MEASURING HOUSE PRICE MOVEMENTS: METHODS, ISSUES AND SOME RECENT EXPERIENCE IN THE AUSTRALIAN CONTEXT

Alexa Olczyk and Geoff Neideck

Australian Bureau of Statistics

A Paper Presented at the Real Estate and Housing Indices Session of the 10th Meeting of the UN International Working Group on Price Indices (The Ottawa Group)

9-12 October 2007, Ottawa

DISCLAIMER

Views expressed in this paper are those of the authors and do not necessarily represent those of the Australian Bureau of Statistics. Where quoted or used, they should be attributed clearly to the authors. *The authors wish to thank the following for their inputs: Merry Branson, Mark King and Ruel Abello.*

For further information, please contact Alexa Olczyk – *telephone*: 02 6252 5854 or *e-mail*: alexandra.olczyk@abs.gov.au

MEASURING HOUSE PRICE MOVEMENTS: METHODS, ISSUES AND SOME RECENT EXPERIENCE IN THE AUSTRALIAN CONTEXT

Alexa Olczyk and Geoff Neideck

Australian Bureau of Statistics

ABSTRACT

In Australia, as in many other countries, movements in house prices have significant macroeconomic impacts. However, measuring house prices accurately over time is not simple or straightforward. It involves complex conceptual, methodological and data issues, and continues to be a challenging area for statistical agencies. Unlike other price indexes, house price indexes are inherently difficult to measure accurately. Houses are sold infrequently, the composition of houses transacted in the market changes over time and the quality of houses (affected by depreciation as well as renovation) is not constant.

There are numerous methods for constructing house price indexes, each with their own advantages and disadvantages. Some of these methods are based on simple summary measures, such as the median price of houses transacted in a particular period. The advantage of these methods lies in their relative simplicity, both in terms of computation and the interpretation of results. However, such simple measures are likely to suffer from compositional and quality problems. Recent advances in more sophisticated methods (such as hedonic regression and repeat sales methods) have enabled price statisticians to adequately account for compositional and quality changes. However, these methods are not without limitations. Most notably, they are data intensive and their practical use is limited by the lack of comprehensive data on housing characteristics.

A practical approach to house price index construction applied to the Australian case by the Australian Bureau of Statistics (ABS) is stratification at the suburb level. This approach involves grouping suburbs together into strata, such that the homogeneity within groups and heterogeneity between groups are maximised. The effectiveness of this approach is affected by the degree of stratification possible, which rests on the amount of dwelling characteristic detail available and the volume of market activity. Comprehensive dwelling characteristic data does not necessarily lead to a feasible fine level stratification if there are insufficient observations per stratum to produce reliable stratum averages.

The aim of this paper is to describe the Australian experience in house price index construction, with a particular focus on the ABS approaches to constructing the established house price index over time.

TABLE OF CONTENTS

1. Introduction

This paper describes Australia's experience in house price index (HPI) construction, with a focus on the methodologies employed by the Australian Bureau of Statistics (ABS). It describes the inadequacies in previous methodologies and how these are being addressed presently through an improved stratification methodology. The importance of stratification in house price index construction cannot be understated because the infrequent turnover of individual houses means that the market prices of each property in the Australian housing stock are not simultaneously observable. The next best alternative is to compare the price changes of houses that are similar from quarter to quarter, where similarity is determined by a set of attributes that strongly relate to prices.

The ABS HPI was originally designed to meet the specific data requirements for the construction of a price measure for mortgage interest charges, which were included in the Consumer Price Index (CPI) from 1986 to 1998. When mortgage interest was removed from the CPI in 1998 as part of the 13th Series CPI Review, the ABS continued to publish the price index of established houses because of user interest in the series.

1.1 Background

The need to address the problems associated with the HPI method stemmed from a review that motivated the ABS to examine the HPI's reliability and timeliness, especially during a period in which public attention was focused on house prices. The surge in house prices from late 1999 until late 2003 stimulated the interest of policy makers and the media in the HPI's ability to accurately measure Australian house price inflation. Issues such as housing affordability, interest rates and growing private debt had increased the pressure for more accurate and timely measures of house price movements [1]. Such interest could easily be understood since housing represented around 65 per cent of household wealth and around 85 per cent of household borrowing in Australia. Property-related revenues (mostly from residential property) accounted for over 35 per cent of total state government own-revenues. The Reserve Bank of Australia, which relies heavily on accurate price information to inform its monetary policy, recognised the macroeconomic importance of housing and supported a strong public policy case for improved data in this area.

