purchases.

a) Cash buyers may be able to purchase properties more cheaply than buyers with mortgages – because they don't suffer from the potential uncertainties and administrative delays of mortgage approval and processing. Though while this might be an advantage in a quiet market, it may be less so in a buoyant market.

b) The *types* of properties bought for cash differ from those bought with a mortgage – as do the *localities* in which they are bought. For instance the proportion of properties bought for cash is highest in Wales and the South West (28 to 29 per cent); and lowest in London (18 per cent). So the profile of cash purchases differs from mortgage purchases.

17. Remortgages are also excluded because the "price" would be a valuation - not a negotiated price;

18. Purchases by sitting tenants are excluded because they are probably buying at a discount to market value.

Inclusions in the index:

19. Purchases for owner-occupation;

20. Buy-to-lets (though BTLs will be omitted from special calculations e.g. for the Retail Prices Index).

AN INDEX BASED ON COMPLETIONS, NOT APPROVALS

21. The index is based only on completions. Whilst an index based on approvals would provide a more timely indicator of house price trends, the House Price Working Group decided that the new index should monitor completion prices for the following reasons:

- not all mortgage approvals go through to completion;
- occasionally the purchase price at the mortgage application stage has changed by the time completion goes through;
- it is appropriate that an official house price index should reflect the final transaction prices;
- within five years or so it may become possible to base the UK's official house price index on Land Registry data. Since the only prices recorded by Land Registry are *completion* prices then it makes sense for this new index to reflect completion prices – as has the old quarterly index dating back to 1968.

TIMELINESS OF THE NEW INDEX

22. The index will be published monthly, just over one month after the end of the reference period (e.g. the November index will be published early January). This is obviously not as timely as the release of the Halifax and Nationwide indices. This is because whilst Halifax and the Nationwide need only analyse in-house data, the SML involves the collection of data from up to fifty different lenders, all of which takes time. Once the new index is up and running we will investigate whether it might be possible to bring forward the publication date.

PARTICIPATION IN THE SURVEY OF MORTGAGE LENDERS

23. About fifty mortgage lenders participate in the SML. Some are supplying a five per cent sample of completions, others 100%. The lenders participate on a voluntary basis. This also applies to whether they choose to provide a five per cent sample of their completions or a 100% sample.

CALCULATING THE NEW INDEX

The main stages may be summarised as follows:

24. **Weights**. Once a year new weights are calculated based on the previous three years' transactions. Two sources of transactions data are used. For England and Wales, Land Registry data (mortgage purchases only) are used to determine weights down to the level of ACORN classification, new/old property, property type, local authority/county. SML data is then used to further split these weights by first-time buyer/former owner occupier and number of rooms/bedrooms. For Scotland and Northern Ireland there is no equivalent Land Registry data available, so only SML transactions data is used.

25. **Cells**. For every combination of main effects with a non-zero weight a matrix cell is created. For instance, at the start of 2003 a total of nearly 100,000 such cells were created.

26. **Hedonic model**. This is the most significant innovation in the methodology for the new index.

- The model is based on the following seven main effects**:**
  - location (local authority district)
  - cluster (an ONS classification of local authorities)
  - type of neighbourhood (ACORN)
  - dwelling type
  - number of rooms
  - old/new
  - first-time buyer/former owner occupier
- The chosen model is a main effects model including all the variables listed above with various other significant two-way interactions between the variables.
- The model uses $\log_{e(}$price) as the dependent variable. A simple transformation applied to the model estimates then ensures that estimated prices correspond to an arithmetic mean and not a geometric mean.

A technical description of the model will appear on the web at a later date

27. **Determination of average price per cell**. In the old index cell prices were determined as the arithmetic average of actual price quotes received for each cell: no modelling or price estimation was carried out. Consequently the number of cells in the matrix was only about 300 (instead of 100,000). This was because we had to ensure that we would receive enough price quotes every quarter to be able to determine an average price for each cell. In the new index it doesn't matter if we receive no prices for a cell – the model generates price estimates for all the cells anyway, whether there are sufficient numbers of actual price quotes for that cell or not.

28. The **benefits of the new model** are as follows:

- With nearly 100,000 cells instead of 300, the quality of the mix adjustment is considerably improved;
- Both partial and complete records can be used in the model (in the old index only *complete* records could be used);
- We no longer have to group the different values for a particular variable. For example, instead of the number of habitable rooms being collapsed into three groups (less than 5, 5 or 6, 7 or more) – the new model can handle the exact number of rooms.
- Greater flexibility, whereby if lenders don't collect number of habitable rooms but they do collect data on number of bedrooms - this can be used in the model instead.
- The inclusion of some additional explanatory variables (e.g. ACORN – which identifies, for each postcode, the most prevalent type of neighbourhood in that postcode area) will further improve the quality of the mix adjustment and of the overall index.

