# Land Registry House Price Index Methodology

---

*Discussion and Description of Methodology*

CALNEA ANALYTICS

# Table of Contents

## Background

Calnea Analytics believes that the new *Land Registry House Price Indices* represent a significant advance in the measurement of house price changes throughout the country – both in terms of national reliability and local coverage.

Unlike other house price statistics produced by various institutions, the *Land Registry House Price Index* is based upon the Repeat Sales Regression (RSR) Method. Under the RSR method, house price growth is measured by looking at houses which have been bought or sold more than once. The use of repeat transactions controls for differences in the quality of the houses comprising the monthly sample – thereby a more constant quality comparison. The Index is produced from the Land Registry's own 'price paid' dataset that contains the transaction details of every residential property sold in England & Wales – of which there are approximately 100,000 monthly. Calnea Analytics has created for the Land Registry a series of local and national indices that provide a picture of what has happened to house prices across the country.

## Introduction

There is a very strong interest in house price movement among various media, public and financial sectors throughout England & Wales.

The measurement of house price movements is by no means an easy task; the primary complication is because houses are heterogeneous goods. No two houses are exactly the same, and even seemingly identical houses within the same small geographic area will follow different (though often correlated) processes of price appreciation. Therefore, it is convenient and useful to capture the overall average price trends followed by a group of houses. In such a situation, it is common practice in economics to propose a single price index, e.g. consumer price index, etc. A house price index is simply "one of many plausible measures of the central tendency of house price appreciation for a particular group of properties" (Araham and Schauman, 1991).

Accurate and reliable house price information is demanded for a range of applications, from economic and social policy to the decisions of private investors.

Information on house price levels and growth rates form the basis of key decisions made by a wide variety of private actors for example; home buyers, home sellers, mortgage lenders, valuation surveyors, and house builders.

An accurate house price index would inform national economic and social policy by providing a key indicator of housing market and wider macro-economic conditions. A measure of house prices is also needed for research into the sources of inflation, economic growth and cycles. House price appreciation can also be an indicator of the strength of the local economy and differences in appreciation rates can reflect differences in regional economic performance. In addition there is a natural interest in house prices amongst the growing home-owning population; housing represents the most valuable asset for many people.

Accurate price series on a large number of assets, such as equities and bonds, are an essential feature of financial market research. Analogous information for local property markets (e.g. relatively small cities and towns) would be useful not only to researchers but also to town officials and homeowners.

# The Logic of Compiling a House Price Index[1]

The key dataset one normally has for compiling a house price index is a set of price observations, either actual transaction prices when a house is sold, or assessed prices according to professional opinion (Clapp and Giaccotto (1992) consider assessed price as useful data to estimate house price index). With abundant price data at various time periods available for the group of properties in a particular housing market, house price indices can be estimated for using statistical techniques.

## Idealized Dataset

Ideally, with "perfect information" we would be able to observe every house's price at every point in time. If we denote the price of house n at time t as $P_{n,t}$, and assume that the population size of properties is N and the total number of periods under consideration is $T+1$, the "perfect information" dataset will come in a matrix form, which is illustrated in ***Table 1***.

Also, denote the underlying unknown index of period t as $I_t$. Our goal is to estimate the whole series of $I_t$ for $t = 0,...,T$ from the matrix of data set of $P_{n,t}$.

***Table 1*: An ideal dataset of house prices**

| | TIME | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | … | **t** | … | **T** |
| **Property 1** | $P_{1,0}$ | $P_{1,1}$ | $P_{1,2}$ | $P_{1,3}$ | … | $P_{1,t}$ | … | $P_{1,T}$ |
| **Property 2** | $P_{2,0}$ | $P_{2,1}$ | $P_{2,2}$ | $P_{2,3}$ | … | $P_{2,t}$ | … | $P_{2,T}$ |
| **Property 3** | $P_{3,0}$ | $P_{3,1}$ | $P_{3,2}$ | $P_{3,3}$ | … | $P_{3,t}$ | … | $P_{3,T}$ |
| **Property 4** | $P_{4,0}$ | $P_{4,1}$ | $P_{4,2}$ | $P_{4,3}$ | … | $P_{4,t}$ | … | $P_{4,T}$ |
| … | … | … | … | … | … | … | … | … |
| **Property n** | $P_{n,0}$ | $P_{n,1}$ | $P_{n,2}$ | $P_{n,3}$ | … | $P_{n,t}$ | … | $P_{n,T}$ |
| … | … | … | … | … | … | … | … | … |
| **Property N** | $P_{N,0}$ | $P_{N,1}$ | $P_{N,2}$ | $P_{N,3}$ | … | $P_{N,t}$ | … | $P_{N,T}$ |
| **Indices** | $I_0$ | $I_1$ | $I_2$ | $I_3$ | … | $I_t$ | … | $I_{T-1}$ |

[1] Part of this section is adopted from Wang and Zorn (1997).

## Real Dataset

However, in reality, house prices are not readily observable continuously. The price of any given property is observable only when it is sold or it is assessed by a professional assessor, both of which happen at infrequent intervals.

In addition, new houses are always being constructed and old houses demolished (Wang and Zorn, 1997). Accordingly, the population of houses we can observe is changing over time.

Consequently, rather than having the convenient data matrix illustrated in ***Table 1***, we have instead a fragmented dataset as shown in ***Table 2***. This fragmented dataset is the starting point of any index estimation.

***Table 2:*** **Dataset of House Prices in Reality**

| | **TIME** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | … | **T** |
| **Property 1** | | $P_{1,1}$ | | | | $P_{1,5}$ | | | | … | |
| **Property 2** | $P_{2,0}$ | | | | | | $P_{2,6}$ | | | … | |
| **Property 3** | | | $P_{3,2}$ | | $P_{3,4}$ | | | | | … | |
| **Property 4** | | | | $P_{4,3}$ | | | | $P_{4,7}$ | | … | |
| **Property 5** | | | $P_{5,2}$ | | | $P_{5,5}$ | | | | … | $P_{5,T}$ |
| **Property 6** | | $P_{6,1}$ | | $P_{6,3}$ | | | | | | … | |
| **Property 7** | | | $P_{7,2}$ | | | | $P_{7,6}$ | | | … | |
| **Property 8** | | | | | | | | | $P_{8,8}$ | … | $P_{8,T}$ |
| **Property 9** | | | | | | $P_{9,5}$ | $P_{9,6}$ | | $P_{9,8}$ | … | |
| **Property 10** | | | $P_{10,2}$ | | | | | | | … | |
| … | … | … | … | … | … | … | … | … | … | … | … |
| **Indices** | $I_0$ | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ | $I_7$ | $I_8$ | … | $I_T$ |

The lesson from the above illustration is that the quality of available dataset (size, accuracy, up-to-date, etc.) is the biggest factor in house price index estimation.