1. Since early 2004 various commentators criticised the timeliness and reliability of measures of house prices published by various organisations, including the ABS. There were several different measures of house prices published regularly, each based on different methodologies, scope, coverage and timing. Users were primarily concerned about the conflicting signals that the different measures were delivering and, in the case of the ABS series, its lack of timeliness (released about 9 weeks after the reference quarter) and the point in the transaction cycle at which it recorded prices (date of final settlement rather than contract exchange date, when contract prices are struck - a timing difference of approximately 2 months).

1.2 Outline of this paper

The paper starts with a short description of the conceptual and practical issues at play when measuring house price changes (section 2), followed by a description of the methods generally used to construct house price indexes (section 3). It then describes the approaches used in the past by the ABS (section 4) as well as examining the current approach and recent developments in house price index (HPI) methodology (section 5). Section 6 describes the way forward and outlines future improvements to the ABS house price index. The final section provides some concluding remarks.

2. Conceptual and practical issues in measuring house price changes

2.1 Heterogeneity and infrequent turnover

The construction of a house price index is inherently more cumbersome than the construction of a price index based on a fixed basket of goods, such as the consumer price index (CPI). The consumer price index can be constructed by taking a representative basket of goods and services and pricing this basket every period. This approach is not feasible for a house price index for a number of reasons.

First, no two houses are absolutely identical. Houses are heterogeneous goods with differing physical characteristics and varied locations. Second, the number of transactions (house sales) is low relative to the housing stock. Third, this infrequent turnover means that instead of being able to observe the price of each house at every point in time the price of a house is revealed only when the house is sold. Thus, even if a representative sample of houses could be assembled, very few of them would transact in any period. A potential solution to this problem would be to use price data based on the independent valuation of each house. However, use of price measures other than actual transaction values leads to subjectivity issues and is a less than ideal solution.

2.2 Compositional change

Since only a fraction of the total stock of housing is traded in any period, HPI measures based on transaction prices are dependent on the mix of properties sold in each period. Given the heterogeneous nature of the housing stock, changes in the average price of property sales from one period to the next will reflect changes in the types of properties transacting, as well as pure price changes. For example, if transactions in one period consist mostly of larger houses or houses in more expensive suburbs, average transaction prices will appear to have increased, though the price for a given type of house may not have changed. This phenomenon is known as compositional change. Thus, although the standard procedure for constructing price indexes is to select a sample of representative items and to reprice

them through time, this approach is not possible in the case of established houses as the observable prices in each time period invariably relate to a different set of dwellings.

2.3 Quality change

To complicate matters further, new houses are constantly built and old houses are sometimes demolished - accordingly, the population of houses changes over time. Housing quality also changes as the existing housing stock is improved through renovations and extensions or is deteriorated by neglect and depreciation. These changes are reflected in house prices, although they are really quality changes, not pure price changes. This issue of quality change may not be so much of a problem when the aim of the HPI is to measure changes in the value of housing stock - in Australia, land accounts for a large part of the established house price and land does not depreciate.

2.4 Timeliness

Another important issue in HPI construction is the lack of timely data, a product of various institutional factors in the housing market. In the Australian context there are four significant dates related to the purchase of a house. The sale progresses in stages, first a verbal agreement is reached to purchase at a negotiated price, second mortgage financing is approved, third the contract is exchanged and finally settlement of the sale occurs. When collecting data on house price sales, it is desirable to choose the earliest date at which the purchase price is set - when verbal agreement is reached. However, measures of housing prices typically use data from land titles offices, which record transactions some time after settlement.

Settlement, recording and processing lags mean that at any point in time the available housing price information relates to prices that were determined some months earlier and that data relating to contracts signed in any particular period are spread out over time. Problems associated with compositional change can be exacerbated by problems of data timeliness if there is a systematic lag between the time when particular sales are agreed to and when they are recorded in a database of transactions. The issue arises when data relating to property sales in a particular period are distributed in a biased way - for example, properties with higher prices often take longer to settle. The implication is that early prices recorded by land titles offices relating to the property sales in a particular quarter are relatively low and it is necessary to obtain an almost complete data set for each quarter before it is possible to calculate an accurate measure of the house prices for the quarter.

