Use of complementary administrative data sets for Australian house price indexes

Dr. Susan Kluth
Australian Bureau of Statistics

This paper describes the Australian experience in house price index construction, with a particular focus on the use of land office register data to construct a price index for established houses. Supplementary data from mortgage lenders are used to increase the timeliness of the index. To account for compositional change, a stratification approach based on location is used, such that the homogeneity within groups and heterogeneity between groups are maximised.

Keywords: *house price index, stratification, compositional change, administrative data*

Background

In Australia, housing represents around 65 per cent of household wealth and around 85 per cent of household borrowing (*Household Wealth and Wealth Distribution, Australia 2005-2006*, ABS cat. no. 6554.0). Hence, timely measures of movements in the prices of Australian houses are, as in many other countries, of great interest to policy-makers, those involved in the housing industry and households. However, accurately measuring the changes in house prices over time is not straightforward. Unlike other price indexes, house price indexes are inherently difficult to construct. This is because standard price index methodology - selecting a sample of representative items and re-pricing them through time - does not apply in the case of houses. It is also extremely difficult to adjust for changes in quality of houses. Thus, measuring the change in house prices continues to be a challenging area for statistical agencies.

Since 1986, the Australian Bureau of Statistics (ABS) has compiled an established house price index (HPI) 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* (ABS cat. no. 6416.0). The method for constructing the HPI by the ABS has developed over time, although it has always incorporated some form of stratification.

The HPI is produced using two complementary administrative data sets. One data set, supplied by the State / Territory Land Titles Office or Valuers-General (VGs) Office[1] contains all house purchase transactions but a considerable time lag is associated with its availability. The second data set, based on loan approval documents from mortgage lenders is timely but not exhaustive in its coverage. It also contains data which is never converted into house purchases. This paper discusses the approach taken to combine the data from these two sources to produce a reliable HPI, and the stratification method applied.

---

[1] Data from these sources are referred to as VGs data in the remainder of this paper

Difficulties in measuring house price movements

In an ideal world, an HPI would be constructed by reference to the current and historical market prices of the entire stock of residential dwellings. In practice, market prices for any particular period are only available for those dwellings that are actually traded (sold/purchased) in the period. Such sales account for only a very small proportion of the total housing stock in any quarter. It is therefore necessary to draw inferences about the price behaviour of the whole stock from these small samples. The assumption behind this procedure is that the median sales price of the houses traded each quarter is indicative of the median price of all houses.

The standard procedure for constructing price indexes is to select a sample of representative items and to re-price the identical items through time. This approach is not viable in the case of established houses as no two houses can be considered identical, so the observable prices in each period invariably relate to a different set of dwellings. In Australia the average turnover of the housing stock is around 6 per cent per year, or just 1½ per cent per quarter, and in other countries the turnover rate of the housing stock is often significantly lower.

Given that the sample of transactions in any period may not be representative of the entire housing stock, if using a simple median or mean price measures, it will be difficult to distinguish in real time between true movements in the price of houses, from spurious movements due to compositional effects. For example, if the proportion of expensive houses sold increased between two periods, but prices remained unchanged, a false price increase would be observed.

The ABS has considered a number of different approaches to control for compositional change, namely, a hedonics approach[2], a repeat-sales approach[3], and a stratification approach[4]. One clear advantage of the stratification approach (grouping dwellings according to similar price-determining characteristics) is its relative simplicity. The effectiveness of this approach is determined by the degree of stratification possible. This is a function of the amount of dwelling characteristics detail available, and the volume of market activity.

More sophisticated approaches like hedonics and repeat-sales have shortcomings. For example, the effectiveness of the hedonic approach is critically dependent on the availability of price determining characteristics data for use in the hedonic model. Currently, the characteristics data required to support the hedonic approach are available for house sales in Hobart and Adelaide only. The effectiveness of the repeat sales approach requires a long time series of price data for individual properties, given their infrequent turnover. Further, this approach makes the assumption that the 'quality' of the individual properties is constant over time which does not take into account any improvements (renovations) made or the effect of depreciation. The nature of the estimation

---

[2]   The hedonic approach utilises regression techniques to estimate a price for each of the characteristics of dwellings and so determine an overall market price.

[3]   The repeat sales approach controls for quality change by maintaining a longitudinal database of properties and, when properties are sold more than once, calculating price changes between successive sales dates. Regression techniques are used to calculate the overall price index for each period.

[4]   The stratification approach involves stratifying the observations in such a way as to group the prices for the 'most like' or 'most similar' dwellings.

technique also means that at least the tail end of the series is potentially subject to significant revision. Again, sufficient data to enable the use of this method are not currently available.

Given the absence of a comprehensive data set to enable the use of either the hedonics or the repeat sales approach, the option taken by the ABS for controlling for compositional effects is the stratification approach.

Data requirements for HPI compilation under a stratification approach

There are four significant dates related to the purchase of a residential property relevant to the construction of an HPI: (1) verbal agreement to purchase at a negotiated price; (2) approval of mortgage financing; (3) exchange of contract; and (4) settlement and registration of property documents. In compiling the HPI series, the ABS selected the exchange of contract (3) as the preferred date to use because it most closely approximates the time at which the market price is determined.

The two sources of administrative data used to compile the HPI are described below.