The two main, theoretically rigorous techniques used to estimate house price indices are hedonic regression and repeat sale regression. Due to historic data unavailability, the more suitable repeat sale regression method was not used in the UK. Now, with the wealth of data on individual transaction prices available to the Land Registry it has become possible to implement this more suitable method.

## Repeat Sales Regression Methodology

Repeat sales regression method (RSR, hereafter) of estimating house price indices was first introduced by Bailey, Muth and Nourse (1963). The main idea behind RSR is that market-wide growth rate for a given period is reflected in "averaging the observed individual growth rates of all properties that were transacted twice in that time period" (Leishman, 2000). Wherever the necessary data has been available, RSR has gained popularity in economic application. RSR is now widely adopted by a number of large private, state and federal organizations in the United States, for example the highly influential Housing Economics and Financial Research Department at Freddie Mac (Federal Home Loan Mortgage Corporation). RSR is also a crucial tool widely used in the study of other markets characterized by infrequent trading, such as the art market (Goetzmann, 1992).

### The Model Setting

The model underlying Bailey et al (1963). method can be written as follows, using the notation we introduced above. Any property $n$ that has been sold twice satisfies the following equation,

$$R_{n,t_1,t_2} = \frac{P_{n,t_2}}{P_{n,t_1}} = \frac{I_{t_2}}{I_{t_1}} \times U_{n,t_1,t_2},$$

where $U_{n,t_1,t_2}$ is an idiosyncratic error term; $t_1 < t_2$, for $t_1 = 0,1,...,T-1$, $t_2 = 1,...,T$

The model means that the ratio of the final sales price in period $t_2$ to initial sales price in period $t_2$ for the $n-th$ property, which is defined as $R_{n,t_1,t_2}$, is equal to the ratio of the (unknown) indices of the corresponding two periods with a property-specific noise term.

The intuition behind the model is obvious. A pair of sales prices of a given property contains information on house price appreciation happening in the market it belongs to within the periods between the first and second sale (Case and Shiller, 1987). Therefore, the observed price appreciation between the two sales of this given property can be attributed to two factors: 1) the general trend of appreciation of the housing market this property belongs to, and 2) some property-specific elements that drive its house price to deviate slightly from the overall trend of the housing market. The first factor is represented in the index ratio $I_{t_2}/I_{t_1}$, while the second factor is captured by the error term $U_{n,t_1,t_2}$, in the above model.

To make the BMN model more practical, one can transform the model into a linear form by taking logarithm of both sides of the equation,

$$\log(R_{n,t_1,t_2}) = -\log(I_{t_1}) + \log(I_{t_2}) + \log(U_{n,t_1,t_2}) \text{, or}$$

$$r_{n,t_1,t_2} = -i_{t_1} + i_{t_2} + u_{n,t_1,t_2}$$

where lower case letters stand for the logarithms of the corresponding capital letters.

In BMN model, it is assumed that the error term $u_{n,t_1,t_2}$ have zero means, constant variance, and are uncorrelated with each other and any $i_t$.

Recall the goal is to estimate $I_t$, or, equivalently, $i_t$, for $t = 0,...,T$. If one makes T dummy variables $x_t$, for $t = 0,...,T$, and rewrites the model above as,

$$r_{n,t_1,t_2} = -\sum_{t=0}^{T-1} i_t b_t + u_{n,t_1,t_2}$$

where, if we denote initial sale and final sale periods as $t_1$ and $t_2$ respectively,

$$x_t = \begin{cases} -1, \text{ if } t = t_1 \\ 1, \text{ if } t = t_2 \\ 0, \text{ otherwise} \end{cases}$$

Therefore, the model becomes multiple linear regression model with $T$ dummy independent variable. In matrix notation, the model is:

$$r_{M\times1} = X_{M\times T} i_{T\times1} + u_{M\times1},$$

if $X$ is the data matrix of the dummy variables $x_t$, and there are altogether $M$ pairs of repeat sales prices in the sample.

## Model Estimation and Results Analysis

The model can be estimated by ordinary least squared (OLS) method, which can be implemented by various statistical packages. The regression output are the estimated parameters $\hat{i}_t$, for $t = 0, \ldots, T$, which in matrix form is

$$\hat{i} = (X'X)^{-1} X'r$$

However, this is still in logarithm form. In addition, it is convention to have an index of 100 for base period $t = 0$, i.e., $\overline{I_0} = 100$. If we denote the indices estimated from OLS as $\tilde{I}_t$, and indices after rebasing as $\overline{I_t}$, for $t = 0, \ldots, T$, the rebase relationship is the following:

$$\frac{\tilde{I}_t}{\tilde{I}_0} = \frac{\overline{I_t}}{\overline{I_0}},$$

which is equivalent to,

$$\log(\overline{I_t}) = \log(\tilde{I}_t) - \log(\tilde{I}_0) + \log(\overline{I_0}) = \hat{i}_t - \hat{i}_0 + \log(100),$$

Therefore, for $t = 0, \ldots, T$,

$$\overline{I_t} = \exp(\hat{i}_t - \hat{i}_0 + \log 100),$$

We hence get house price indices $\overline{I_t}$, for $t = 0, \ldots, T$, after rebasing $\overline{I_0} = 100$.

## Features of Repeat Sale Regression Method

### Separating Quality from Price

The greatest strength of RSR is its method of "separating quality from price" (Araham and Schauman, 1991).

The difficulty hedonic method facing is the fact that it relies heavily on the correct specification of both the functional form of the model and the set of property characteristics (Meese and Wallace, 1997). Case and Quigley(1991) illustrated hedonic model in a general form, $P_t = f(x,t)$, i.e., house price is a function of time $t$ and the vector of all physical and locational characteristics $x$. This requires $f$ to be correctly specified, and the vector $x$ is correctly chosen and accurately measured, none of which can be guaranteed. These can potentially introduce what is known as 'misspecification bias' (Bailey, Muth and Nourse, 1963; Case and Shiller, 1987).

