

# A guide to the compilation of Registers of Scotland statistics

## Contents

<b>1. Introduction and background</b> .....	<b>1</b>
1.1 UK House Price Index.....	1
1.2 Quarterly statistical release.....	2
<b>2. Administrative procedures / background data</b> .....	<b>3</b>
2.1 Registration .....	3
2.2 Land class .....	3
2.3 Cash / mortgage transactions .....	4
2.4 New builds.....	4
2.5 Analysis by region .....	4
<b>3. Data quality</b> .....	<b>4</b>
3.1 Data validation / records excluded from the analysis.....	4
3.2 Data management.....	5
<b>4. Summary of RoS House Type Methodology</b> .....	<b>5</b>
4.1 Assignment of House Type to Sale .....	5
4.2 Allocation of Unassigned Sales.....	6
<b>5. Revision Policy</b> .....	<b>8</b>
<b>6. Summary content table for the HPI and quarterly statistical release</b> .....	<b>8</b>

## 1. Introduction and background

The 2010 National Statistician’s Review of Official House Price Statistics recommended the creation of a “single definitive house price index produced by the official statistics producer community”, replacing the various indices covering the UK at that time. Addressing this recommendation, the producers<sup>1</sup> of official house price statistics worked together to develop the new UK HPI, which is based on a more comprehensive set of data sources covering both cash and mortgage sales.

Registers of Scotland (RoS) has been producing quarterly house price figures for Scotland for many years and, following user consultation, will continue to publish these in parallel with the monthly UK HPI.

This guide, therefore, provides details of the compilation of both the Scotland element of the UK HPI and our existing quarterly statistical publication. The base data used for the publications is comprehensive, capturing all changes of ownership, including cash sales. This base data is then subject to quality assurance (see section 4 for further information about this quality assurance).

The table in paragraph 7 provides a summary of the key differences between the two publications.

### 1.1 UK House Price Index

To deliver the Scottish element of the UK HPI, RoS will extract details of all market value residential sales on a monthly basis. The extraction will be based on the Date of Entry provided in the application for registration and will be published approximately six weeks in arrears, i.e. January data will be first published in mid-March. Additional markers will be added to each line of data to identify mortgage or cash transactions and house type.

<sup>1</sup> HMLR, Registers of Scotland, Land and Property Services Northern Ireland (LPSNI), Office for National Statistics (ONS)

The data will be incorporated by the Office for National Statistics into their HPI model. The HPI applies a Hedonic regression<sup>2</sup> model, into which the following price-determining characteristics will feed:

- Local authority district in GB and housing market area in Northern Ireland
- ACORN area classification variable
- Property type (i.e. detached, semi-detached, terraced, flat, or unassigned if the property type cannot be identified)
- Floor area (metres squared)
- Number of rooms
- New or old property

Information about buyer status (first-time or former owner-occupier<sup>3</sup>) and financing status (cash or mortgage) will also be published each month, but have not been included as price-determining characteristics.

The UK HPI is mix-adjusted to allow for the fact that different houses are sold in different periods by annually updating the fixed basket of properties. The resulting mix-adjustment weights will be calculated annually and will be derived using the latest complete year's worth of transactions.

The average prices will be calculated using a geometric mean<sup>4</sup>, which was deemed to be the most appropriate method for the HPI, since it is less distorted by high values and is in line with international best practice.

Due to the volatility of sales figures for Scotland's local authorities on a month by month basis, particularly in the smaller local authority areas where there may be very few sales, Scotland's local authority figures will be based on 3-monthly moving averages. This will create much smoother figures that are less influenced by high or low value sales within each 3-monthly period.

## 1.2 Quarterly statistical release

Similar to the HPI publication, the RoS quarterly statistical release extracts details of all market value residential sales. However, for the quarterly release, this is done on a quarterly basis and is extracted based on date of registration – the date that the application was received for registration – rather than date of entry. The registration date is generally within a few weeks of the date of sale. The extraction is also limited to applications where the selling price of the property lies between £20,000 and £1,000,000. The quarterly release is published approximately three to four weeks following the end of the quarter to which the data relates.

