ADVISORY PANEL ON CONSUMER PRICES – TECHNICAL

Redeveloping Private Rental Market Statistics

Status: Final
Expected publication: Alongside minutes

Purpose

1. This paper presents the progress on and proposed methodology for the production of new price statistics for the private rental market. These statistics will be used in future to produce the owner occupiers' housing costs (OOH) element of Consumer Prices Index including OOH (CPIH). The OOH measure accounts for around 17% of CPIH, ONS's headline measure of inflation.

Actions

2. Members of the Panel are invited to:
   a) comment on the suitability of each proposed hedonic regression model
   b) identify their preferred hedonic regression model for utilisation in the development of private rental market statistics

Background

3. Currently, the Office for National Statistics (ONS) publishes two private rental prices statistical outputs: the UK Index of Private Housing Rental Prices (IPHRP) and Private rental market summary statistics in England (PRMS).

4. IPHRP publishes a monthly index of private rental prices and annual percentage change for the UK, its countries and English regions. PRMS are point-in-time (twice yearly) rental price estimates for England, English regions and English local authorities. Current methodology limitations prevent compositional changes from being taken into account, so it is not appropriate to compare PRMS estimates to infer trends in the rental market and a price index cannot be produced.

5. The objectives of the rents development project are:

   a. To ensure published rental statistics remain relevant to users and better meet user needs by producing statistics that include average rental price and change in rent over time at increased geographical granularity

   b. To produce a fully consistent set of measures of housing market statistics, which will inform OOH, using the latest available data sources and the most robust available best practice methods

   c. To reduce time and resource required to publish private rental market statistics by transforming two publications (IPHRP and PRMS) into one over-arching publication, which will in the long-term, enable further efficiency increases by consolidating all housing market statistics into one unified publication

6. The new publication will make better use of available data sources to produce private rental prices statistics comparable over time and down to lower geographic levels than currently