Valuers-General data

Each property transaction, regardless of type or location, must be registered with the VGs Office, to enable the government authority to maintain a record of property ownership. The information contained in these records varies from one jurisdiction to another, and includes not only data on the sale of the property (including suburb, date and price) but also, in some jurisdictions, information on the physical characteristics of the property. The high coverage and low respondent burden offered by VGs data make these data the obvious choice of an administrative data source for HPI calculation. The main disadvantage of these data sets is that lengthy delays are often involved in all the data becoming available.

Different jurisdictions have different legislation governing the reporting responsibilities of parties to property transfers. In general, the requirement is for the property transfer to be registered within 60-90 days of settlement. When combined with the lag between exchange of contracts and final settlement, the complete process can take around six months or so for all transactions relating to a particular quarter to be received by the ABS. A further concern for reliable price measurement is that properties with higher prices generally take longer to settle.

The implication is that details received by the ABS relating to the property sales in a particular quarter are distributed in a biased way. The resulting bias of early reported data is always downwards but its magnitude is not consistent, either between cities or over time within any one city. As a result, it is necessary to obtain an almost complete data set for each quarter before it is possible to obtain an accurate measure of the house prices for the quarter.

Clearly, it is not possible to produce a timely HPI from this administrative data source, although it is the preferred source as far as comprehensiveness is concerned.

<u>Mortgage lenders' data</u>

As a large percentage of house sales involve mortgages, loan approval documents created by mortgage lenders are a timely source of house price data. Applications for finance to purchase residential dwellings are generally processed, at the latest, shortly after the exchange of contracts. Although the data do not cover all house sales, they have sufficient coverage for them to be used as a means of estimating the recent movements in the established house prices. The ABS also found that estimates of house price using loan approval data are comparable with estimates based on VGs data.

A shortcoming of this data set is that the loan documents systems used by most mortgage lenders do not capture the actual date of exchange. However, it has proved possible to model the loan approval date to emulate the exchange date.

Mortgage lenders' data are also biased. Loan documents do not necessarily record the actual sale price of the property in all cases. Some loan approvals data contain the security valuation. Analysis has shown that median prices derived from mortgage approvals data are always higher than median prices derived from the complete VGs data set.

Stratification

Stratifying the housing stock is based on the idea that the next best alternative to comparing the price movements of identical houses is to compare the price movements of houses that are similar in their attributes. The stratification approach achieves this objective by combining suburbs into strata that group together (or 'cluster') according to a set of fundamentals that determine house prices. The aim is to (1) maximise the level of homogeneity of the suburbs contained in the same stratum, and (2) have a sufficient number of sales observations in each stratum in each reference quarter. The finer the level of stratification available, the more similar or homogenous the houses within the cluster will be.

However, the finer the level of stratification, the fewer observed property sales will occur. So the clusters defined have to balance the homogeneity of housing characteristics and the number of observations required to produce a reliable median price.

Given the complexity of stratification, the ABS has undertaken considerable research in this area and has refined its method of stratification over time. The original approach was based on geographically stratified median prices and evolved into a stratification method based on geography and structural, locational and neighbourhood characteristics. A new framework is about to be introduced which offers a simpler stratification method – based on the long-term median house price and the neighbourhood characteristics of a suburb.

Combining administrative data sets to produce an HPI

VGs data are used to compile the HPI up to the point for which a complete set of data is available on an exchange date basis. Typically, complete sets of data are available for all cities up to the quarter ending two quarters prior to the most recent quarter. For example, if the current quarter is the September quarter 2008, VGs data is available up to the March quarter 2008. The series based on these data is referred to as the benchmark series.

To produce a preliminary index for the two most recent quarters, the benchmark series is projected using a mixture of supplementary mortgage lenders' data and early VGs data. As the full set of VGs benchmark data becomes available, the preliminary figures are revised. No attempt is made to adjust for changes in quality of houses.

The HPI is compiled as a weighted average of indexes derived from movements in median house prices, stratified by cluster, for each city. The national index is compiled in the same way as that for the benchmark series (i.e. as a weighted average of the eight capital cities). The weights used are the value of the housing stock in each cluster in each city. The values were obtained by combining Population Census house counts with a selected quarter's mean prices.

Summary

The Australian HPI has been compiled since 1986 and the method for constructing the HPI has developed over time. In compiling the HPI, a key challenge for statistical agencies is to construct a measure that abstracts from compositional change, and the ABS solution is the stratification approach.

Administrative data plays a key role in constructing the Australian HPI. Two complementary sets of administrative data are used. One provides an exhaustive collection of transactions at the expense of timeliness, and the other is timely but is not of ideal/perfect coverage. Combining these data sets yield a reliable measure of house price change.

References

Australian Bureau of Statistics, *Household Wealth and Wealth Distribution, Australia 2005-2006,* cat. no. 6554.0

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

Australian Bureau of Statistics, Refining the Stratification for the Established House Price Index, cat. no. 1352.0.55.093

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

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

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

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

Olczyk, A. and Neideck, G. (2007), *Measuring house price movements: methods, issues and some recent experience in the Australian context,* 10th Meeting of the UN International Working Group on Price Indices (The Ottawa Group), Ottawa, 9-12 October 2007

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

Olczyk, A and Lane, S (2008) *Refining the Stratification for the Established House Price Index, cat. no.* 1352.0.55.093