3. Methods for constructing house price indices

House price indices can be constructed using a number of different methods. The simplest approach is to use a summary measure (such as a mean or median) of all transactions each period for the region of interest. An index based on the simple average of transaction prices has a clear interpretation - it measures the average price of houses sold in that period. However, if interest lies in measuring the rate of change in the price of the housing stock or the price of representative house, the data must be composition and quality adjusted - the mix of houses sold each period is unlikely to be a consistently reliable indicator of the mix of houses in the housing stock. The summary measure method can be improved by calculating summary price measures for smaller geographic regions to partly account for the heterogeneity caused by locational differences. However, this does not account for differences in the sample of houses transacted - changes in average prices over time may reflect changes in the mix of houses being sold rather than in the value of the stock of dwellings or the price of a typical house. The central issue in HPI construction is thus how to utilise prices for an essentially heterogeneous set of dwellings to construct measures of price change for homogeneous dwellings. There are three general approaches that may be used: stratification, hedonics and repeat sales methods.

3.1 Stratification

The most commonly used method for compositional adjustment is the simple weighted average method, sometimes called mix-adjustment or stratification. The stratification approach involves stratifying the transaction prices in such a way as to group the 'most similar' dwellings. Generally, stratification groups objects according to a set of defining characteristics, such that objects in each group are relatively closer to each other than to objects belonging to other groups. The objective is to minimise the heterogeneity of dwellings within each stratum. In each period a summary price measure is calculated for each stratum and used to construct a stratum level price index. The aggregate index is then calculated by weighting together the individual strata indexes with weights representing the relative significance of the stock of dwellings in each stratum (in value terms).

The effectiveness of stratification is limited by the degree of stratification possible. This depends on the amount of dwelling characteristic detail available and the volume of market activity. Not only is data on dwelling characteristic hard to come by, some major price determining characteristics (e.g. a view of the water) are difficult, if not impossible, to measure. Even with comprehensive dwelling characteristic data it may not be feasible to employ fine level stratification if there are insufficient observations per stratum to produce reliable stratum averages.

3.2 Hedonic methods

One approach to overcome this limitation of the stratification method is to increase the number of characteristics used to describe a house price observation. The outcome of this is a method that resembles a hedonic model. Indeed, stratification and hedonic regression can give very similar results if they adjust for the same property characteristics. The idea behind the hedonic model is that the price of a house can be accurately estimated from its price determining characteristics. A hedonic model is one where the independent variables relate to the quality of the product and hedonic regression is a way of estimating the value that the market places on each house characteristic.

The effectiveness of hedonic modelling is critically dependent on the availability of dwelling characteristic data (at least for those characteristics that have a significant influence on price) and the estimation of the underlying hedonic model itself. The detail on house characteristics required to support the hedonic approach is generally much greater than that required for the stratification approach. Hedonic regression relies heavily on the correct specification of the functional form of the model and the set of property characteristics. Like the stratification method, the hedonic method requires the inclusion of an appropriate set of house attributes in the modelling process. Failure to account for these attributes can lead to inconsistent estimates of the implicit prices of characteristics in hedonic methods and the inefficient grouping of observations in stratification. The adoption of hedonic methods requires considerable data collection effort and therein lays its most serious practical drawback and the reason why a repeat sales regression method can often be seen as superior - as previously mentioned, many characteristics that have an impact on house prices cannot be measured, e.g. view, ambience, aesthetics, neighbours.

3.3 Repeat sales method

Repeat sales methodologies focus on price changes rather than on prices themselves. This approach controls for compositional change by maintaining a longitudinal data set of properties and, when properties are sold more than once, calculating the price changes between successive sale dates. By using only repeat sales properties all contributing factors to variation in price growth are controlled for. However, it is important to note that there is no guarantee that a house sold in one period is in the same condition in a later period. It may have depreciated or been renovated between sales and if repeat-sale houses differ from single-sale houses in the sample, the index may be biased. Sample selection bias could also arise, caused by differences in the characteristics of houses that sell more often than once and those that sell only once - that is, repeat sale houses might not be representative of the housing stock.

