The Whenfresh house price indices are different from all other indices in the UK market. They are different in many ways and for many reasons. This document will explain these differences and the reasons for them, and why the indices are extraordinarily well suited for valuations.

## Price indexing

Indexing, at its simplest, applies a numeric description of how nominal prices for a constant good or service vary across space or time. When applied to residential real estate, at a high level, the function is quite simple:

$$V_t = \frac{i_t}{i_0} \cdot V_0$$

The value at any time is a product of the original value and the ratio of index values at present and the time of the original value.

While theoretically sound, this method makes many critical assumptions which are not always perfectly true in messy reality. Some of these **assumptions** are:

### 1: The object in question has not changed over time.

This is rarely true, in reality, for any good: Objects, if used, suffer wear. An exception might be special collectible objects which are preserved in their condition, but even then, special care is taken to prevent the otherwise inevitable time-based decay. In the case of residential real estate, this criterion will be, arguably, almost always false: A lived-in (or uninhabited) property will suffer decay due to elements and normal wear and tear, over any time period. A property un-maintained will have changed. Conversely, a property that is maintained **also** will have changed: Was the roof replaced? The walls re-painted? The brickwork re-pointed? All of these represent changes to the nature of the object in question.

This reality compels us to make a minor assumption: That the property in question hasn't changed in an **economically meaningful** way since its last sale. Put another way, a property's relative value in the market now is not terribly dissimilar to its place in the market at the time of its last transaction.

The real trouble comes when a property has, in fact, been substantially changed. Some are gutted and modernized. Some have additions of living space. We have seen cases where a prior structure has been demolished or an existing property wholly rebuilt. In these cases, the prior price is not well connected to the present value.

### 2: The index applied is appropriate for the good in question

It is obvious that a price index for Jamaican sugar cane would not be relevant for finding the present value of UK steel. While the value of both goods may have changed over time, and both may have been subject to common global inflation factors, we need an index that is measuring price changes in goods that are most closely related to the good we want to value. A global steel index would be good. A European steel index would be better. A UK steel index would be best. An indexed estimate is only as accurate as the chosen index is appropriate for the valuation task.

The challenge in residential real estate is that micro geographies in close physical proximity

can represent very different markets. Yet calculating a robust index for a small geography is hard. It's sort of the Heisenberg uncertainty problem for housing economics: A smaller geography for an index has higher specificity for the properties within it, which ought to improve the index's accuracy. However, a smaller geography means fewer data points available which would diminish the accuracy of the index. There inevitably comes a trade-off between geographic specificity and data abundance.

The challenge is also relevance: Markets are not only geographically defined, but also by types of properties and market segments. Supply and demand for 2 bed flats and 5 bed houses could be at very different balance points in the same geography, and we'd expect these differing tensions to be reflected by price indices generated for these different segments in the same geography.

### Creating House Price Indices

There are many methods. The Land Registry publishes an index which takes the geometric mean transaction price observed over time. The Halifax has a long running index generated by the weighted average observed price of homes selling in its mortgage business, where the weightings standardize the basket of transactions by house type, size and geography.

These are good indices for describing price trends for macro economists and entities wishing to understand the UK market as a whole, and even regional sub-markets. But for the purposes of modelling predictions, they are simply too broad in their scope to make the subtle distinctions between micro geographies that translate to critical improvements in indexed estimate accuracy for specific properties.

Therefore, we at Whenferesh decided to create our own house price indices from our own raw ingredients.

### Why making a house price index isn't trivial

At first glance, one might think creating a price index is straight forward. Isn't it simply the average price trended over time?

While one could trend a simple average over time, one would be plagued by many difficulties:

- The mix of properties changes each month. No matter the unit of aggregation, the underlying mix of the kinds of properties in question will shift each month merely due to random fluctuations. Last month was 36% flats, this month it's 33%. Last month 12% of the sales came from SW11, this month it's 14%. In the immediate aftermath of the Grenfell Tower fire, the volumes of completions of flat sales crashed. Failure to compensate for this will result in index volatility due not to a change in prices but due to random change in the mix of the kinds of properties used to calculate the index.
- **Outliers happen.** The more specific an index, the smaller the sample of data for defining it. Small samples are intrinsically less reliable for estimating the mean, and outliers in small samples can strongly move the mean. One needs a way to dynamically reduce outlier influence to assure that changes in mean price are due to changes in nominal price levels and not random sample fluctuations.

• **Availability bias.** Higher turnover properties will provide data points to the index more often than lower turnover properties. The basket of properties that sold doesn't necessarily represent the population of properties in the market to be represented by the index.

