

Article

The redevelopment of private rental prices statistics, intended methodology: March 2022

Overview of the methodology that we intend to be used for our redeveloped private rental prices statistics.

Contact:
Natalie Jones
hpi@ons.gov.uk
+44 1633 456400

Release date:
25 March 2022

Next release:
To be announced

Table of contents

1. [Main changes](#)
2. [Overview](#)
3. [Existing private rental prices methods](#)
4. [Methods](#)
5. [Researching the methodology](#)
6. [Future developments](#)
7. [Glossary](#)
8. [Related links](#)

1 . Main changes

- We are undertaking a [transformation of our private rental prices' statistics](#), including making better use of existing data sources, improving methods, and developing systems.
- We currently produce the Index of Private Housing Rental Prices (IPHRP) and the Private rental market statistics (PRMS); however, in the future we envisage our rental price statistics to be a new, single, monthly publication that provides a more detailed insight into the rental market.
- The new methodology will use a hedonic regression model; this is similar to the approach used to calculate the UK House Price Index.
- The development outlined is part of a [continuous programme of improvement for consumer price statistics](#); our ambition is to bring in new data sources for further areas of the inflation basket and continue to improve these statistics over the coming years.

2 . Overview

This article provides an overview of the methods that we are expecting to use for our new rental prices estimates. An overview of the project aims and timelines can be found in our [Private Rental Prices Development plan](#).

Work began at the end of 2019, when we first received access to the data that were necessary to begin redeveloping the private rental prices' statistics.

We researched existing methodologies used for rental prices statistics and began developing our future methodology. We have engaged with stakeholders (such as the [Advisory Panels for Consumer Prices](#)) and with international experts in property price statistics, including members from the Economic Statistics Centre of Excellence (ESCoE).

We have also worked closely with Office for National Statistics (ONS) methodology experts to quality assure multiple aspects of the methodology. This engagement allowed us to decide on our methodology towards the end of 2021.

We have now begun the build of a high-quality [reproducible analytical pipeline](#) for our monthly production of price statistics for private rentals, [which we expect to begin publishing in 2024](#).

Please note, the detail within these methods could change as we build a new production system, however, the overall methodology should remain the same.

3 . Existing private rental prices methods

Currently, the Office for National Statistics (ONS) publishes two private rental prices statistical outputs:

- Index of Private Housing Rental Prices (IPHRP)
- Private rental market summary statistics in England (PRMS)

Index of Private Housing Rental Prices

IPHRP measures the change in price of renting residential property from private landlords. A rental price index and its annual percentage change for the UK, its countries and English regions is published.

IPRHP's method is based on a matched-pairs approach, which splits the collected rental transaction data into a sample and a substitution pool. More detail on the methodology can be found within its [Quality and Methodology Information](#).

Existing private rental prices' statistics are used to inform the owner occupiers' housing (OOH) costs element of the Consumer Prices Index including OOH (CPIH), the ONS's lead measure of consumer prices inflation, as well as "actual rentals for housing" aspect of Consumer Prices Index (CPI) and CPIH, and "rent" in the Retail Prices Index (RPI). It is anticipated that the outputs from this development work will eventually be used to measure these.

Private rental market summary statistics in England

PRMS publishes point-in-time arithmetic mean and median rental price estimates for England, English regions and English local authorities.

The current methodology uses all rental transaction data, but limitations prevent compositional changes from being taken into account, so it is not appropriate to compare PRMS estimates year-on-year to infer trends in the rental market and a price index cannot currently be produced. Further detail on the methodology can be found within the [PRMS publication](#).

4 . Methods

Note: The detail within these methods could change as we build the system, however, the overall methodology should remain the same.

We will measure the change in price of renting residential property from private landlords. [Other measures of private rental prices](#), such as those published by Homelet, Rightmove and Zoopla, are produced using only newly advertised rentals, whereas we will produce a measure that reflects both the newly "agreed" rents and existing rents. Therefore, we aim to reflect the stock of rents and not the "flow" of new rents. This is how the current Index of Private Housing Rental Prices (IPHRP) is measured.

The [Johnson Review](#) (2015) points to research that suggests a flow measure may be worth considering; that is, only new lets. We investigated the feasibility of measuring the flow of rents, and we concluded that we do not currently have data sources available to us to disaggregate new rents from existing rents.