To be effective, the repeat sales approach requires a long time series of price data for individual properties. This method does not maximise the use of available data since houses sold only once during the time interval of interest are omitted and the method violates temporal fixity (i.e. the price index must be revised retrospectively when new data becomes available). Revision is an undesirable characteristic for official price indexes. An advantage of the repeat sales approach is its less strict data requirement, with the method relying only on data relating to transaction prices and dates of subsequent transactions.

4. ABS' past experience in constructing house price indices

Since 1986, the Australian Bureau of Statistics has compiled an established house price index for each of Australia's eight capital cities as well as an index at the national aggregate level (a weighted average of the city level indexes). The HPI is compiled and published quarterly in *House Price Indexes, Eight Capital Cities* (Cat. No. 6416.0).

The method for constructing the HPI has developed over time, but has always applied some form of stratification. The original approach was based on geographically stratified medians and evolved into a stratification method based on geography and other house price determining characteristics, with a foray into hedonic methods along the way. The ABS approach to HPI construction is based on measuring price movements, looking at changes in price levels over time. There are in effect two issues affecting the ABS HPI: how to measure price changes and how to weight these changes. Generally speaking, weights are based on the total housing stock and the index is constructed by measuring changes in median prices. Median prices are used because they are considered to be the best measure of central tendency and are less affected by outliers than other summary measures, such as means.

4.1 Use of median price

Originally, the ABS-constructed HPI was based on median measure of price, with a partial attempt to control for compositional change made by stratifying houses within each capital city by region. Depending on the size of the city, transactions were stratified into several geographic regions (the SSD) [2]. There were seven regions each for Sydney, Brisbane and Perth, six regions for Melbourne, five regions for Adelaide, four regions for Hobart and three regions each for Darwin and Canberra. Median price measures were calculated for each of the regions and then these region indexes were

2. A statistical subdivision (SSD) is a general purpose spatial unit of intermediate size between the statistical local area (SLA) and the statistical division (SD) in the Australian Standard Geographical Classification (ASGC). SSDs consist of one or more SLAs. In aggregate, they cover Australia without gaps or overlaps. SSDs are defined as socially and economically homogeneous regions characterised by identifiable links between the inhabitants. Moreover, in the non-urban areas (i.e. outside the capital cities or areas with population clusters of 25,000 or more people), an SSD is characterised by identifiable links between the economic units within the region, under the unifying influence of one or more major towns or cities (ABS 2005).

weighted up to the city level (to form a capital city price index). A national house price index was calculated by aggregating the eight capital city price indexes.

Weights are important components in calculating price indexes, both at the city level and in the construction of the national HPI. The region weights used to derive capital city indexes were estimated based on the information extracted from the Population Census. The estimation was based on the distribution of housing stock among the regions, with representative weights based on the number of established houses and their average value.

The indexes for each of the eight capital cities were arithmetically weighted together to form a national HPI. The weights were estimated based on the value of secured (individual) finance commitments for the purchases of newly erected and established houses, with data sourced from the ABS Survey of Housing Finance for Owner Occupation.

For Sydney, Melbourne, Brisbane and Adelaide, a 'trimean' method was used in the calculation of prices in the region indexes. Broadly speaking, this method involved the following steps:

- in each region, “extreme” or “unrepresentative” prices, which were either suspiciously low or unreasonably high, were removed from the sample;
- prices were divided into 3 quantiles, representing low, medium and high valued houses;
- an unweighted average price was calculated for each quantile;
- a weighted arithmetic mean was used to aggregate the three quantile prices to form a regional price index. The price for the medium quantile was weighted by 0.5 and the prices for the low and high quantiles were weighted by 0.25 each;

A simpler method was used in the calculation of region price indexes for Perth, Hobart, Darwin and Canberra. For these cities, after removing the “extreme” or “unrepresentative” values, an unweighted average price was calculated for each region.

By stratifying sales based on broad geographic region, the ABS made an attempt to partly control for compositional change in the type of properties being sold. However, this was only a limited solution, which did not fully control for compositional change or changes in the quality of houses sold in different periods.