---

<sup>2</sup> Hedonic regression is widely used for estimating house prices. For a hedonic regression the properties are defined in terms of a set of features or characteristics, each of which contributes to the value of the property. For example, the number of bedrooms or the location of the property will contribute to its value, but these features cannot be priced in isolation. A regression model is used to estimate the value of each of these features from the set of properties sold during a particular period. For example, the model might estimate the effect that every additional bedroom and each different location have in the sale price in a certain month. Then, the price of a particular property can be calculated by combining the values assigned to each of its features.

<sup>3</sup> Data provided from CML.

<sup>4</sup> There are different ways of calculating average prices. These include the arithmetic mean, the geometric mean and the median. The geometric mean involves multiplying the prices together, and then taking the cube (third) root of the total. An arithmetic mean, where all the prices are added up and divided by the number of sales) can result in those areas with the highest priced properties tending to dominate the index, whereas the geometric mean is more equal in its weighting of the change in price of expensive and cheaper properties.

Registrations with a value of over £1m are excluded to ensure that a single large value sale does not distort the average and also to ensure that sales of a block of properties are excluded. Sales of a block of properties that have been identified from the existence of multiple addresses in a title are also excluded.

Similarly, registrations with a value of less than £20k are excluded from the calculation to ensure that, as far as practical, a low value sale, or the sale of part of the title to a residential property, does not distort the average. The transfer of part of a title may happen when one of the joint owners of a shared property buys out the share of the other owner.

Additional markers are added to each line of data to identify house type. Where the house type cannot be identified, additional methodology is applied to place a proportion of these unassigned properties into a particular house type. Further details about the house type methodology are provided in section 4.

Average prices are calculated using arithmetic mean and median, and published to local authority level within Scotland.

## **2. Administrative procedures / background data**

The majority of the information provided in our data extractions for both the HPI and the quarterly statistical release is derived from data gathered as part of the land registration process. The exception to this is the house type methodology, which is covered in more detail in section 4 below.

### **2.1 Registration**

The house price data extractions come from information supplied to RoS with new applications to register a house sale. Registration occurs when a transfer of title takes place, regardless of the amount of money involved. As part of the registration process an applicant's agent (usually a solicitor) records the 'Consideration' for which the title has been reassigned. It is possible for the consideration to be nil or below the true market value of the title. In such a case the applicant's agent (or solicitor) notes the basis of the consideration and also provides a value. Below market price transactions such as these are excluded from the statistics.

The registration process will also provide property address details, date of entry to the property, registration date, title number information, land class, deed type, buyer and seller information and details of any applications for registration of a mortgage over the property. RoS can therefore use the information gathered to identify residential sales, cash sales and new builds and the geographical location of the property.

### **2.2 Land class**

RoS records six land classes: Residential, Commercial, Land, Agriculture, Forestry and Other. The land class information is provided by the submitting agent on the RoS application form when an application for registration is submitted.

Only residential sales are included in the data extractions for the HPI and quarterly statistical release.

As part of our data validation (see section 3.1 below), RoS will also reclassify properties where the land class is incorrect, for example commercial premises that have been classified as residential.

## **2.3 Cash / mortgage transactions**

Every deed submitted to RoS for registration is classified by deed type. It is possible, therefore, to identify if an application for registration of a mortgage security is received at the same time as, or in the weeks following, an application for registration of a residential sale.

A cash sale is defined as a residential sale for which no mortgage application has been received either at the time the sale is registered or within the period of up to three months following the registration of the sale. For example, if an application for registration of a residential sale with no mortgage security deed is received on 20 January and no application for registration of a mortgage security is received in either January, February or March, then the sale will be classed as a cash sale.