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

a. An index of private rental growth over time

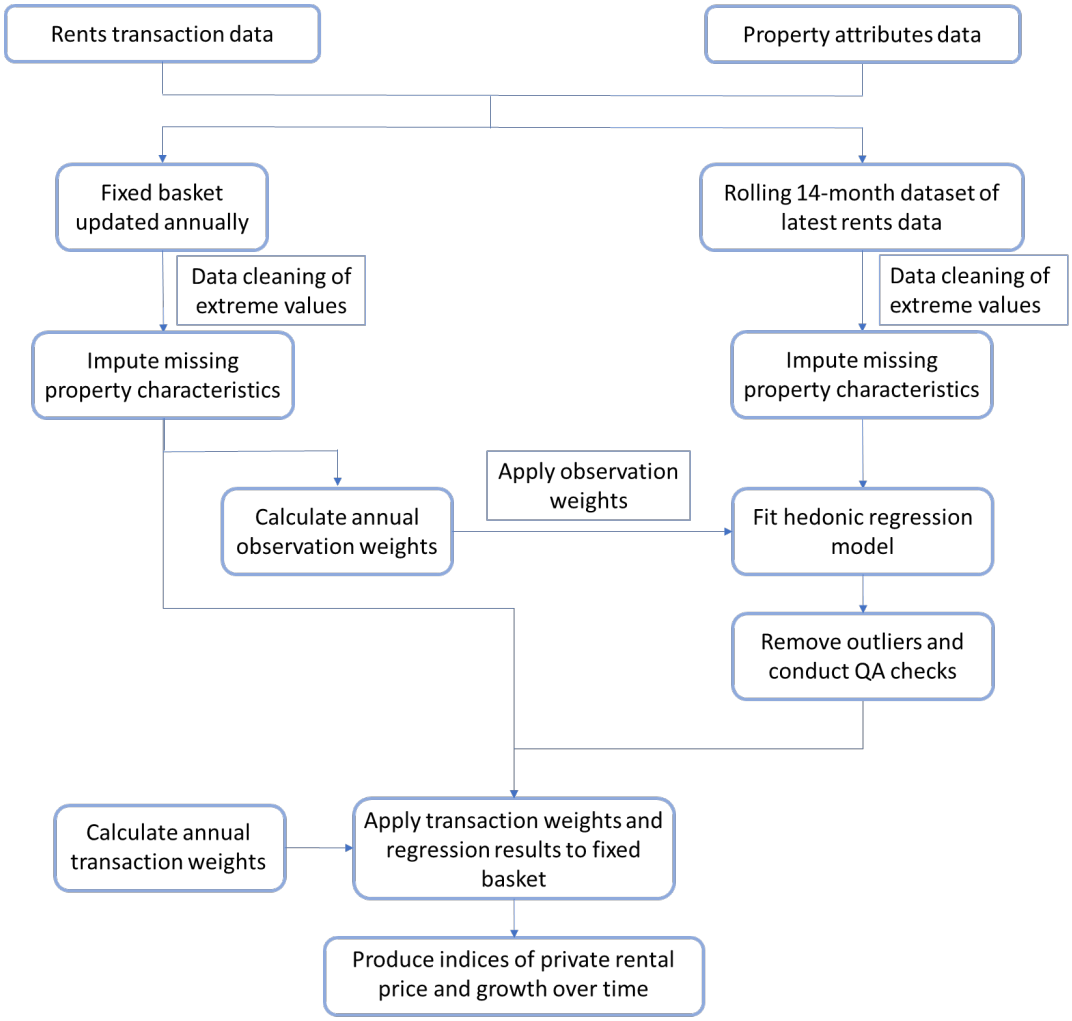
b. Annual percentage change over time

c. Private rental price over time

d. A breakdown of private rental price by geography and bedroom category (studio, one bedroom, two bedrooms, three bedrooms, and four or more bedrooms)

Proposed potential methodology

Figure 1: Proposed Rents Development methodology



7. A summary of the proposed methodology is shown in Figure 1. Key stages are:

   a. Link property-level price data with property attributes and location data

   b. Populate missing data using a univariate decision tree imputation, as recommended by ONS's Editing and Imputation Expert Group

   c. Model dependence of rent price on property characteristics using a hedonic regression model, similarly to HPI methodology. The hedonic regression is run each month on the latest 14-month rolling dataset. Identified key property characteristics include:

i. Number of bedrooms (England & Wales); number of rooms (Scotland & Northern Ireland)
ii. Floor area (in m^2)
iii. Property type (Flat/Maisonette, Bungalow/House:Detached, Bungalow/House:Semi-detached, Bungalow/House:Terraced)
iv. Furnished status (Y/N)
v. ACORN Group classification
vi. Local Authority property resides in
vii. Property age (England, Wales & Northern Ireland)

d. Elementary aggregates (combinations of furnished status, local authority and property type) are produced using geometric means
e. Elementary aggregates are weighted together and then chain-linked to produce a rental price index series over time for the UK, its countries, English regions and local authorities and breakdowns
f. The corresponding average rental price series is derived by applying the index to a base set of rental prices from the reference period. This ensures the price series is consistent with the index – a key requirement

8. Following consultation with ONS' internal regression methodology experts, and after actioning feedback from prior consultation with the APCP-T Panel in January 2021, there are currently three alternative regression models under consideration for Rents Development, on which we seek feedback from the APCP-T Panel.

Proposed regression models

Model 1 – Weighted Least Squares without interaction terms

Regression formula:

$$ln(rent) \sim bedrooms + ln(area) + type + furniture + acorn\ group + local\ authority + age$$

Model fit: R^2 ~ 83%

~96% of 375 regression coefficients are statistically significant with a p-value < 0.01.

Advantages:

- Easy to interpret model outputs
- Well recognised regression technique and already utilised in the UK HPI
- Transparent methodology and easy to explain to users
- Resultant England index agrees very closely with the England index produced by Model 2. This is worth noting because Model 2 is able to account for interactions, while Model 1 does not.

Disadvantages:

- Interactions between explanatory variables are not accounted for by this model

Model 2 – Random Forest with shrinkage

In the random forest model, 200 decision trees are ran, each with a maximum of 20 levels (questions). The model rent price is predicted from the average rent price of those 200 trees. The model decides how to split the data at each level by choosing the split that results in the greatest decrease in model error. Shrinkage is applied to prevent over-fitting; a minimum of 50 observations is required for a node to split to a further level and it is required that, upon a split, there must be a minimum of 25 observations in each of the resultant nodes.

