**Economic research**

# Covid-19 and house prices in the UK: what can be learned from web-scraping data?

Official statistics on the housing market generally provide aggregated and/or lagged data. In addition, a lot of freely accessible granular information can be collected in real time from real estate websites using web-scraping techniques (automated retrieval of Internet data). Using the United Kingdom as an example, this article illustrates the value of a daily, granular and real-time database, particularly for analysing the behaviour of sellers. These data confirm that during the first lockdown period, activity stalled and sellers adopted a wait-and-see attitude. The granular approach also reveals significant regional differences: while advertised prices were stable or even increased after the first lockdown in more rural areas, they declined continuously in London.

**Jean-Charles Bricongne, Nicolas Chatelais, Baptiste Meunier\* and Sylvain Pouget\*\***
**Directorate General Statistics, Economics and International**
Economic Affairs and International Cooperation Directorate

**Jérôme Coffinet\* and Etienne Kintzler**
**Directorate General Information Technologies**
Data and Analytical Services Directorate

\*Directorate at the time this article was written.
\*\*Internship at the Banque de France when this study was carried out.

JEL codes
E01, R30

## 80%
decline, at least, in the number of new property listings in the first UK lockdown

## 90%
of listings still available after one month on the Internet at the height of the first lockdown, compared to about two-thirds outside it

### New weekly listings
(number)



Sources: Zoopla and authors' calculations.

# Bulletin
de la Banque de France

235/5 - MAY-JUNE 2021

Economic research

BANQUE DE FRANCE
EUROSYSTÈME

2

## 1 Data from real estate websites: a useful supplement to official real estate statistics

**Using alternative data makes it possible to construct indicators more rapidly** (real-time daily indices vs. monthly/quarterly with a lag) and with **greater granularity** (e.g. at the postcode level) than is possible with official statistics. In this case, thanks to the higher level of granularity; we can detect imbalances in local housing markets. From a broader perspective, this article reflects recent attempts by economists to construct high-frequency indicators (e.g. Lewis et al., 2020). The availability of these indicators has become especially urgent due to the Covid-19 crisis and the sudden fluctuations in activity (see for example Bricongne et al., 2020). Indeed, it is even more important to make use of these data in real time during crisis episodes such as the Covid-19 pandemic, as the delays in the publication of traditional statistics means that economic downturns are only captured with a lag.

For instance, a large amount of data is publicly available in real time on real estate websites. This is particularly true for the UK housing market where 92% of estate agents advertise on Internet.[1] The economic literature includes examples of the use of alternative data to monitor the housing market, such as Kulkarni et al. (2009) with Google Trends or, more recently, the use of web-scraping to analyse the housing market (e.g. Hanson and Santas, 2014 or Bricongne et al., 2019). This paper focuses on the sellers' perspective (which is precisely what is achieved by real estate web-scraping, where the data are posted by the sellers, as opposed to final transaction data that result from the interaction between sellers and buyers) and derives indicators from their perspective. It covers the UK market in this unique period of Covid-19 and Brexit – although this approach can be transparently extended to other geographical areas.

**Official statistics on residential real estate are relatively incomplete and available with a lag.** In the United Kingdom, the main indicators (price indices, construction sector surveys, new build index, detailed transaction data, etc.) are generally published monthly, at least 10 to 15 days after the end of the month, or more depending on the source and type of data. This lag time is generally longer in most other OECD countries, which would make alternative statistics even more useful. Furthermore, in the UK, statistics are not always available at a very detailed sub-national level, while discrepancies between urban – notably London – and rural areas have been identified in the literature (e.g. Poon and Garratt, 2012). However, individual transaction data may be available, but this is still limited to a few countries, including France.[2]

## 2 Two important innovations: real-time monitoring and monitoring from the seller's perspective

Using data from web-scraping, **daily statistics** on the housing market are available ahead of official statistics. Interestingly, location data can also be used to construct these indicators at a very granular level. This makes it possible to monitor several common indicators (e.g. prices, number of new construction projects) with improved speed and maximum granularity compared to official statistics.