For the RSR method, researchers control for hedonic characteristics by examining only those properties that have been sold more than once during the period under consideration. Case and Quigley(1991) provided a general expression of RSR model as: $P_{t_1} / P_{t_2} = g(t_1, t_2)$, which obviously highlights the feature of RSR that it only depends on price data and transaction dates, both of which can be measured accurately. The functional form $g$ is also unique in theory, which is a clear advantage over hedonic regression.

### Less Strict Data Requirements

RSR's less strict data requirement is advantageous.

As mentioned above, while hedonic method is a valid technique, its onerous data requirements limit the datasets that can be used. Data on many of the attributes that can be important determinants of the price of a property, particularly qualitative attributes such as neighbourhood, location, quality of workmanship, etc., are often not available (Case, Pollakowski and Wachter, 1991). Usually, large databases containing information on both quantitative and qualitative attributes are built by mortgage lenders. Using such databases can introduce two potential sources of bias: a) "the sample includes only those loans and properties that are subject to historical agency purchase patterns and loan amount restrictions"; b) "the inclusion of refinancing transactions whose property values represent estimates by appraisers can bias a derived price index" (Cho, 1996). The type and location of properties over which these institutions grant mortgages are not randomly distributed. The data available to the lenders are from their respective customers bases concentrated in the North and South of the country respectively.

RSR requires only data on transaction prices and dates of two consequent transactions, and does not require data on physical attributes. Thus, RSR is able to capture the information value from much larger nationwide datasets.

In addition, RSR enables the creation of local house price indices. This is due to the fact that the data requirements of RSR are greatly simplified when compared with hedonic regression methods. This simplicity gives RSR great prospect of application and significant potential for application to local housing markets. UK based academics have proposed the application of RSR in developing a "system" of local house price indices (Costello and Watkins, 2002) (Leishman Watkins and Fraser, 2003). As frequently cited, one of the main problems with the existing indices is their lack of geographic focus. Munro & Maclennan (1986) point to the need to examine house price appreciation rates at neighbourhood level and caution against making assumptions about the aggregate nature and behaviour of markets.

## Dataset

As Araham and Schauman (1991) pointed out, the biggest obstacle in creating a repeat-sales index is getting the necessary large amount of raw data. The fact that the RSR method is not in use in the UK yet (Fleming and Nellis, 1994) is due to the limited availability of data on a large scale. Leishman (2000), Leishman and Watkins (2002), and Leishman, Watkins and Fraser (2002) have done some pioneering research in applying RSR method in UK. But due to the limited availability of data, their work was confined to experimenting with RSR using data from a number of cities in Scotland.

The problem with respect to dataset restriction was eased by one significant factor - the restoration of 'price paid' into the Land Register in England and Wales from April 1st, 2000. Land Registry has utilised outside economics and valuation expertise to create a whole system of national and regional house price indices, and has launched the its system of indices using the RSR method in the UK.

This section below focuses on the Land Registry database and the sampling technique used by Calnea Analytics.

## Size

The database used is the Land Registry *price paid* dataset. Since every residential property transaction in England & Wales is required to file with the Land Registry, the Land Registry database contains every open-market property transaction since April 1st, 2000. Between April 1st, 2000 and April 30, 2006 approximately ~7,000,000 transaction records were available for analysis. This figure increases by approximately 100,000 records per month. The Land Registry database is the largest and most comprehensive database of house selling prices available for England & Wales.

It is clearly a major advantage to have the largest database, as "for a given level of aggregation, more data means tighter standard errors about the estimated mean of the price series" (Araham and Schauman, 1991). The large size of the data also gives us the ability to develop a system of indices for finer geographic and price bands than any other index.

## Recent

House prices take between 1 and 3 months from the date of sale before being registered with the Land Registry. The Land Registry provides Calnea Analytics with monthly updates for all transactions occurring a month prior to the date.

## Cleansing

A data cleansing process is necessary in order to spot potential coding errors and outliers in the database. Conservative editing standards exclude all transactions where doubts over validity exist. Similar to the RSR practices employed by US institutions; both Land Registry and Calnea Analytics employ a cleansing process to filter address errors and various other outliers.

## House Improvement Adjustment

RSR requires that "the property has undergone neither a significant enhancement in value, such as remodeling, nor substantial physical deterioration" (Araham and Schauman, 1991), so that the single property price appreciation can be attributed solely to trends of market price movement. It is clear that all properties experience physical depreciation and in addition many properties are improved prior to sale.

There are two contrasting approaches in the academic literature. One approach advocates not adjusting for this issue. It could be potentially argued that overtime in the long-run the value of house improvements will equate to the value of depreciation such that this factor will hold constant. If it is viewed that the main component of value is space, i.e. square footage then the argument regarding the irrelevance of depreciation or improvements gains strength.

The alternative approach is to make an adjustment to the index to reflect the average value of improvements minus depreciation. Araham and Schauman (1991) believe that it is possible to correct the index directly from knowledge of the value of improvements nationwide.

A perfect RSR model would require the absence of systematic property deterioration or improvement across the sample. It is worth noting that this same bias affects hedonic model variants in so far as the home improvement or deterioration is not perfectly captured by the hedonic variables. It is the view of Calnea Analytics that in so far as data on systematic property deterioration or improvement is available, this information can and should be practically incorporated in the model. One valuable source of such information is the English House Condition Survey (EHCS) annually conducted by the Office of the Deputy Prime Minister.

## Technical Considerations

There are a number of technical issues that arise from implementation of the RSR method. It must be recognized that no index estimation method is perfect, and the RSR method, while we believe to be extremely robust and value adding, is not by nature bias free. A description of various technical issues follows below.

## Heteroskedasticity

Recall in the model of Bailey, Muth and Nourse (1963), $u_{n,t_1,t_2}$ is by convention assumed to have constant variances. Case and Shiller(1987, 1989) argued that $u_{n,t_1,t_2}$ has varying variances, and this is called heteroskedasticity in econometric literature.

The solution Case and Shiller(1987, 1989) provided was a “weighted repeat sale” model. However, it has been shown that the effect of the model is ambiguous. Leishman and Watkins (2002) using Scottish data, applied both the normal RS method and weighted RS method and concluded that the normal RS method was preferred.