The new measures of rental prices bring together several rich administrative data sources. The methodology will use a hedonic regression model, which will allow for mix-adjustment of the monthly price data to control the effect of the changing composition of collected rental properties; this is similar to the approach used to calculate the UK House Price Index. However, the exact detail of each stage in the methodology will be tailored to suit the rental data. These stages are:

1. Quality assurance checks are completed on the data
2. Data are cleaned and property records are linked
3. On an annual basis, a "fixed basket" of properties is created
4. On a monthly basis, data are fitted to a hedonic regression model to quantify the relationship between property characteristics and associated rental price for each calendar month
5. Using the coefficients from the model's output, imputed prices for properties within the fixed basket are calculated
6. Elementary aggregates are produced at a local authority level by taking the ratio of the geometric means of the predicted prices in the base month and the current month
7. Elementary aggregates are weighted together (Lowe index) and then chain-linked annually to produce a rental price index series over time

The corresponding average rental price series is derived by applying the index to a base set of rental prices from the reference period. For example, if the average rental price in the reference period was £500, and the index in the current month was 110.0, a 10% growth would be applied to the reference period average rental price. So, the average rental price in the current period would be estimated at £550. This ensures the price series is consistent with the published index.

Data

The new measures of rental prices bring together several rich administrative data sources. The data sources used fall into two distinct categories: price data and property attributes data. Combining the detailed property attributes data with the price data provides a comprehensive dataset required for use in a hedonic regression model.

The Valuation Office Agency (VOA), Scottish Government, Welsh Government and Northern Ireland Housing Executive (NIHE) deploy rental officers to collect information on the prices paid for privately rented properties, along with some characteristics of the properties. Data for Northern Ireland also include data provided by propertynews.com.

Annually over 450,000 private rental prices are collected in England, 30,000 in Wales, 25,000 in Scotland and 15,000 in Northern Ireland, which make these sources of data rich. Further information on the data sources can be found within our [Index of Private Housing Rental Prices quality and methodology information publication](#).

To strengthen our methods, we are now able to link these rental prices data to other property attributes data (for example, age of the property and floor area), such as from Council Tax data. Separately, we can also link to a geo-demographic segmentation, which will help control for differences in smaller areas.

When a rental price is collected, it will be assumed to be valid either for 14 months from its entry date into the system, or until an update is received. A 14-month validity period will be used as it balances typical contract lengths (which tend to be either 6, 12, 18 or 24 months) against operational practices.

On an annual basis, expenditure weights are calculated to ensure the estimates are representative of the UK. To calculate expenditure weights, dwelling stock data are multiplied with average rental prices. Dwelling stock data come from the Office for National Statistics, Department for Levelling Up, Housing and Communities (DLUHC), Scottish Government, Welsh Government, Department of Finance Northern Ireland and Northern Ireland Housing Executive (NIHE). Dwelling stock estimates are split by the proportion of property types rented privately in Wales, Scotland and the nine regions of England using data from the English Housing Survey and equivalent sources from other countries.

To calculate timely expenditure weights, the most recently available data are used. For a given year, y , the dwelling stock data are based on the period $y-3$, while average prices are based on the period $y-1$.

Valuing a rental property

A regression model is used to estimate the value of each characteristic from the set of properties during a period. For example, the model might estimate the effect that every additional room and each different location have on the rental price in a certain month. Then, the rental price of a property can be calculated by combining the values assigned to each of its features. This method allows us to estimate the prices of properties with every combination of features (such as number of rooms and local authority), even if that combination was not collected in the period.

The price-determining characteristics that we expect to use for England (this may change for other countries depending on property attribute availability) are:

- number of bedrooms
- floor area
- property type (detached, semi-detached, terraced, flat or maisonette)
- furnished status
- geo-demographic segmentation ([ACORN](#))
- local authority
- property age bracket

Mathematically, a semi-log ordinary least squares (OLS) model will be used:

$$\log(p_i) = k + \sum_j \beta_j x_j^i + e_i$$

where:

- p_i is the rental price of property i
- K is a constant
- β_j is the coefficient associated with characteristic j
- x_j^i indicates whether property i has the characteristic j (such as detached property); if so, it takes the value 1, otherwise it takes the value 0 (except for floor area where it takes the floor area in square metres)
- e_i is the statistical error term

The logarithm of the rental price paid is used because rental prices tend to be log-normally distributed, meaning the frequency distribution of the log of the rental price is bell-shaped.

Fixed basket

The rents development methodology is mix-adjusted to control for the fact that different types of rental properties might be collected in different periods. The process of mix-adjustment requires that, in each January, a fixed basket of properties is updated to reflect changes in the composition of rental properties. This basket is then used to produce imputed rental prices for the current year, before the basket is then updated again in the subsequent January.