Furthermore, constructing data from listed/online offers gives the **sellers' perspective** – which is not the case with the usual statistics based on transaction data from comparing supply and demand. This innovative approach allows us, for example, to monitor the **number of new listings** which, adjusted for seasonal variations, indicates whether sellers wish to put their properties on the market.[3] A deviation from its "normal" value would mean that sellers are anticipating (or facing) a housing market shock. These data also enable us to monitor **changes in the price of a listing over time**. A deviation from the "normal" adjustment path (because even under stable economic conditions, sellers can adjust their prices over time to adapt their initial valuations) could signal an

---

1 Source: *National Association of Realtors,* available at this link: *https://www.nar.realtor/sites/default/files/documents/2018-real-estate-in-a-digital-world-12-12-2018.pdf*
  The coverage of the sites used may vary from region to region, but in all cases it is sufficiently high to ensure good representativeness at sub-national level.
2 See: *https://app.dvf.etalab.gouv.fr/*
3 It is also possible to ascertain the buyers' point of view through, for example, Internet searches.

# Bulletin
de la Banque de France
235/5 - MAY-JUNE 2021

**Economic research**

BANQUE DE FRANCE
EUROSYSTÈME

3

BOX

## A very granular daily database

The data are extracted through a daily automatic upload of data on the Internet from the major real estate classified ad sites in the UK: Rightmove, Zoopla and OnTheMarket. To improve coverage of more specific regions, PropertyPal – a property site specialising in Northern Ireland – and S1Homes, its counterpart for Scotland, are also included. Through these websites, more than 1.5 million real estate listings are downloaded on average every day, corresponding to the total stock of available listings. About two-thirds of these are sales offers (see table). Zoopla data have been extracted since the beginning of March 2020, while data from other websites have been available since July 2020.

### Average number of web-scraped listings per day
(per single URL)

| Name of site | Residential | | Commercial | |
| --- | --- | --- | --- | --- |
| | Sale | Rental | Sale | Rental |
| Zoopla | 300,000 | 200,000 | 15,000 | 35,000 |
| Rightmove | 500,000 | 200,000 | 20,000 | 25,000 |
| OnTheMarket | 250,000 | 100,000 | 5,000 | 10,000 |

Sources: Websites Zoopla, Rightmove and OnTheMarket; authors' calculations.
Note: The URL, or Uniform Resource Locator, commonly refers to an Internet address, in this case corresponding to a listing ad.

For each listing, information describing in detail the dwelling, the type of transaction and the price, and especially its location, is retrieved at a very granular level. Information may differ between sites, but for the main provider Zoopla, the following information is available for residential listings: city, postcode, address, size, type of transaction (sale, rental), type of building (new or existing), number of bedrooms, number of bathrooms, number of living rooms, general description (e.g. "nice flat with terrace in a safe area"), and type of property (e.g. flat, duplex, freehold, etc.). By having access to detailed information, not only objective, but also possibly more subjective, it is possible to better identify the factors that can affect the price of a listing (location, type of property, neighbourhood, etc.). In particular, the general description contains keywords that refer to additional facilities (e.g. "balcony", "terrace", "garage") that may increase the price of the property.

Once the data are extracted, they are quality-checked. Data on surface areas are missing for a significant share of listings of certain types of property. The surface area is usually given for higher quality dwellings likely to be of interest to international buyers, such as large flats in London. Due to this selection bias, the authors do not use surface area data and focus on absolute house prices rather than prices per square metre, as is the case in most official statistics. It can therefore be implicitly assumed, when analysing changes in absolute prices, that average surface areas change little over time, even if certain atypical episodes such as lockdowns can lead to possibly non-negligible composition effects. Any duplicates that may exist on the same site are also stripped out. For dwellings with the same address, description and price, only one listing is kept. However, this filtering is only carried out for existing dwellings. In the case of new dwellings, an apparently similar listing may correspond to different properties because several dwellings with comparable features may be posted simultaneously on the same site.

…/…

# Bulletin
de la Banque de France

**235/5 - MAY-JUNE 2021**

**Economic research**

BANQUE DE FRANCE
EUROSYSTÈME

4

The data are also filtered according to their category because we chose to focus on residential real estate and not include commercial real estate, even though this is feasible and useful, at least for some segments, given the financial stability issues. Commercial real estate has distinct characteristics and cycles and gives rise to fewer observations. Assets such as car parks, mobile homes or land are also excluded. Furthermore, a number of properties are sold by auction in the United Kingdom: in this case, the advertised price is in fact the reserve price. Therefore, such listings are not included. Other outliers are stripped out by winsorising[1] at 1%.