Model fit: R^2 ~ 85%

Advantages:

- Resultant England index agrees very closely with the England index produced by Model 1
- Important interactions between explanatory variables should be accounted for by this model
- Removes the assumption of linear models that the relationship between the dependent variable and explanatory variables are linear and independent
- In general, random forests are able to better learn relationships in complex data compared to linear models and so have greater predictive power/accuracy

Disadvantages:

- The operation through which predictions are made is more of a "black box" than Model 1 or Model 3 and it is difficult to explain to users how a random forest works
- Unlike WLS models, random forests do not output regression coefficients because property characteristics contribute to rent price differently for every property. So it is difficult to interpret contributions to the rent price from different property characteristics in a random forest model, unlike regression coefficients
- The model splits the nodes in such a way as to result in the greatest decrease in model error. However, it is difficult to know in practice what split decisions are made and the in-tree decisions made to obtain one month's fit may be completely different to the decisions made in the following month
- Difficult to visualise the random forest due to large tree depth and many trees
- Although use of random forests is established in academic research, random forests are less commonly used for production of official statistics compared with general linear models and may therefore be considered more experimental

Model 3 – Weighted Least Squares with some interaction terms

Regression formula:

$$ln(rent) \sim bedrooms + ln(area) + type + furniture + acorn\ group + local\ authority + age + acorn\ group * bedrooms + acorn\ group * \ln(area) + acorn\ group * type$$

Model fit: R^2 ~ 84%

~87% of 524 regression coefficients are statistically significant with a p-value < 0.01. The percentage of significant combinations of the acorn_group*bedrooms interaction varies between months.

Advantages:

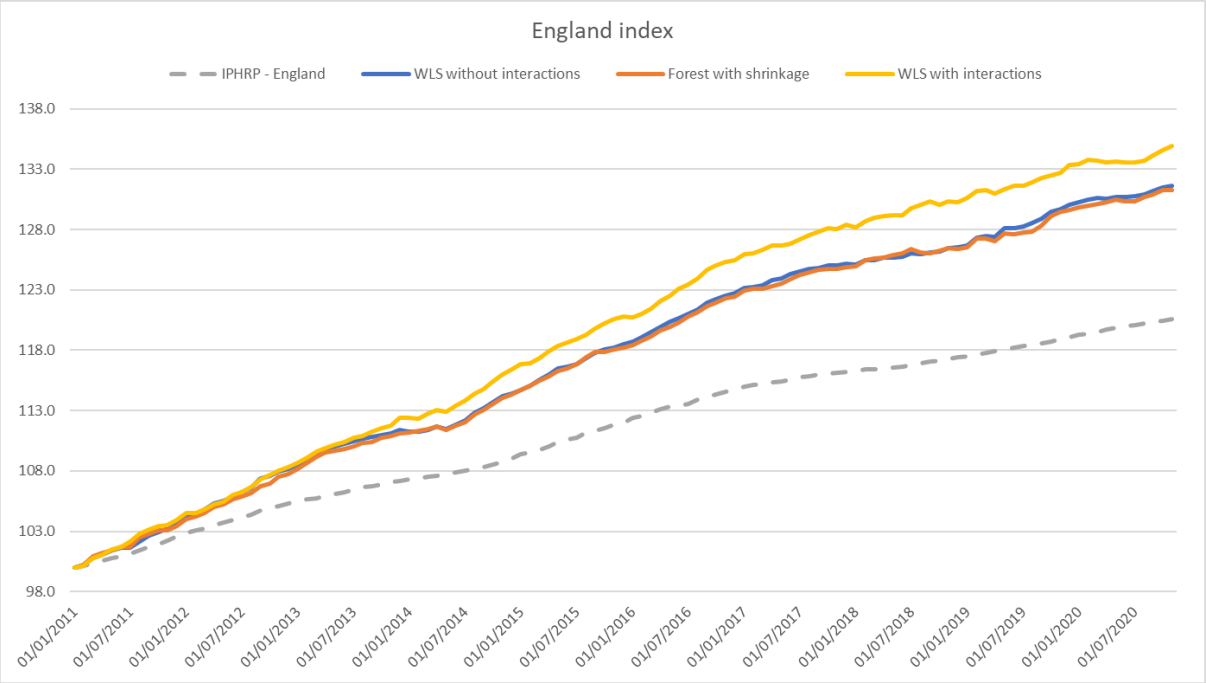
- Accounts for statistically significant interactions between some explanatory variables
- Easy to interpret model outputs
- Well recognised regression technique and already utilised in the UK HPI, although no interaction terms are used in the UK HPI (see Model 1)
- Transparent methodology and fairly easy to explain to users