Since the HPI will be based on the date of entry rather than the registration date, a deed received for registration in January could have a date of entry in December or earlier. Our revision process (see section 5 below) would mean that a property received for registration on 20 January, but with a date of entry in December would be added to the revised data for December that is published in March. The cash sale status of this property would then be updated in the April publication but would not be updated again.

## **2.4 New builds**

The majority of new build properties are sales of plots from a builder's title. RoS can therefore identify sales from builders' titles and flag them up as new builds. These sales are subject to a quality assurance process. Individual new build plots cannot yet be identified within the registration data, but methods of identification are being investigated.

## **2.5 Analysis by region**

The HPI and the quarterly statistical release provide data down to local authority level. Properties are allocated to local authority area through use of a grid reference applied to an automated Geographical Information System. As a result of incomplete, or wrong, grid references, a small number of properties cannot be allocated to a local authority area. The HPI will exclude these unallocated properties from the analysis, while the analysis within the quarterly statistical release identifies these separately as "Unallocated".

# **3. Data quality**

## **3.1 Data validation / records excluded from the analysis**

Prior to producing the data extractions for the HPI and quarterly statistical release, a validation check is carried out on the database of registrations and the following actions are taken:

- Registrations in which the Consideration is less than the Value are identified so that they can be excluded from the extractions. This ensures that the analysis excludes registrations which are not true, market value sales.
- The land class provided is compared manually with the address data with the aim of ensuring correct classification of residential properties.
- Sales of multiple properties in one application are identified so that they can be excluded from the extractions.
- Duplicate lines of data are identified and removed.
- Dates of entry are checked for obvious errors and missing dates of entry added.
- Any blank fields within the data are identified and populated if appropriate.
- All considerations under £20,000 and in excess of £750,000 are checked for accuracy.



- Right to buy sales of council houses to sitting tenants are included in the final RoS statistics, but the quality assurance aims to identify sales of blocks of properties and sales that are not true, market value sales, e.g. transfers of ownership between family members, sales of pro indiviso shares, etc, so that they can be excluded from the extractions.
- Applications which are subsequently cancelled are identified so that they can be excluded from the extractions.
- Foreign currency considerations are identified so that they can be excluded from the extractions.

### **3.2 Data management**

The spreadsheet-based analysis methodology used to produce the summary tables was developed using procedures recommended to RoS following an independent audit of a pilot methodology. Key features, designed to ensure orderly and consistent outputs, are:-

- Archiving of the base analysis data set for each set of tables showing full details of records included and excluded from the analysis.
- Detailed documentation of processes used for data editing.
- Validated, automated spreadsheet-based processes for producing summary tables from edited base data.
- Inbuilt consistency checks in spreadsheet summaries.

## **4. Summary of RoS house type methodology**

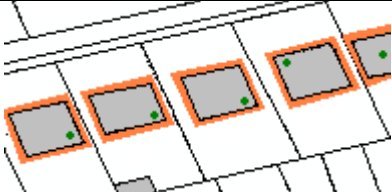

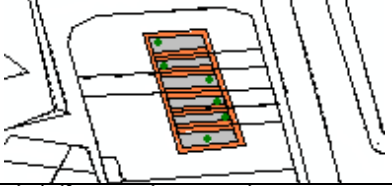

### **4.1 Assignment of house type to sale**

The house type is identified by using a house type classification system developed by RoS. RoS registers all residential property sales in Scotland, and we initially assign one of the four house types using Ordnance Survey (OS) mapping and address products.

The RoS methodology automatically classifies dwelling types according to the four pre-defined types and uses counts of addresses and neighbouring building polygons<sup>5</sup> to classify each property. This method of classification is detailed in the table below.

---

<sup>5</sup> Building polygons are building shapes identified on the OS digital map base.

House Type	Example	Method of Classification
Detached		A building with no adjacent addressed building.
Semi - Detached		A building with 1 adjacent singly addressed building.
Terraced House		A building with one or more singly addressed neighbours.
Multiple Occupancy		Any building with 2 or more unique addresses in any single building.