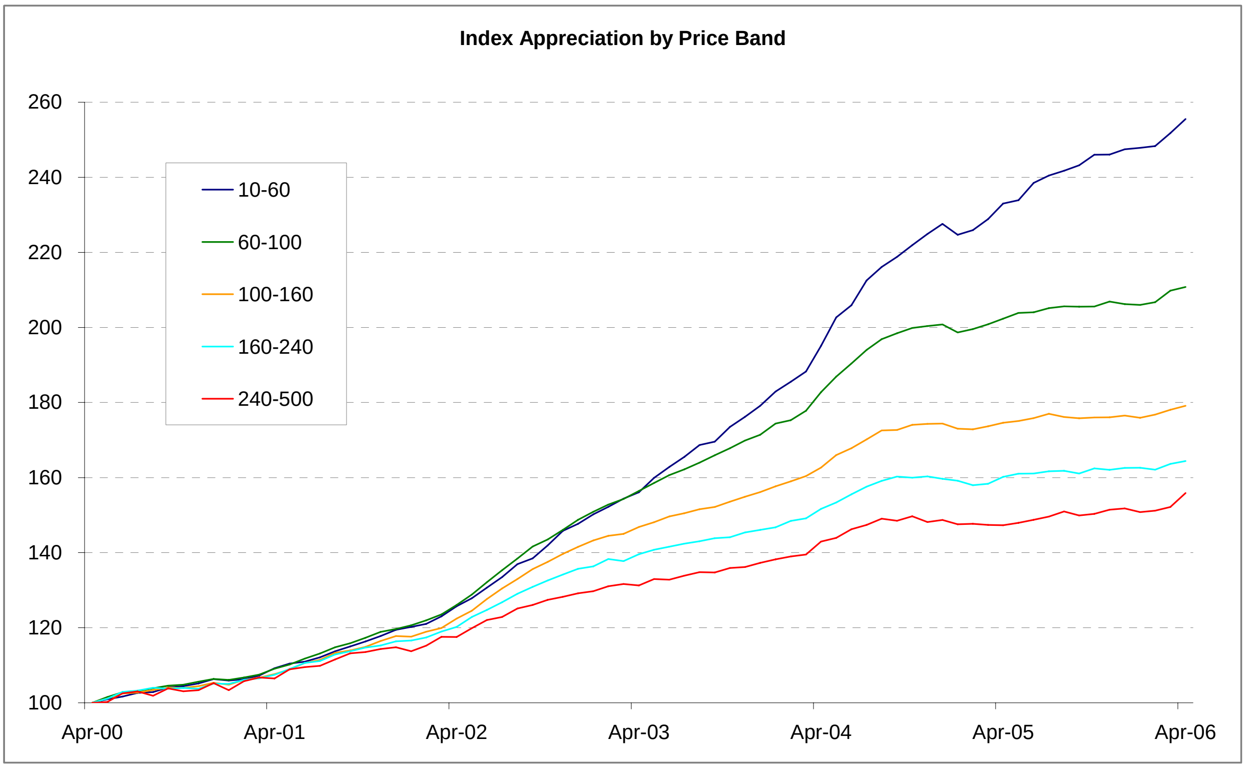
## Sample selection

An important feature of the RSR method is that the sample used in RS regression only includes houses that have been sold more than once, and therefore suffers from ‘sample selection bias’ (Case, Pollakowski and Wachter, 1991; Cho, 1996; Gatzlaff and Haurin, 1997; Meese and Wallace, 1997; Steele and Goy, 1997).

The empirical study of Clapp, Giacotto and Tirtiroglu (1991) found no systematic differences between the RS sample and the full sample of all transactions over the long run. They argued that “arbitrage typically forces prices for the repeat sample to grow at the same rate as those for the full sample”. Another study of Wallace and Meese (1997) also arrived at the conclusion that the sub-sample of RSR is actually representative of all home sales during the period under consideration.

The RSR index is naturally more reflective of properties that transact more frequently. In so far as a differential in price appreciation exists between properties based on the relative frequency of transactions, the RSR measure will be naturally weighted towards the more frequently transacting subset of properties.

There are a variety of reasons why the holding duration of properties might be unevenly distributed. The increase in transaction costs for more expensive properties due to stamp duty may result in a decreased turnover of more expensive homes. ‘Life-cycle’ theories on property holding period posit that less expensive properties are traded more frequently - when people move up the property ‘ladder’ they tend to move home less often. In addition the Buy-to-Let market is more active in the lower price brackets. Policy-makers need to be aware of the price appreciation differentials between submarkets, especially when there is systematic variation in the frequency of transactions between these submarkets. The chart that follows illustrates differentials in price appreciation according to price bracket.




## Inconstancy of attribute appreciation

If one takes a hedonic perspective to consider a house as a bundle of separate attributes, both qualitative ones and quantitative ones, the setting of RS method implicitly assumes that the prices of all these attributes move at the same rate overtime, which may not be the case (Case, Pollakowski and Watcher, 1991). However negligible this is another potential bias of house price indices.

## Multicollinearity

Multicollinearity refers to situations where there is an approximate linear relationship among independent variables (Kennedy, 2003). This is not a rare phenomenon in econometrics. Although the Gauss-Markov Theorem still ensures a best linear unbiased estimator, some problems can be caused in applied research.

1. Small changes in the data produce wide swings in the parameter estimates.

2. Coefficients may have very high standard errors and low significance levels even though they are jointly significant and the $R^2$ for the regression is quite high.

3. Coefficients may have the 'wrong' sign or implausible magnitudes. (William, 2003)

Unlike other model specification problems, the problem of multicollinearity is caused by the specific sample used in the regression (Kennedy, 2003).

Cho (1996) pointed out that this problem tends to arise with small sample sizes. With only a small percentage of transactions, two columns of the data matrix are by construction similar and hence highly correlated. However, this problem does not present itself materially in the UK housing market. The housing market liquidity in England & Wales is greater than most other countries. With ~100,000 residential property transactions per month and close to a 70% rate of home-ownership, the national sample data does not suffer from material multicollinearity.

## Inefficiency

One complaint about the RSR method is that it only uses a portion of the transaction dataset (i.e. it only uses matched pairs and ignores other transactions) therefore suffering from inefficiency. Such comments often fail to note that the explanatory and informative power of the remaining portion of data (i.e. the matched pairs) is notably superior to a similar sized dataset used by hedonic methods. A RSR index based on 100 price observations is superior to a Hedonic index based on 100 price observations – ceteris paribus. The reason for this is that the hedonic methods also suffer from information inefficiency. There is an important distinction is between an index utilising 100% of transactions versus an index utilising 100% of the possible information.