4.2 Investigation into hedonic methods

The surge in Australian house prices beginning in late 1999 engaged the close attention of policy makers and commentators on the HPI and its ability to indicate (or even predict) changes in economic activity. This heightened interest in the

methodology for HPI construction, in particular the difficulty in controlling for quality and composition changes, led the ABS to experiment with hedonic methods.

In 2003, the ABS undertook a study to investigate the feasibility of applying hedonic methods to the construction of the HPI. The main motivation for the study was to test the efficacy of hedonic methods to account for the impact of housing attributes on the HPI. The analysis was based on data from Hobart and Adelaide only, because the information required for hedonic methods was not available for other cities.

Three broad types of housing attributes [3] were considered in the hedonic investigation:

- *Structural* (S), which refers to characteristics such as size of the house or land, number of rooms, construction materials used, age of house, etc.;
- *Locational* (L), such as distances of the house to the CBD, local shopping centre, hospital or schools etc.; and
- *Neighbourhood* (N), encompassing characteristics related to social and environmental conditions.

A general hedonic function of the form:

$$P = f(S, L, N)$$

was adopted, where P represents the prices of individual houses.

The partial derivative of the above hedonic function with respect to any one of the attributes can be interpreted as the marginal price of the attribute, *ceteris paribus*.

The data for this investigation, sourced from the South Australian Valuer-General's Office and the Tasmanian Department of Primary Industries, Water and Environment (DPIWE) contained some information on structural attributes (S) of individual houses, but did not have any information on neighbourhood attributes (N) or locational attributes (L), other than the address details. To fill the information gap, two additional sets of data were obtained from sources within the ABS. The first set included information on locational attributes (L), measured as the distance of specific streets to places such as the CBD, nearest school, shopping centre or hospital, etc. The second set, sourced from the Australian Population Census, contained information on neighbourhood attributes (N). This included a Census-based indicator of an area's socio-economic standing, the Socio-economic Index for Areas (SEIFA). The SEIFA is based on a principal components analysis of a wide range of socio-economic variables representing a neighborhood's socio-economic status.

3. In this paper, "attributes", "characteristics" and "qualities" are used interchangeably and they all represent factors that influence house prices.

Results from the hedonic regressions were consistent with *a priori* expectations, i.e. that the price of a house would be higher if the:

- block of land was larger;
- floor area of the house was larger;
- house was newer;
- house was in better condition;
- house had brick walls;
- house had more than one bathroom;
- house was closer to the CBD; and
- house was located in a better socio-economic environment (as measured by the SEIFA).

Unfortunately, firm conclusions as to the usefulness of hedonic methods in HPI construction could not be drawn due to limitations caused by the data. In particular, the following issues were not overcome:

- only data from Hobart and Adelaide was available. Different results might be obtained when hedonic methods are applied to large cities (such as Sydney and Melbourne) where the situation (in terms of the housing market, socio-economic conditions and consumer preferences) is more complex.
- the data only covered a short period of time and this was a period when housing markets were booming everywhere. The results could be unique to this stage of the housing market cycle.
- a hedonic function depends on the availability of well-defined attribute variables (i.e. S, L and N). The attribute variables used in estimating the hedonic functions could be improved in terms of coverage and definition.

5. The current approach to and new developments in house price index construction

In 2004 the ABS initiated a review of the HPI, following a recommendation in the 2003 Reserve Bank of Australia Annual Report. The RBA advised that the ABS should give "...give priority to developing an efficient, accurate and timely method for the collection of data relating to levels and movements in prices of residential real estate in Australia." [4] The RBA collects evidence on house prices from a number of sources and noted that all of these sources had a 'significant lag' in information reporting. The result was house price data that referred to the growth rate in the previous quarter

4. The Parliament of the Commonwealth of Australia (2004), p.g. 18.

and covered settlements based on contracts that had been entered into up to three months earlier, leading to a total delay of possibly six months.

The review was intended to produce a better index by improving its capability to predict turning points through exploring alternative data sources and improving compilation methods, such as the treatment of outliers, geographic stratification, etc. [5] The review focussed on determining the best means of stratification given available data sources and looked into data sources that would lead to improved timeliness. The outcome was a change in the way that the HPI is constructed, applying a structural, locational and neighbourhood based stratification, as well as supplementing the data by incorporating loan approval data from mortgage lenders to improve the timeliness of index. This enabled the index to be published 5 weeks after the reference quarter instead of the previous 9 week lag. The index was also made more timely through the use of contract date rather than settlement date to determine the date of the transaction - aligning it more closely with the point at which price was determined, rather than when the purchaser paid.