Disadvantages:

- Difficult to interpret and explain the contribution of interaction terms to the rent price to users
- Not all possible interactions between explanatory variables could be tested because there are over 300 local authority categories, 18 acorn group categories and 10 age categories. Memory errors are obtained if it is attempted to include more interaction terms. It is likely not all statistically significant interactions are included in this model
- The included interactions increase the long-term growth, but when no interactions are included in the WLS model, the result agrees more closely with the random forest model. Since the latter model should be accounting for most important interactions, it is unclear if inclusion of some, but not all interactions is detrimental
- The regression coefficient values may not be stable from month to month
- The resultant index from this model is more different than the other two models. This model consistently predicts higher growth for all English regions and Wales using this model

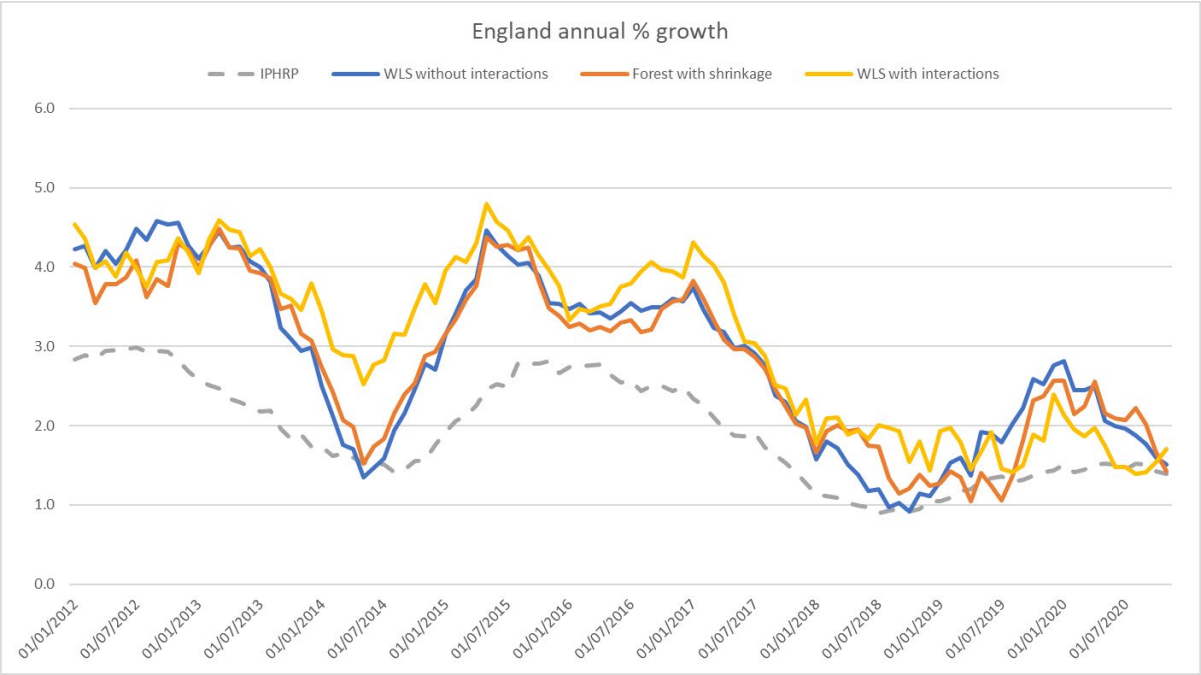
Preliminary results comparing proposed regression models

England



Over nearly 10 years to November 2020, IPHRP estimates a growth of 20.6% for England. This is lower than the growth predicted by all Rents Development models. WLS without interactions, Random Forest with shrinkage and WLS with interactions estimate a growth of 31.6%, 31.3% and

34.9% for England, respectively. This is an additional 1.1%, 1.1% or 1.5% growth per year for England compared with IPHRP.