29. Once the monthly price estimates for all cells have been determined by the model, they are combined with their appropriate weights to produce that month's mix-adjusted average prices for all the required output categories.

30. The published house price indices will be temporal indices. We do not plan to publish any spatial series, though by publishing mix-adjusted average prices and the median prices by region, users will be able to compare house price levels across regions.

31. There will be complete coherence within each year. New weights applied at the beginning of each year will remain fixed until the start of the following year.

32. The relevance of the underlying hedonic model will be reviewed towards the end of each year and any modifications to the model that are deemed necessary following this review will be implemented at the start of the following year. However, the broad index methodology will be left unchanged for a number of years.

33. The new index is not being launched as a National Statistic. To qualify as a National Statistic the index and its underlying methodology will have to undergo a number of audit checks. This will happen during the autumn of 2003 – and it is hoped that the new index can be re-named the "National Statistics House Price Index" in early 2004.

RANGE OF MONTHLY INDICES AND MIX-ADJUSTED AVERAGE PRICES

34. We plan to publish monthly indices, the underlying mix-adjusted average prices and the year-on-year inflation rates for:

- all dwellings
- new dwellings
- old dwellings
- dwellings bought by first-time buyers
- dwellings bought by existing home-owners

for

- UK
- England and Wales
- Each of the four home countries
- UK ex-London
- UK ex-London and South East
- every Government Office Region

Rather than overburden the monthly statistical release with all the above tables, only a selection of tables will be included in the release. All the tables will be put up on the web site. The monthly release and web tables will be designed to be user-friendly and to maximise accessibility. Background information will be made available on the underlying methodology. In particular, in the longer term, we will undertake to determine and publish standard errors for the published indices.

Implications for the current quarterly house price index series

35. The quarterly report for Q2 2003 was the final report to be issued in the old format. Most other outputs (live web tables) will continue unchanged – but these will be reviewed. Once the new model is up and running we may well add to the range of web tables.

Reservations/caveats regarding the data and/or methodology

36. We remain dependent on lenders continuing to supply data on a voluntary basis.

ODPM, September 2003

CURRENT CONTACT POINT FOR FURTHER ASSISTANCE

Dorothy Anderson, Office of the Deputy Prime Minister

1/E2 Eland House, Bressenden Place, London SW1E 5DU

email: HPRICES.RSLS@odpm.gsi.gov.uk
phone: 020 7944 3306
fax: 020 7944 3279

ANNEX

The calculation of the monthly house price index for every aggregate area is carried out in several stages:

1. Determining the cells and weights

Every January the cells and weights for the next 13-month period (January this year to January next year) are derived from the numbers of mortgage transactions reported during the previous three calendar years. For England and Wales two data sources are used: Land Registry and the SML. For Scotland and Northern Ireland, SML data only is used.

A single cell is created for every combination of explanatory variables that has a non-zero weight. Note that both the numbers of cells and the weights for these cells can change from year to year.

2. Nomenclature

Assume that there are r aggregate regions for which a house price index is required. These might be basic Government Office Regions or combinations of regions such as “England & Wales” or “UK excluding London”.

The months from January in the current year to January in the following year are numbered 1 to 13.

The current year is identified as y.

The number of cells for region r in year y = n_{ry}

The average price for cell j in region r in month m of year y = p_{rmyj}

The weight for cell j in region r in year y = w_{ryj}

3. Calculating the mix-adjusted average price

Based on the above nomenclature, the mix-adjusted average price for region r in month m in year y can now be specified as:

$$P_{rmy} = \frac{\sum_{j=1}^{n_{ry}} w_{ryj} * p_{rmyj}}{\sum_{j=1}^{n_{ry}} w_{ryj}}$$

where

p_{rmy} = average price in region r in month m in year y
w_{ryj} = weight for cell j in region r in year y
p_{rmyj} = average price for cell j in region r in month m in year y

4. Calculating the index

The index for month m in year y (Jan of current year = 100.0) is then derived as:

$$100 * \frac{P_{rmy}}{P_{r1y}}$$

In other words the mix-adjusted average price for month m is divided by the equivalent average price at the start of the current 13-month period.

The above process gives a series of index numbers that start at 100.0 in January and end at some (usually higher) level the following January.

The whole process is then repeated from the following January.

5. Chain-linking

But this isn't much good as a publishable output. What users want to know is the rate at which house prices are rising year-on-year. This is where chain-linking comes in. Essentially all the separate annual indices (all starting at 100.0 in January) are linked together to form a single continuous series.