1 The statistical procedure for reducing the impact of extreme or unrepresentative values in the estimation of a distribution parameter of the mean or standard deviation.

episode of stress. This new approach is based on listings, i.e. at an early stage in the process of a property sale, which should in principle allow for a faster detection of market turnarounds. Conversely, official statistics are compiled using completed transactions. In extreme cases, if sellers find it more difficult to sell their homes, transaction data could become unavailable while plentiful information on downward adjustments to the prices of listed properties would still be available. It should be noted, however, that part of the price adjustment can occur at the time of the transaction, reflecting the negotiation margin, which provides information on the state of the market and which can be reconstructed by matching, as far as possible, listings and transactions (see below).

Statistics calculated from the seller's point of view should therefore not necessarily be compared with official statistics, as they provide a complementary viewpoint.[4]

Furthermore, the innovations of real-time tracking and the seller's perspective are not specific to the UK case, as they would be applicable to most other countries.

## 3 Illustration with the UK residential housing market during the Covid-19 crisis

A first illustrative indicator is the number of new listings – which reflects whether sellers are selling their properties at the right time. Chart 1 below shows the weekly number (which has the advantage of eliminating seasonal variations within a week) of new listings. There was a wide variability in this number, which was five to six times lower during the (first) lockdown than after or before it.

Another indicator is the number of listings still available after one month on the Internet (see Chart 2): the impact of lockdown is visible with around 90% of listings published at the end of March/beginning of April still available one month later. After lockdown, the proportion gradually decreases and stabilises around its pre-lockdown level, i.e. about two-thirds of the total.
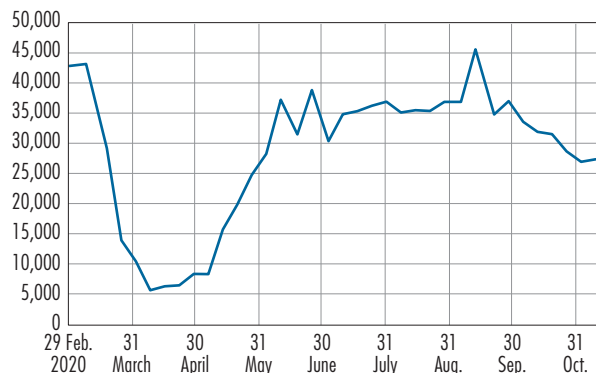
The granularity of the data makes it possible to analyse prices at the local level. Chart 3 shows price trends in four regions of England: the east of England, the

4 For a possible comparison between the two approaches, it should also be borne in mind that the figures published in the UK generally show median rather than average developments, which can lead to significant differences, especially in areas such as London with a high proportion of expensive properties. A calculation of median prices from web-scraping data provides orders of magnitude closer to the official aggregated figures.

# Bulletin
### de la Banque de France
**235/5 - MAY-JUNE 2021**

**Economic research**

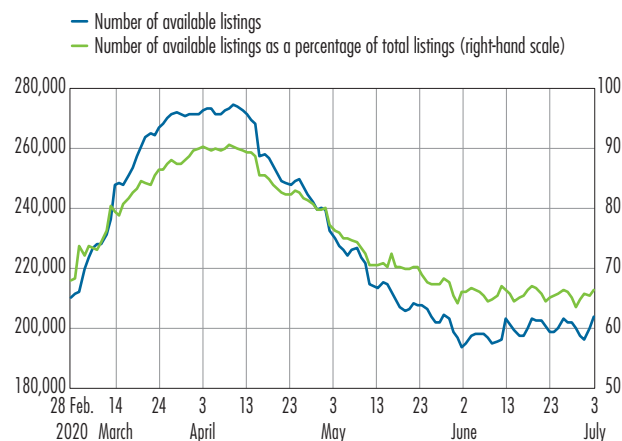BANQUE DE FRANCE
EUROSYSTÈME

5

## C1 New weekly property listings

(number)