5.1 S, L, N stratification

The aim of the new stratification was to group houses that were similar according to their attributes, so that the level of homogeneity between houses contained in the same stratum was maximised and there was a sufficient number of sales observations in each stratum in each reference quarter to construct robust price relatives. This involved a trade-off, with the level of homogeneity limited by having a sufficiently large number of sales observations to calculate price estimates. Individual houses could be grouped together on the basis of attributes to enable a comparison of like against like, however the absence of adequate unit-level information, in addition to the computational burden, made this approach unfeasible. A compromise was reached to treat larger areas as the building blocks for the stratification - there was sufficient information available to apply a suburb level stratification. [6]

It was decided that stratification should be based on structural, locational and neighbourhood characteristics. Analysis results determined that four structural variables, four locational variables and one neighbourhood variable were the most relevant. The structural characteristics of each suburb were derived at the household level from the Census of Population and Housing. This information was summarised at the suburb level to create four variables capturing the general structural attributes of dwellings within each suburb. These were:

---

5 The ABS indices were typically released about nine weeks after the end of the reference quarter, with this delayed timing designed to ensure that the indices captured most of the sales settled in that quarter. The index based on mortgage lenders' and land title office data is currently released five weeks after the reference quarter.

6. This involved the implicit assumption that houses within a suburb were homogeneous.

- the percentage of houses in the suburb with two or fewer bedrooms;
- the percentage of houses with four or more bedrooms;
- the percentage of houses that are owner-occupied; and
- the percentage of dwellings in the suburb that are houses.

Locational variables were obtained from data sets created by the Geography Section of the ABS and included the distances of the midpoint of any given street to the:

- central business district (CBD);
- nearest shops;
- nearest hospital; and
- nearest emergency services.

As with structural variables, data at the street level was used to calculate the average distances of streets within suburbs to derive average distances for suburbs.

The Socioeconomic Index for Areas (SEIFA) was used to represent neighbourhood characteristics [7]. This ABS-produced index is a continuum used to indicate the socioeconomic status of areas, with low values indicating disadvantage and high values indicating areas of advantage. The index takes into account variables such as the proportion of families with high incomes, people with a tertiary education, and employees in skilled occupations (see Information Paper: Census of Population and Housing - Socio-Economic Indexes for Areas (Cat. No. 2039.0)).

Nine variables were available for the clustering process, but the ability to visualise the solution post hoc to assess its plausibility was seen as important. Therefore, principal components analysis was used to reduce the eight structural and locational variables into two principal components, one each for the structural variables and the locational variables. The stratification was undertaken at the SSD level in the hope that sub-city level indexes could be constructed in the medium term and to preserve the geographical homogeneity of suburbs. The implication of this constraint was that only suburbs within the same SSD were permitted to be grouped into the same cluster.

The S, L, N stratification based HPI was compiled using weights relating to the stock of established houses. The weights were expressed in terms of stock values (currently derived from the 2001 Census of Population and Housing) - with an initial value of the established housing stock in each cluster estimated by aggregating suburb counts to clusters and valuing them at March quarter 2002 mean prices. The new stratification improved the quality of the HPI by improving the homogeneity of houses whose prices were compared from quarter to quarter. This finding was evidenced by the

7. In particular, a standardised Index of Relative Socioeconomic Advantage/Disadvantage was applied in the stratification.

strata having significantly different median price levels, which reflected significant between-group variation in price levels. In general, the current stratification, seeking to isolate various homogenous sub-populations in the housing stock, led to more regular distribution properties at the stratum level than at the SSD level (on which the regions for the simple median method were based).