## Model improvement

Some academic literature has been devoted towards proposing hybrid models, which tend to combine the merits of both hedonic method and repeat sale method (Meese and Wallace, 1997; Case, Pollakowski and Wachter, 1991). Although a theoretical attractive idea, the hybrid models still suffer from the model specification and variable measurement problems of normal hedonic models. Empirical studies have yielded ambiguous results, and the superiority of hybrid models to RSR models at least at the moment lacks of supportive evidence.

Clapp and Giaccotto (1992) tried to use an assessed value method to improve the efficiency of the model, but found the method gave substantially the same estimates of price trends as RSR gave over the long run.

# Revision

## Revision of Historical RSR Indices

Historical published data is revised for two main reasons:

1. Data capture lag - there is a time-lag between the sale of a property and the subsequent registration of this information with Land Registry. This means the transaction data for the previous month will not be 100% complete when the monthly report is prepared. The missing data is included as it becomes available.

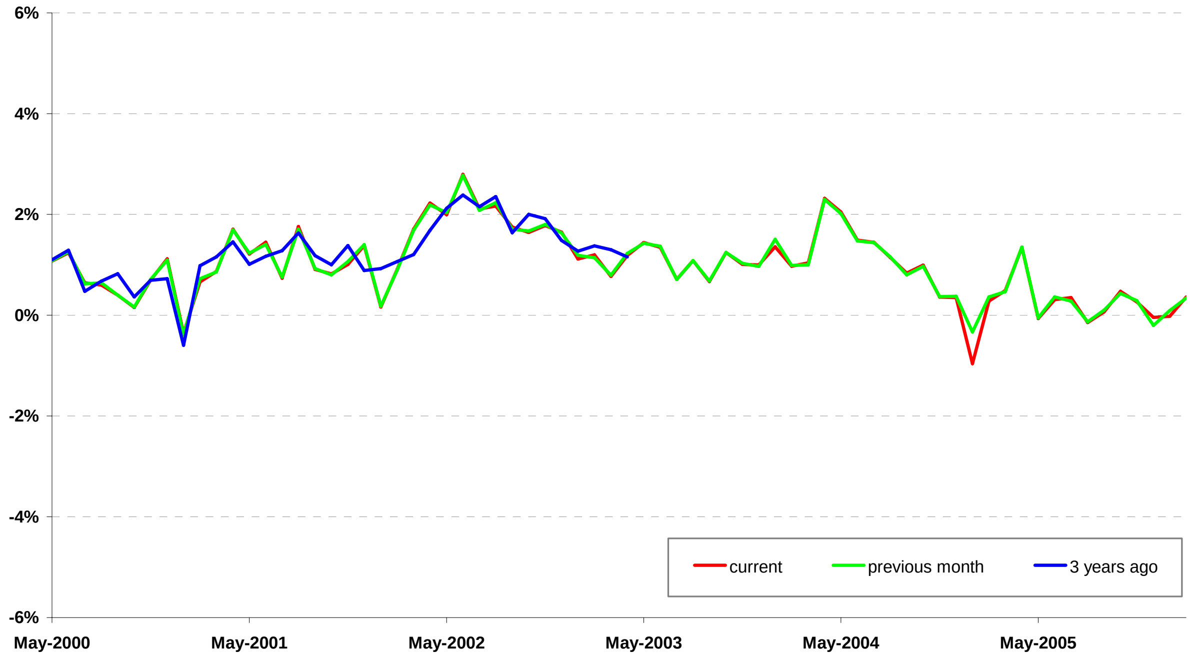
2. HPI based on repeat transactions - When Land Registry publishes its HPI reports each month, the index represents the view on historic house price movements at the time. As new information becomes available, the published indices are revised to reflect any new data.

Clapp and Giaccotto (1999) pointed out that any RSR price index will be revised every time new information becomes available. This is a feature of any price index that is in continual receipt of additional information. As such this is classed as a practical issue rather than a methodological problem. Levels of understanding in all fields are revised in the presence of additional information. Due to the nature of statistical sampling, most of the important indices and other economic indicators are subject to revision (Baumohl, 2005). Shiller (1994) in his book argues that revisions are a necessary aspect of any price index, whether the index is based on a repeat-sales or hedonic model.

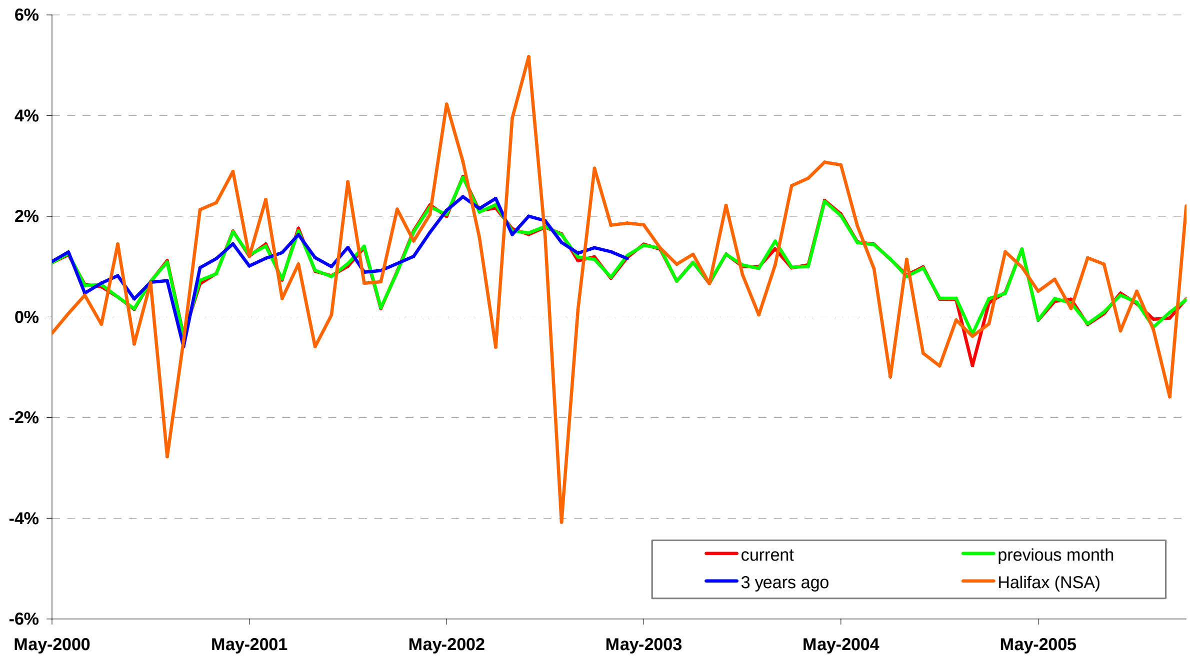
## Revision Size

Whilst revisions are to be expected, policy-makers should be made aware that revisions are relatively small. The following charts show a current price index alongside one created in the previous month as well as one created with data limited to three years prior. All price indices shown are prior to seasonal adjustment. It is apparent that the degree of revision does not represent a fundamental change in estimates. The Halifax HPI is displayed alongside the revisions in the second chart in order to show how minimal revisions are in relation to the volatility of another index.