The majority of house types can be identified using the above method to an accuracy rate of 96.4%. When all sale applications have completed the registration process, the house type can be assigned to one of the four house types by this method.

#### 4.2 Allocation of unassigned sales

Sales which cannot be identified using this method (approximately 15-30% of the monthly total) are initially categorised as Unassigned. The Unassigned House Types fall into three categories as follows:

- **Unknowns:** - Represents approximately 4% of the monthly total. These are predominately New Build properties without OS address points<sup>6</sup>. The house type will be identified when the property transfer has been fully registered by RoS. Cancelled applications are also included in this category.
- **Unaddressed:** - Represents approximately 1% of the monthly total. These are applications with a RoS Title seedpoint<sup>7</sup> which falls within a building we have identified as 'unaddressable' e.g. garages, sheds and buildings without OS address points. These applications need to be examined manually to determine the accuracy of the sales data and OS address point data.
- **None:** - Represents between 10-25% of the monthly total and is linked to the date of the extraction of the application data. The closer the date of extraction is to the date the application was received by RoS the higher this figure will be. Approximately 80% of the "None" category have not yet passed through the Land Registration process.

<sup>6</sup> An "address point" is a point marker used to identify an address within a building on the OS basemap.

<sup>7</sup> A "Title seedpoint" is a point marker used to identify the position of a Land Register Title on the OS basemap.



All of these unassigned categories will be classified as unassigned with the HPI extractions. For the quarterly statistical release, however, a significant proportion (over 95%) of the unassigned categories can be allocated to house type using the following methodology.

All sale applications where RoS **can** identify the house type are split into the following price bands for each Local Authority area. An example of this process is shown in the table below. The respective volumes, average price and proportion of property types are calculated from the previous 12 months of sales.

Price Band	Total Number of Sales	Market Value of Sales	Volume of assigned Sales	Average Price of assigned Sales	Proportion of Assigned Sales in Price Band				Volume of unassigned Sales	Value of unassigned Sales
					detached	semi-detached	terraced	multi-occupancy		
£20,000-£60,000	50	£2,770,000	40	£55,000	5%	10%	40%	45%	10	£570,000
£60,000-£125,000	107	£9,791,000	100	£95,000	5%	10%	55%	30%	7	£291,000
£125,000-£250,000	160	£26,500,000	150	£165,000	20%	20%	30%	30%	10	£1,800,000
£250,000-£375,000	110	£30,550,000	100	£275,000	60%	20%	5%	5%	10	£3,050,000
£375,000-£500,000	58	£22,132,000	55	£400,000	65%	20%	5%	5%	3	£1,320,000
£500,000-£750,000	30	£18,125,000	25	£600,000	70%	25%	4%	1%	5	£3,125,000
£750,000-£1,000,000	15	£17,000,000	10	£900,000	75%	24%	1%	1%	5	£8,000,000

For each price band:

- The proportion of each house type is calculated;
- Unassigned sales are then allocated to each house type using these ratios.

Allocation using this method transfers the volume and average sale price rather than the actual sale price. For example, in the table above:

- There are 10 unassigned sales within the "£20,000-£60,000" price band, with an actual sale price total of £570,000;
- These would be allocated as 1 Detached, 1 Semi-detached, 4 Terraced and 4 Multi-occupancy properties with an estimated sale price total of £550,000 (10 x average price of assigned sales) to maintain the average price of £55,000;
- The actual market value in this band is £2,770,000 (40 x £55,000 plus £570,000);
- However, if we multiply the average price by the volume of sales in this band (now 50 x £55,000) the estimated market value would be £2,750,000, a difference of £20,000.