5.2 Use of mortgage lenders' loan approval data

As noted previously, it takes several months for all transactions relating to a particular quarter to be finally settled, recorded by the relevant land titles office and then passed on to the ABS. Clearly, it is not possible to produce a timely HPI from this administrative data source, although it is the preferred source as far as coverage and comprehensiveness are concerned. However, based on investigations into data from mortgage lenders (banks, etc.) the ABS concluded that loan approval documents could be used to produce preliminary estimates for the two most recent quarters. As a large percentage of house sales involve mortgages, loan approval documents created by mortgage lenders were identified as a source of timely house price data. Although the data do not cover all house sales, they have sufficient coverage for them to be used as a means of estimating the movements in the established house prices for the latest two quarters.

The ABS concluded that the best approach would be to use two different data sets to compile the HPI. The first is to use the prices from the land title offices to compile the price indexes up to the point for which a complete set of data can be obtained on an exchange date basis. Substantially complete sets of data are available for all cities up to the quarter ending two quarters prior to the most recent quarter. The series based on this data is referred to as the benchmark series. The second approach uses the mortgage lenders' data combined with early land title office data to project the HPI for the two most recent quarters to provide a more timely indicator of changes in house prices. As the full set of land title office benchmark data becomes available, it is used to replace the leading indicator component. As a result, estimates of the HPI for the two most recent quarters are preliminary and subject to revision.

5.3 New developments - price based stratification

In 2006 the Reserve Bank of Australia (RBA) proposed a simple measure of house price growth [8] that addressed the problem of compositional change by stratifying individual transactions into different groups. It is reasonable that stratum boundaries should be defined using information on all relevant variables that influence the characteristic being measured, with the most useful variables for stratifying the data being those that are highly correlated with the variable of interest. For this reason, the

8. Prasad and Richards (2006)

RBA suggested grouping small geographic regions (suburbs) into strata based on the variable most likely (on an *a priori* basis) to explain the price in any transaction - the long-term level of prices for the suburb in which the house is located. This methodology produced a measure of price growth that substantially improved on median prices and one that was highly correlated with more sophisticated (but more computationally intensive) measures.

The RBA analysis was based on data from Australian Property Monitors. The data set contained virtually the entire population of housing transactions that occurred over the period of interest - data from land title offices as well as auction price data. The RBA method involved the following steps:

- suburbs were ranked according to their median transaction prices over the 2000-2004 period;
- based on these rankings, suburbs were grouped into strata; suburbs in the five largest cities were divided into ten groups (deciles) with an approximately equal number of suburbs in each group, while houses in Canberra and apartments in Sydney and Melbourne were divided into 5 groups (quintiles);
- a median price was calculated for each stratum for each quarter, then changes in median prices from each stratum were weighted together to calculate the growth in city-wide prices.

The price-based stratification was assessed by examining how well it addressed the problems with conventional unstratified median measures and whether it outperformed the change in the *X-12* seasonally adjusted median price. The price based stratification was also compared with regression-based techniques on a correlation basis. The price based result was a considerable improvement on the simple median approach and performed better in real time with limited data samples. The growth rates produced by the price stratification also lined up closely with more advanced hedonic measures.

The RBA also considered the extent to which price movements in a particular segment of a capital city market, as proxied by the median price of each quantile, were correlated with price movements of other market segments. A reasonable amount of co-movement in housing price growth across different cities was found, suggesting the existence of a national housing cycle. Interestingly, within-city correlations tended to be even higher - suggesting the existence of significant regional effects. The results showed a reasonably high degree of correlation in house price movements in cases where there was a higher degree of similarity between two market segments in terms of location, dwelling type and 'economic background'. This provides evidence of socioeconomic factors on a national level that have different impacts on price growth in higher- and lower-priced suburbs.

6. Further improving the ABS established house price index

The ABS first published the HPI based on the S, L, N stratification in December 2005, with the series backdated to March quarter 2002. An assessment of the method's strengths and practical limitations has now been made and several key issues requiring investigation have been uncovered. The price based stratification introduced by the RBA has also paved the way for an investigation into potential improvements to the current stratification method through the use of long term median price as a stratification variable.

6.1 Limitations of the current S, L, N stratification

The HPI compilation process has brought to light a problem concerning the number of clusters per capital city. In some cities, there seem to be too few clusters and the sheer number of observations leads to large value ranges, affecting the volatility of the overall house price index. In other cities, the stratification suffers from too many clusters and as a result of this, some clusters have too few sales observations in some quarters and are sensitive to outliers.