The fixed basket contains all the rental prices collected in the previous year. If any rental property had been collected more than once in the year, the most recently collected data would be used.

A single univariate decision tree imputation routine is used to impute missing property characteristics in the fixed basket, as recommended by the Editing and Imputation Expert Group within the Office for National Statistics (ONS). Several reasons were given for choosing to impute using a univariate decision tree including that they are fast to implement and re-train, and they are easy to interpret.

Calculating an index

The Ordinary Least Squares model creates coefficients, which are used to calculate an imputed rental price for each property within the fixed basket. These imputed rental prices are then averaged using a geometric mean, which involves multiplying the "n" imputed prices together, and then taking the n^{th} root.

A fuller description of this method and other alternative methods for calculating residential property prices can be found in the [Handbook on Residential Property Price Index](#).

Output

We aim for these measures to be available for the UK, its constituent countries, English regions and local authorities. The new publication will contain:

- an index of private rental prices
- annual rates of change
- average private rental prices
- a breakdown of private rental prices by geography and bedroom category (studio, one bedroom, two bedrooms, three bedrooms and four or more bedrooms)

5 . Researching the methodology

Data cleaning

Within our data, sometimes a rental property can have missing property characteristics. In the fixed basket, a single univariate decision tree imputation routine is used to impute these missing property characteristics. On a monthly basis, any properties with missing characteristics are dropped from the model.

We treat missing property characteristics differently in both scenarios because when testing different imputation options, we found that imputing the sample that is used in our hedonic regression each month using a single univariate decision tree gave differing outputs to other imputation methods. This meant the differences in the indices were driven by the imputation method. The imputation rate is currently low and dropping data points with missing property characteristics is unlikely to introduce bias into the data. This decision will be monitored and reviewed as necessary.

Model choice

Following engagement with experts (Advisory Panel for Consumer Prices - Technical, and international property prices experts), we considered multiple models in our methodological research phase. These included Ordinary Least Squares (OLS), Weighted Least Squares (WLS) and Random Forest approaches. We also considered the use of interaction terms in the model.

By using K-fold cross-validation and generalised variance inflation factor, we decided to use an OLS model without interaction terms. As well as being methodologically robust, the chosen model must also balance different aspects of quality.

Accessibility and clarity

The behaviour of some models are easier to describe, understand and extend. Linear models are well established in existing literature, while others currently require specialised expertise in machine learning.

Coherence and comparability

The House Price Index uses a WLS, and so although we propose using OLS for the rental prices, this does have internal similarity with other housing statistics produced by the Office for National Statistics (ONS). Separately, [StatCan recently moved from a matched pairs model to a hedonic model](#) using a log-linear regression when calculating their rental prices.

Timeliness and punctuality

Models that are easier to troubleshoot and computationally less intensive tend to be easier to implement in the production of official statistics. The monthly production round provides us with around one week to produce, quality assure and evaluate our statistics. More complex models can take more time to run and evaluate.

Longer-term support

It is usually easier to find support for, and to maintain, simpler, well-understood models.

6 . Future developments

We have recently begun the build phase of our production system:

- in February 2022 we published an update to our development plan
- throughout 2022 we will be publishing research articles
- throughout 2022 we will continue the build of the new production system
- we will then begin our user consultation period in 2023, where we will publish an initial analysis of the rental indices
- at the end of 2023 we will publish a decision on the incorporation of these measures into official statistics
- we will publish the first statistical bulletin using the new methodology in 2024, subject to continued systems development

7 . Glossary

Administrative data

Administrative data are data that people have already provided to the government through day-to-day activities, for example, health records, social security payments or educational attainment information.

Interaction term

Interaction effects occur when the effect of one variable depends on another variable. An interaction term within a model accounts for these interaction effects.

8 . Related links

[Private rental prices development plan, UK: updated February 2022](#)

Article | Released 8 February 2022

Overview of our plans for the statistical development of rental prices statistics, including a timeline for development.

[Index of Private Housing Rental Prices Quality and Methodology Information](#)

Methodology | Released 9 November 2018

Quality and Methodology Information for Index of Private Housing Rental Prices, detailing the strengths and limitations of the data, methods used and data uses and users.

[House Price Index Quality and Methodology Information](#)

Methodology | Released 17 November 2021

Quality and Methodology Information for the UK House Price Index, detailing the strengths and limitations of the data, and methods used.

[Transformation of consumer price statistics - Office for National Statistics](#)

Article | Released 9 November 2021

Our plans to transform UK consumer price statistics by including new improved data sources and developing our methods and systems for production from 2023.