Therefore, the estimated market value calculated by multiplying the house type average price by the number of sales will differ from the actual market value of all sales in the period. The difference between the estimated value and the actual value indicates the accuracy of the allocation methodology. Please note that if a large proportion of unassigned sales are present the Annual Change in Average Property Prices for Scotland and for each house type will differ significantly and will over/understate the level of decrease/increase.

A small proportion of sales cannot initially be allocated in this way as these properties have no geographical reference or seed point and therefore cannot be matched to a specific Local Authority. As more information is provided to RoS during the registration process the sale can be assigned a house type. These sales are described as Unallocated in the published statistical tables within the quarterly statistical release.

## 5. Revision Policy

Registers of Scotland statistics are governed by the Code of Practice for Official Statistics. One of the principles of the code is to publish "a Revisions Policy for those outputs that are subject to scheduled revisions" and where applicable, to "correct errors discovered in statistical reports, and alert stakeholders promptly".

It is our policy to use all the data available at the time of publication. However this means that previously published quarterly figures are subject to minor revision which does not alter the interpretation that user would draw from the statistics.

The HPI will be updated on a rolling 3-month basis. For example, data published for January will be updated when data for February and March is published. The March data will be the final version of the January HPI.

In relation to the quarterly statistical release, individual statistical reports will not be updated. However, the RoS Quarterly Statistics Time Series will be revised each quarter to ensure that time series data reflects the most up to date data, including taking account of any revisions. The data will only be revised up until one year after the first publication. This Quarterly Statistics Time Series is available at the following link: <https://www.ros.gov.uk/property-data/property-statistics/quarterly-house-price-statistics>.

More substantial revisions which would significantly alter the interpretation of the data will be dealt with on a case by case basis. The action taken may range from publishing guidance to inform users of the revised figures or re-publishing the data.

There are no Registers of Scotland statistics which are subject to regular scheduled revisions.

If you require further details on our methodology please get in touch with the RoS Data Team: [data@ros.gov.uk](mailto:data@ros.gov.uk)

## 6. Summary content table for the HPI and quarterly statistical release

Key features	House Price Index	Quarterly statistical release
Frequency of publication	Monthly	Quarterly
Base data source	Applications for registration in the land register plus house types derived using RoS methodology and base mapping data (Ordnance Survey ©Addressmap)	Applications for registration in the land register plus house types derived using RoS methodology and base mapping data (Ordnance Survey ©Addressmap)
Date range of extraction	Based on the Date of Entry to the property for each calendar month	Based on applications received for registration in the land register within each quarter
Price range	All data with a consideration	£20,000 - £1million
Average house prices	Based on geometric mean – Scotland figures will be based on 3-monthly moving averages to reduce volatility	Based on arithmetic mean and median
Case mix adjustment	Yes – based on room numbers provided via EPC data	No





Key features	House Price Index	Quarterly statistical release
Seasonal adjustment	Yes <sup>8</sup>	No
Revisions	Three month cycle – final figures at 3-month update	The RoS Quarterly Statistics Time Series is updated on a quarterly basis up until one year after the first publication
Volume of sales	Yes	Yes
Property types	Yes – detached, semi-detached, terrace, flat	Yes – detached, semi-detached, terrace, flat
Cash sales	Yes	Cash sales are included but not separately identified
New builds	Yes	New builds are included but not separately identified
Buyer analysis	Yes – statistics relating to first time buyers and former owner occupiers is extracted from CML data	First time buyers and former owner occupiers are included but not separately identified
Results	Average prices presented as geometric means are typically closer to the median than the arithmetic mean	Typically the use of an arithmetic average will lead to a higher price than the use of other measures. This is because an arithmetic approach to calculating the average price takes full account of the extreme value properties that have been sold and as a result the average price can be sensitive to high value property.

<sup>8</sup> The level of geography at which the seasonally adjusted series will be calculated and published has yet to be finalised, and will be assessed once a sufficient period of new data is available from the new HPI to carry out a formal seasonal adjustment review. Initially it will be at government office region level and above.