Sources: Zoopla and authors' calculations.
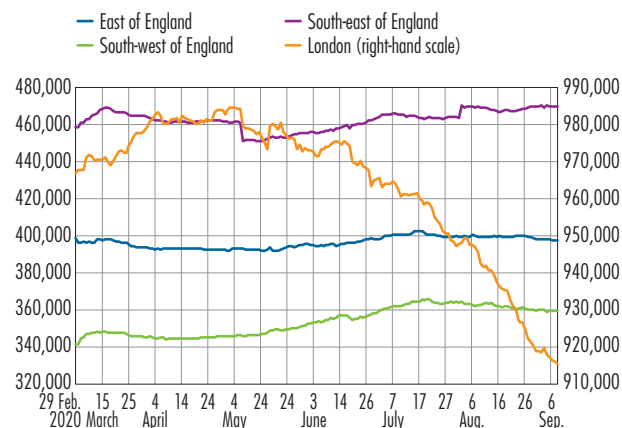
## C2 Listings still available after one month on Internet

(in units and as a %)

— Number of available listings
— Number of available listings as a percentage of total listings (right-hand scale)



Sources: Zoopla and authors' calculations.

## C3 Average listing price by region

(in GBP by property)

— East of England      — South-east of England
— South-west of England   — London (right-hand scale)



Sources: Zoopla and authors' calculations.

south-east, the south-west and London. While the trends are relatively similar for the first three, with a certain stability or even a slight upward trend, London is different, with a continuous decline since May. As these prices are absolute and not per square metre, this decline may reflect a decrease in the listed price per square metre, a compositional effect with properties sold that are smaller or of lower quality, or a combination of both effects. However, it is reasonable to assume that these compositional effects remain relatively limited in magnitude over fairly short periods of time, and in any case smaller than the spectacular trends shown in Chart 3.

The prices displayed in Chart 3 are average prices, whereas official published data are generally closer to median prices. The aim of this study is not to reproduce the official statistics, but rather to supplement them. Nevertheless, a calculation of median prices would yield orders of magnitude close to the published statistics. It should be noted that the price trends are identical when using median or average prices. Similarly, to offset compositional effects between different types of property (studio flats, two bedroom flats, etc.), price indices by type of property are calculated: during the Covid-19 crisis, trends are also comparable between property categories.

Using scraped data, it is also possible to analyse how the price fluctuates for each listing. For instance, for each listing on a given day, the authors check whether it is still available the next day and whether its price has changed; the same process is then repeated on a daily basis. Comparing the listings for which the price changes to the total number of listings, it seems that only a minority (between 8 and 10%) experience price changes, 90% of which are downward adjustments. Chart 4 below provides an overview of the share of listings undergoing price changes after 7, 14, 21, 31 and 45 days on the market. This chart shows that at the beginning of lockdown, sellers were not yet ready to change their prices and preferred to adopt a wait-and-see approach: a significantly lower share of listings underwent price changes compared to the pre-lockdown situation, even after a prolonged period. However, after the first lockdown, the share of listings undergoing a price change (mainly downward adjustments) increased rapidly and stabilised at higher levels than in the pre-lockdown period, especially for listings that had been on the market for longer periods.

# Bulletin
de la Banque de France
**235/5 - MAY-JUNE 2021**

**Economic research**

BANQUE DE FRANCE
EUROSYSTÈME

7

# References

Bricongne (J.-C.), Coffinet (J.), Delbos, (J.-B.), Kaiser (V.), Kien (J.-N.), Kintzler (E.), Lestrade (A.), Meunier (B.), Mouliom (M.) and Nicolas (T.) (2020)
"Tracking the economy during the Covid-19 pandemic: the contribution of high-frequency indicators", *Banque de France Bulletin*, No. 231/5, September-October.
*Download the document*

Bricongne (J.-C.), Turrini (A.) and Pontuch (P.) (2019)
"Assessing house prices: insights from 'Houselev', a dataset of price level estimates", *European Commission Discussion Paper*, No. 101.

Hanson (A.) and Santas (M.) (2014)
*Field experiment tests for discrimination against Hispanics in the US rental housing market, Southern Economic Journal,* Vol. 81, No. 1, pp. 135-167.

Kulkarni (R.), Haynes (K.), Stough (R.) and Paelinck (J.) (2009)
"Forecasting housing prices with Google Econometrics", *GMU School of Public Policy Research Paper,* No. 2009-10.

Lewis (D.), Mertens (K.) and Stock (J.) (2020)
"Monitoring real activity in real time: the weekly economic index", *Federal Reserve Bank of New York Liberty Street Economics*, No.20200330b.

Poon (J.) and Garratt (D.) (2012)
"Evaluating UK housing policies to tackle housing affordability", *International Journal of Housing Markets and Analysis*, Vol. 5, No. 3, pp. 253-271.