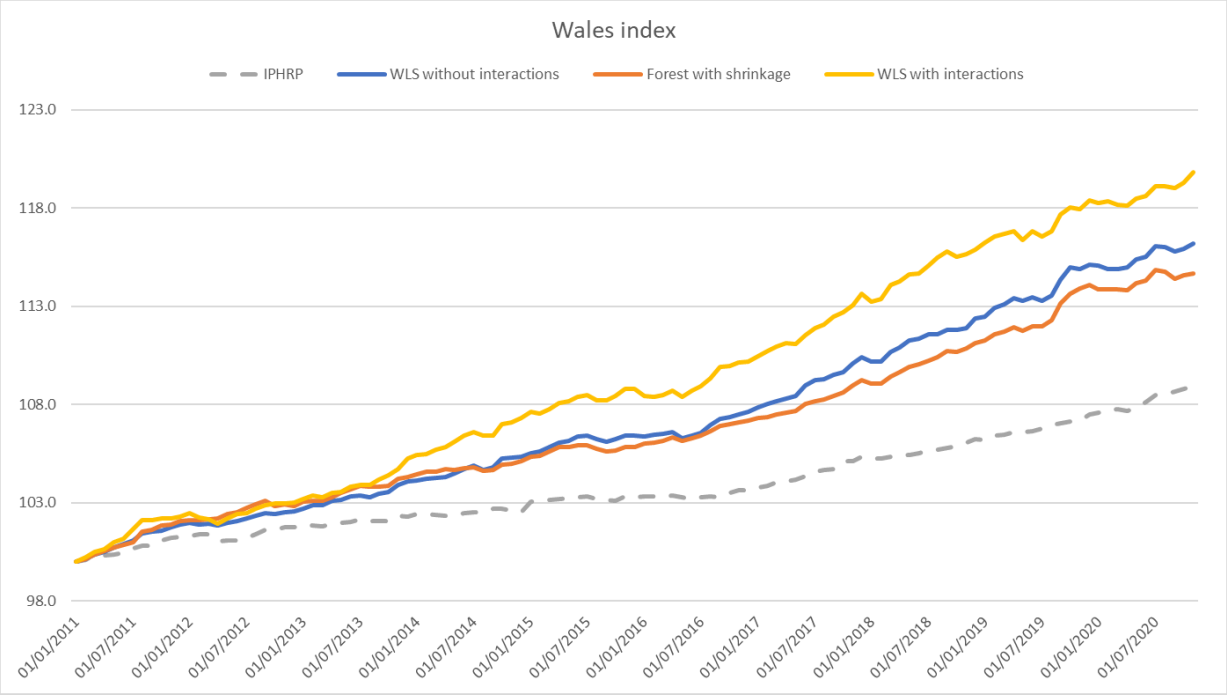


The England indices resultant from using WLS without interaction terms and the Random Forest model are in close agreement. All three models show a similar pattern in annual percentage change and this pattern is broadly similar to the pattern shown by IPHRP. However, although Rents Development models predict similar annual percentage growth in times of lower growth, Rents Development predicts higher annual growth than IPHRP in times of higher growth.

Rents Development indices are more sensitive to changes in the market than IPHRP. This is likely because all collected price data is utilised in the latest month's measure in Rents Development, whereas IPHRP's matched pairs approach results in around 40% of the latest price data being utilised in the latest month's measure (increasing to ~60% during the year due to substitutions). This means that a large proportion of the latest price data is not included in IPHRP's latest estimates and it takes longer for changes in the private rental market to be observed in IPHRP compared with Rents Development's measure.