## Index Revisions - Monthly Change




## Index Revisions - In Perspective



## RSR Index Revision

| | 3 years ago | | | previous month | | | current | | Halifax | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Index | Change | Revision | Index | Change | Revision | Index | Change | Index | Change |
| Apr-00 | 100.0 | | | 100.0 | | | 100.0 | | 100.0 | |
| May-00 | 101.1 | 1.10% | 0.0% | 101.1 | 1.08% | 0.0% | 101.1 | 1.08% | 99.7 | -0.33% |
| Jun-00 | 102.4 | 1.29% | -0.1% | 102.4 | 1.26% | 0.0% | 102.3 | 1.24% | 99.7 | 0.07% |
| Jul-00 | 102.9 | 0.47% | 0.2% | 103.0 | 0.62% | 0.0% | 103.0 | 0.65% | 100.2 | 0.44% |
| Aug-00 | 103.6 | 0.68% | -0.1% | 103.7 | 0.64% | 0.0% | 103.6 | 0.60% | 100.0 | -0.14% |
| Sep-00 | 104.4 | 0.82% | -0.4% | 104.1 | 0.40% | 0.0% | 104.0 | 0.40% | 101.5 | 1.45% |
| Oct-00 | 104.8 | 0.36% | -0.2% | 104.2 | 0.16% | 0.0% | 104.2 | 0.15% | 100.9 | -0.53% |
| Nov-00 | 105.5 | 0.69% | 0.0% | 105.0 | 0.71% | 0.0% | 104.9 | 0.70% | 101.6 | 0.68% |
| Dec-00 | 106.3 | 0.72% | 0.4% | 106.1 | 1.09% | 0.0% | 106.1 | 1.12% | 98.8 | -2.78% |
| Jan-01 | 105.7 | -0.59% | 0.2% | 105.7 | -0.39% | 0.0% | 105.7 | -0.36% | 98.3 | -0.48% |
| Feb-01 | 106.7 | 0.98% | -0.3% | 106.5 | 0.73% | -0.1% | 106.4 | 0.66% | 100.4 | 2.13% |
| Mar-01 | 107.9 | 1.16% | -0.3% | 107.4 | 0.86% | 0.0% | 107.3 | 0.87% | 102.7 | 2.27% |
| Apr-01 | 109.5 | 1.45% | 0.3% | 109.2 | 1.69% | 0.0% | 109.2 | 1.71% | 105.7 | 2.89% |
| May-01 | 110.6 | 1.01% | 0.2% | 110.6 | 1.23% | 0.0% | 110.5 | 1.21% | 106.9 | 1.20% |
| Jun-01 | 111.9 | 1.17% | 0.3% | 112.1 | 1.41% | 0.0% | 112.1 | 1.45% | 109.4 | 2.34% |
| Jul-01 | 113.3 | 1.28% | -0.5% | 113.0 | 0.76% | 0.0% | 112.9 | 0.73% | 109.8 | 0.36% |
| Aug-01 | 115.2 | 1.64% | 0.1% | 114.9 | 1.71% | 0.0% | 114.9 | 1.76% | 111.0 | 1.05% |
| Sep-01 | 116.6 | 1.18% | -0.3% | 116.0 | 0.93% | 0.0% | 115.9 | 0.91% | 110.4 | -0.59% |
| Oct-01 | 117.7 | 1.00% | -0.2% | 116.9 | 0.80% | 0.0% | 116.9 | 0.82% | 110.4 | 0.03% |
| Nov-01 | 119.4 | 1.38% | -0.4% | 118.1 | 1.06% | -0.1% | 118.1 | 1.01% | 113.4 | 2.69% |
| Dec-01 | 120.4 | 0.89% | 0.5% | 119.8 | 1.40% | 0.0% | 119.7 | 1.38% | 114.1 | 0.67% |
| Jan-02 | 121.5 | 0.92% | -0.8% | 120.0 | 0.18% | 0.0% | 119.9 | 0.16% | 114.9 | 0.70% |
| Feb-02 | 122.8 | 1.06% | -0.2% | 121.1 | 0.89% | 0.0% | 121.0 | 0.91% | 117.4 | 2.14% |
| Mar-02 | 124.3 | 1.20% | 0.5% | 123.1 | 1.69% | 0.0% | 123.1 | 1.71% | 119.1 | 1.51% |
| Apr-02 | 126.4 | 1.68% | 0.5% | 125.8 | 2.19% | 0.0% | 125.8 | 2.23% | 121.6 | 2.04% |
| May-02 | 129.1 | 2.12% | -0.1% | 128.4 | 2.04% | 0.0% | 128.3 | 2.00% | 126.7 | 4.23% |
| Jun-02 | 132.1 | 2.39% | 0.4% | 131.9 | 2.77% | 0.0% | 131.9 | 2.79% | 130.6 | 3.08% |
| Jul-02 | 135.0 | 2.15% | 0.0% | 134.7 | 2.08% | 0.0% | 134.7 | 2.11% | 132.7 | 1.58% |
| Aug-02 | 138.2 | 2.35% | -0.2% | 137.7 | 2.23% | -0.1% | 137.6 | 2.17% | 131.9 | -0.60% |
| Sep-02 | 140.4 | 1.64% | 0.1% | 140.0 | 1.71% | 0.0% | 140.0 | 1.75% | 137.1 | 3.95% |
| Oct-02 | 143.2 | 2.00% | -0.4% | 142.4 | 1.67% | 0.0% | 142.3 | 1.64% | 144.2 | 5.17% |