6.2 Potential for further improvements in stratification

The current stratification is restricted by an SSD level constraint so only suburbs within the same SSD can be grouped into the same cluster. It has been hypothesised that by relaxing this constraint and allowing stratification across SSD boundaries the number of clusters will decrease and each cluster will contain a relatively broader group of more similar suburbs. This will lead to a sufficient number of sales observations per stratum per quarter and result in stratum medians that are less sensitive to outliers.

In order to improve the HPI further, the research carried out by the RBA will be taken into account. The possibility of including the long-term median price of a suburb in the stratification will be explored and its impact on the HPI will be examined.

The key questions to be answered include:

- what will be the impacts of removing the SSD constraint from the stratification?
- will the stratification be improved by incorporating the long-term median price?
- does including long term median price as a clustering variable and/or removing the SSD restriction lead to a significant change in the size and number of strata? does the homogeneity of suburbs within a stratum increase?
- what are the implications of including price as a stratification variable on HPI movements over time? Compare and contrast to other methods - past (median

price), current (S, L, N stratification) and other (RBA approach of long term median price stratification).

In light of the above and other developments, the current ABS research aims to improve the HPI by including the long-term median house price of a suburb in the stratification process and removing the SSD constraint.

6.3 Other plans for the HPI

At present the scope of the HPI is limited to detached houses located in Australia's capital cities. However, key users of the HPI have indicated interest in an expanded scope. Of priority is extending the coverage of dwellings to include apartments, units and townhouses. The ABS plans to investigate the feasibility of this extension in the near future. Extension of the coverage of the HPI to regional cities and rural areas is also of interest, but of a lower priority.

7. Concluding remarks

Research from both the ABS and RBA in house price measurement shows the results from stratification line up quite closely with more advanced regression-based measures of price growth. Overall, this indicates that it is possible to develop computationally simple estimates of price growth that control for a significant proportion of compositional change. Australian research indicates that the vast majority of the explanatory power in standard hedonic regressions comes from the location of the dwelling (suburb) along with some indicator of social and economic well-being in the suburb.

One clear advantage of stratification approaches is their relative simplicity. However, more sophisticated approaches like hedonics and repeat-sales are possible. These approaches are not without shortcomings, for example, hedonic regressions will only be as good as the data on housing characteristics that are available, and repeat-sales estimates are likely to have significant problems in real time and can be subject to nontrivial revisions. Stratification approaches produce a measure of price growth that substantially improves upon standard unstratified city-wide median measures. In particular, when compared with a city-wide median measure, stratification measures of price growth are considerably less volatile, are not subject to seasonality and perform better in real time with limited data samples.

8. References

Australian Bureau of Statistics (2005) *Australian Standard Geographical Classification*, cat. no. 1216.0.

Australian Bureau of Statistics (2006) *House Price Indexes, Eight Capital Cities*, cat. no. 6416.0.

Branson, M. (2006) ***The Australian Experience in Developing an Established House Price Index***. OECD-IMF Workshop on Real Estate Price Indexes, Paris, 6-7 November 2006.

Chen, L., Zhao, S., Romanis, P. And Lim, P.P. (2004) *Exploring Hedonic Methods for Constructing a House Price Index*. Australian Bureau of Statistics Research Paper (Methodology Advisory Committee), cat. no. 1352.0.55.067.

Gatzlaff, D.H and Ling, D.C. (1994) "Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies", *Journal of Urban Economics*, 35:221-244.

Hansen, J. (May 2006) *Australian House Prices: A Comparison of Hedonic and Repeat-sales Measures*. Reserve Bank of Australia Research Discussion Paper 2006-03.

McCarthy, P., Branson, M. and King, M. (2005) *Renovating the Established House Price Index.* Australian Bureau of Statistics Information Paper, cat. no. 6417.0.

Prasad, N. and Richards, A. (2006) *Measuring Housing Price Growth: Using Stratification to Improve Median-based Measures.* Reserve Bank of Australia Research Discussion Paper 2006-04.

Reserve Bank of Australia (2003) *Annual Report,* available at www.rba.gov.au under Publications & Research.

The Parliament of the Commonwealth of Australia (2004) *Review of the Reserve Bank of Australia Annual Report 2003.* House of Representatives Standing Committee on Economics, Finance and Public Administration, Canberra, August 2004.