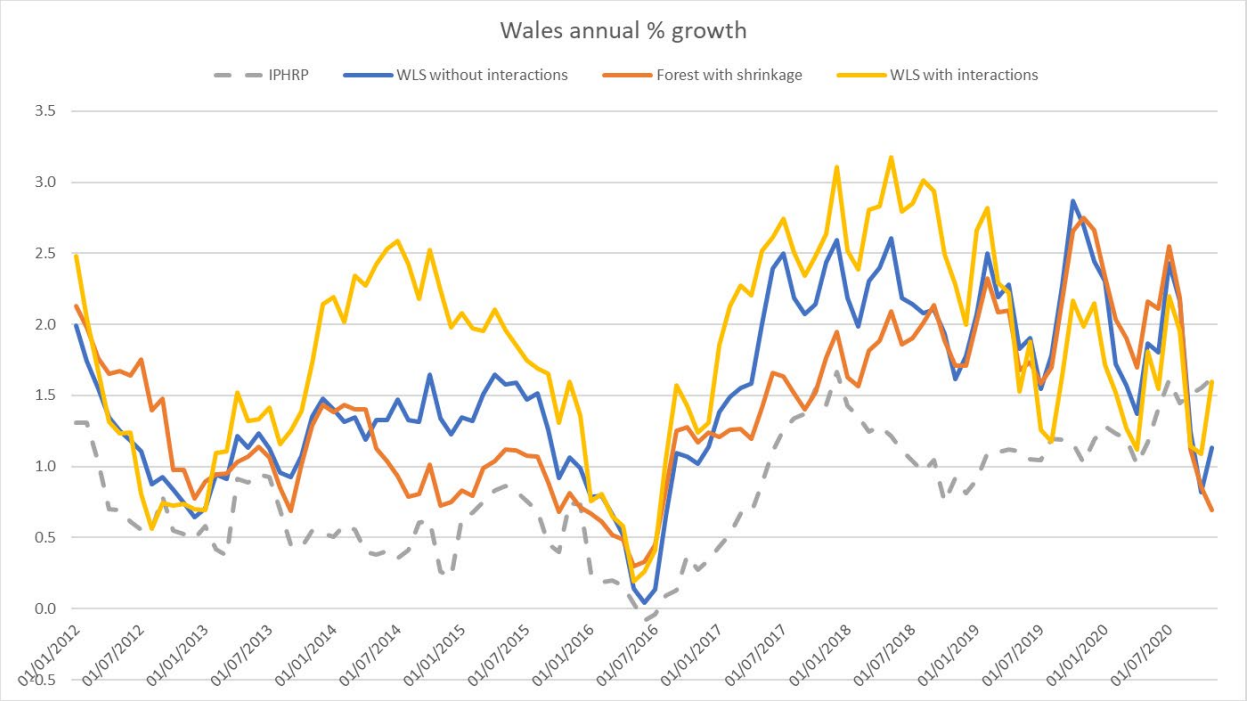
For instance, during the Coronavirus-19 pandemic in 2020, the annual percentage growth in England remains steady in IPHRP. However, in Rents Development it is shown to decline from 2.1-2.8% in Jan-2020 to 1.4-1.7% in November 2020. This slowdown in annual growth aligns with private measures of rent prices. From the England index, the decline in annual growth was not driven by falling England prices, but by a slowdown in rent price increases. Rent prices largely stagnated from April 2020, before starting to rise again from July/August 2020.

Wales



Over nearly 10 years to November 2020, IPHRP estimates a growth of 8.9% for Wales. This is lower than the growth predicted by all Rents Development models. WLS without interactions, Random Forest with shrinkage and WLS with interactions estimate a growth of 16.2%, 14.7% and 19.8% for Wales, respectively. This is an additional 0.7%, 0.6% or 1.1% growth per year for Wales compared with IPHRP.

The increase in rent price growth in Wales from mid-2016, observed in Rents Development and IPHRP, is driven by rising prices in Cardiff. The greater increase in growth observed in Rents Development can be partly attributed to weighting. In IPHRP, weights are calculated at region level for England and at Wales level for Wales. Since Cardiff is under-sampled relative to other local authorities in Wales, IPHRP under-represents Cardiff in its estimates. Therefore, the contribution of rising Cardiff prices to Wales' growth is under-reported in IPHRP. In contrast, Cardiff is more accurately represented in Rents Development since weights are calculated at local authority level, and consequently the full impact is captured in Wales' price index.



There is a larger difference between the WLS without interactions and Random Forest models for Wales than for England. However, these two models' estimates are still in much closer agreement than with WLS with interactions. Again, all three models show a similar pattern in annual percentage change and this pattern is broadly similar to the pattern shown by IPHRP.

The impact of the Coronavirus-19 pandemic is more difficult to assess in Wales for both IPHRP and Rents Development as the index displays more volatility. In general, Rents Development predicts that annual percentage change declined during 2020 with signs of a possible recovery in November 2020, while IPHRP reported annual growth remain fairly stable between January and August 2020 (slight decrease in April 2020 before recovering by June), followed by an increase in annual growth.

Aimee North
Prices Division, Office for National Statistics
February 2021