| Nov-02 | 146.0 | 1.91% | -0.1% | 144.9 | 1.80% | 0.0% | 144.8 | 1.78% | 146.5 | 1.58% |
| Dec-02 | 148.1 | 1.48% | 0.2% | 147.3 | 1.64% | 0.0% | 147.2 | 1.65% | 140.5 | -4.08% |
| Jan-03 | 150.0 | 1.27% | -0.2% | 149.1 | 1.19% | -0.1% | 148.9 | 1.12% | 140.7 | 0.15% |
| Feb-03 | 152.1 | 1.38% | -0.2% | 150.8 | 1.14% | 0.1% | 150.6 | 1.20% | 144.9 | 2.96% |
| Mar-03 | 154.1 | 1.30% | -0.5% | 152.0 | 0.79% | 0.0% | 151.8 | 0.77% | 147.5 | 1.82% |
| Apr-03 | 155.9 | 1.16% | 0.0% | 153.8 | 1.23% | 0.0% | 153.6 | 1.19% | 150.3 | 1.86% |
| May-03 | | | | 156.0 | 1.43% | 0.0% | 155.8 | 1.44% | 153.0 | 1.83% |
| Jun-03 | | | | 158.2 | 1.37% | 0.0% | 157.9 | 1.35% | 155.1 | 1.37% |
| Jul-03 | | | | 159.3 | 0.71% | 0.0% | 159.0 | 0.71% | 156.7 | 1.05% |
| Aug-03 | | | | 161.0 | 1.08% | 0.0% | 160.8 | 1.08% | 158.7 | 1.25% |
| Sep-03 | | | | 162.1 | 0.68% | 0.0% | 161.8 | 0.67% | 159.8 | 0.66% |
| Oct-03 | | | | 164.1 | 1.25% | 0.0% | 163.9 | 1.25% | 163.3 | 2.22% |
| Nov-03 | | | | 165.8 | 1.03% | 0.0% | 165.5 | 1.01% | 164.7 | 0.84% |
| Dec-03 | | | | 167.4 | 0.97% | 0.0% | 167.2 | 1.00% | 164.7 | 0.04% |
| Jan-04 | | | | 169.9 | 1.51% | -0.2% | 169.4 | 1.36% | 166.4 | 1.03% |
| Feb-04 | | | | 171.6 | 0.99% | 0.0% | 171.1 | 0.97% | 170.8 | 2.61% |
| Mar-04 | | | | 173.3 | 1.00% | 0.0% | 172.9 | 1.04% | 175.5 | 2.75% |
| Apr-04 | | | | 177.3 | 2.30% | 0.0% | 176.9 | 2.32% | 180.9 | 3.07% |
| May-04 | | | | 180.9 | 2.02% | 0.0% | 180.5 | 2.05% | 186.4 | 3.02% |
| Jun-04 | | | | 183.6 | 1.48% | 0.0% | 183.2 | 1.49% | 189.7 | 1.81% |
| Jul-04 | | | | 186.2 | 1.44% | 0.0% | 185.8 | 1.45% | 191.5 | 0.95% |
| Aug-04 | | | | 188.4 | 1.16% | 0.0% | 188.0 | 1.15% | 189.3 | -1.19% |
| Sep-04 | | | | 189.9 | 0.80% | 0.0% | 189.5 | 0.83% | 191.4 | 1.15% |
| Oct-04 | | | | 191.7 | 0.97% | 0.0% | 191.4 | 1.00% | 190.0 | -0.72% |
| Nov-04 | | | | 192.4 | 0.37% | 0.0% | 192.1 | 0.36% | 188.2 | -0.97% |
| Dec-04 | | | | 193.2 | 0.37% | 0.0% | 192.8 | 0.35% | 188.1 | -0.06% |
| Jan-05 | | | | 192.5 | -0.33% | -0.6% | 190.9 | -0.96% | 187.4 | -0.38% |
| Feb-05 | | | | 193.2 | 0.36% | -0.1% | 191.4 | 0.28% | 187.1 | -0.14% |
| Mar-05 | | | | 194.1 | 0.47% | 0.0% | 192.4 | 0.49% | 189.5 | 1.30% |
| Apr-05 | | | | 196.7 | 1.35% | 0.0% | 195.0 | 1.35% | 191.4 | 0.99% |
| May-05 | | | | 196.6 | -0.05% | 0.0% | 194.8 | -0.06% | 192.4 | 0.51% |
| Jun-05 | | | | 197.3 | 0.36% | -0.1% | 195.4 | 0.31% | 193.8 | 0.75% |
| Jul-05 | | | | 197.9 | 0.28% | 0.1% | 196.1 | 0.35% | 194.2 | 0.17% |
| Aug-05 | | | | 197.6 | -0.13% | 0.0% | 195.8 | -0.15% | 196.5 | 1.17% |
| Sep-05 | | | | 197.8 | 0.10% | 0.0% | 196.0 | 0.06% | 198.5 | 1.05% |
| Oct-05 | | | | 198.7 | 0.43% | 0.0% | 196.9 | 0.47% | 198.0 | -0.27% |
| Nov-05 | | | | 199.3 | 0.29% | 0.0% | 197.4 | 0.26% | 199.0 | 0.51% |
| Dec-05 | | | | 198.9 | -0.20% | 0.2% | 197.3 | -0.04% | 198.5 | -0.25% |
| Jan-06 | | | | 199.0 | 0.09% | -0.1% | 197.3 | -0.02% | 195.3 | -1.59% |
| Feb-06 | | | | 199.7 | 0.34% | 0.0% | 198.0 | 0.36% | 199.6 | 2.20% |

# Seasonal Adjustment

## Seasonality in time series

The seasonal component in time series corresponds to the regular movements observed in monthly time series during a twelve-month period.  The rate of house price inflation and property market transaction volumes are examples of this.  Other examples include increases in retail sales data associated with the Christmas period or the fall in industrial activity during holiday periods.

## Why publish seasonally adjusted data

Presenting a time series from which the seasonal movements have been eliminated allows the comparison of data between two months for which the seasonal pattern is different. Also seasonal effects on non-adjusted or original data make it difficult to made valid comparisons over time using these data, particularly for the most recent period. Consequently, seasonally adjusted data are always used in economic modelling and cyclical analysis.

Presentation of data on a seasonally adjusted basis allows the comparison of the evolution of different series which have different seasonal patterns and is particularly pertinent in the context of house price index comparisons since asking price, mortgage approval price indices and actual selling price indices may be in different seasons at identical periods of the year.

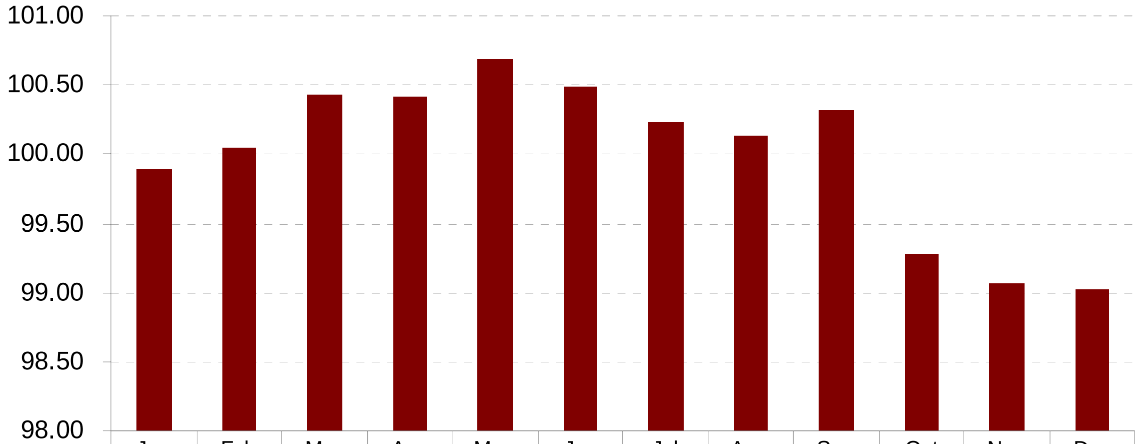
In data analysed by Calnea Analytics both raw and seasonally adjusted figures are shown for a large number of indicators.

## Adjustments for seasonal variation

Seasonal adjustment is calculated using Classical Seasonal Decomposition (Census Method 1) otherwise known as the ratio-to-moving-average method.  The purpose of the seasonal decomposition method is to isolate those components, that is, to de-compose the series into the trend effect, seasonal effects, and remaining variability.  For house price indices and transaction volume analysis we adopt the multiplicative model as the amplitude of the seasonal variation is proportional to the level of the series.

Seasonal adjustment factors are recalculated on a monthly basis.  Those who wish to employ their own adjustment software (e.g. X-12-ARIMA, X11, TRAMO-SEATS etc.) are able to use the non adjusted data provided.

**Seasonal Factors (%) May-2006**




| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Seasonal Factor | 99.89 | 100.04 | 100.44 | 100.41 | 100.69 | 100.49 | 100.23 | 100.14 | 100.31 | 99.27 | 99.07 | 99.02 |

## References


Araham, J.M. and Schauman, W.S. (1991), "New evidence on house prices from Freddie Mac repeat sales", *AREUEA Journal*, 19(3), 333-352.

Bailey, M., Muth, R. and Nourse, H. (1963), "A regression method for real estate price index construction", *Journal of the American Statistical Association*, 58, 933-942.

Baumohl, B. (2005), *The Secrets of Economic Indicators*, Pearson Education.

Case, B., Pollakowski, H. and Wachter, S. (1991), "On choosing among house price index methodologies", AREUEA Journal, 19(3), 286-307.

Case, B. and Quigley, J.(1991), "The dynamics of real estate prices", *Review of Economics and Statistics*, 73(1), 50-58.

Case, K. and Shiller, R. (1987), "Prices of single-family homes since 1970: new indexes for four cities", *New England Economic Review*, Sep/Oct 87, 45-56.

Cho, M. (1996), "House price dynamics: a survey of theoretical and empirical issues", *Journal of Housing Research*, 7(2), 145-172.

Clapp, J. and Giaccotto, C. (1992), "Estimating price indices for residential property: a comparison of repeat sales and assessed value methods", *Journal of the American Statistical Association*, 87, 300-306.

Clapp, J. and Giaccotto, C. (1999), "Revisions in repeat sales price indices: here today, gone tomorrow?", *Real Estate Economics*, 27(1), 79-104.

Clapp, J., Giacotto, C. and Tirtiroglu , D. (1991), "Housing price indices based on all transactions compared to repeat subsamples", AREUEA Journal, 19(3), 270-85.

Costello, G. and Watkins, C. (2002), "Towards a system of local house price indices", *Housing Studies*, 17(6), 857-873.

Fleming, M.C. and Nellis, J.G. (1994), "The measurement of UK house prices: a review and appraisal of the principal sources", *Housing Finance*, 24, November.

Gatzlaff, D.H. and Haurin, D.R. (1997), "Sample selection bias and repeat-sales index estimates", *Journal of Real Estate Finance and Economics*, 14, 33-50.

Goetzmann, W. (1992), "The accuracy of real estate indices: repeat sales estimators", *Journal of Real Estate Finance and Economics*, 5, 5-53.

Goetzmann, W. and Peng, L. (2002), "The bias of RSR estimator and the accuracy of some alternatives", *Real Estate Economics*, 30(1), 13-39.

Greene, W.H. (2003), *Econometric Analysis*, 5th Edition, Prentice Hall, 56-59.

Kennedy, P. (2003), *A Guide to Econometrics*, 5th Edition, Blackwell Publishing, 205-206.

Leishman, C. (2000), Estimating local housing market price indices using Land Registry data", Housing Finance, 47, August, 55-60.

Leishman, C. and Watkins, C. (2002), "Estimating local repeat sales house price indices for British cities", *Journal of Property Investment and Finance*, 20(1), 36-58.

Leishman, C., Watkins, C. and Fraser, W.D. (2002), "The estimation of house price indices based on repeat sales regression using Land Registry data", Report to RICS Education Trust, London.

Meese, R. and Wallace, N. (1997), "The construction of residential housing price indices: a comparison of repeat sales, hedonic regression, and hybrid approaches", Journal of Real Estate Finance and Economics, 14(1/2), 51-74.

Munro, M. & Maclennan, D. (1986) Intra-urban changes in housing prices: Glasgow 1972-1983, *Housing Studies*, 2, pp.65-81.

Steele, M. and Goy, R. (1997), "Short holds, the distribution of first and second sales, and bias in the repeat-sales price index", *Journal of Real Estate Finance and Economics*, 14, 133-154.

Shiller, R.J. (1994), *Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks*, Oxford University Press, Chapter 8.

Wang, T. and Zorn, P.M. (1997), "Estimating house price growth with repeat sales data: What's the aim of the game?", *Journal of Housing Economics*, 6, 93-118.