A well documented method for constructing house price indices is called Repeat Sales Regression (RSR). It's the gold standard in academic circles for describing property markets. The idea is to pair sales of the same property at two points in time. With enough paired points, all time periods are covered. The paired sales data is transformed into a specially designed matrix for the analysis, which proposes a price index that best explains all the value changes over time observed among all the pairs in the sample. This approach is a clever solution to a tough problem, but it still has many issues when applied to our specific purpose:

- Requires a lot of data. A transaction data point for a property is of no use (and no consideration!) unless and until that same property has another transaction data point. As such, many transaction data are discarded from use. Limiting the pool of useful data inherently limits the market specificity which can be described accurately by this method. And as we've established, there will be a trade-off between specificity and data prevalence. By diminishing the data set, RSR will only be able to describe larger markets. This is fine for macro economists, and not well suited to making predictions for individual addresses.
- 2. More signal from higher turnover properties. Properties that turn over more often will provide more sets of paired sales to the sample, and hence have more influence on the index than properties with lower turnover, that may participate less, or not at all if they lack a paired set of sales.
- 3. Assumes properties have static attributes. It's difficult to avoid this in any index, but it's not true that a change in price for a property over time will be solely due to market conditions. One can assume that there is a constant proportion of underlying asset change present in the index across all time. But assuming away the problem doesn't remove it: There will be paired sales implying value change vastly different from the true market effect. Failure to identify these pairs of sales and remove them from consideration will distort our picture of the market which will in turn diminish the accuracy of the index in predicting present values from past sales.
- 4. It still requires judgement. One can work to address the above difficulties and others via prudent rules. But the rules still need to be stipulated and not all of them can be objectively defined. How long a time interval between sales should be required to trust that both sales are true market transactions for a similar good and not a case of buy-renovate-flip? How much of a deviation from the expected value change is enough to disqualify a pair of sales as being an outlier? How much data is needed to construct a valid index for a market? What to do if a valid index cannot be constructed for a market (Detached homes in EC)? There are many defensible approaches to these issues, and each approach will have an effect on the performance of the model when applied to making predictions.

#### How do we build an index?

To build an index for a micro market, we do the following:

- To correct for random fluctuations in the mix of properties that sell each month, each month's input data is weighted to standardize the mix by property type and geography such that the data inputs represent the distribution of properties in the area to be represented.
- To correct for outlier cases, highly improbable values are dynamically detected and suppressed in the index creation process.
- To correct for data scarcity, gaps in the data series are infilled via interpolation.
- To correct for random fluctuations, a smoothing algorithm tunes each series individually to bring volatility within acceptable parameters. This effectively pools data across time to generate the effect of more robust sampling.

We make one of the index layers, that at postal area plus property type, available for client use. This index is often the heaviest weighted index in the index model, meaning relying upon it heavily leads to the most accuracy. It represents generally an ideal geographic specificity: There are around 120 viable postal areas, but only 12 UK Regions. And while the postal area level gives great specificity alone, by adding an additional layer or property type within that geography, we now have nearly 500 distinct and specific house price indices to apply to past prices to estimate their present values. Postal areas are generally large enough that there are enough transactions to generate a viable index for each Area and Type combination (an exception example would be Detached houses in Central London). This profoundly greater specificity generates much greater estimate accuracy at the individual property level, and hence also much greater accuracy at the portfolio level.

The indices' property type specificity is important, as shown in the following examples:

In the Bradford Postal area, we can see that Flats have had a very different price history than the houses have had for the duration of the index.



Alternatively, in Bristol, Terraces have outperformed the rest of the local market. This may well be because a Terraced House in Bristol is very different to a Terraced House in Huddersfield.



- In both of these cases, using just an area index would have missed important distinctions.
- Alternatively, within the same region, and for the same type of property, there are often very important differences in local markets. In the example of Terraced houses in the East Midlands Region, we can see that, since the housing crash of 2008-9, there have been starkly different price movements between Postal areas within the same region. Doncaster (DN) prices have barely climbed above their post-crisis peak, while Leicester (LE) prices have shot more than 50% above their pre-crisis peak in the same timeframe. Applying an "East Midlands Terraces" index to properties in Leicester (light brown) would under-represent

their appreciation, and would over value properties in Doncaster (green).



#### Region: East Midlands Derby, PROPERTYTYPE: TERRACED

#### **Rent Price Indices**

Every month we compile our property-level data about market rents. We remove extreme outlier data and then generate monthly averages, which are appropriately smoothed to generate valid indices. Our rent price indices capture very well changes in market conditions. A recent example of this can be seen in the differences in rent indices for distinct London postal areas during 2020 and into 2021. The indices show the steep decline in rental prices for the central London postal areas, but less decline for the less central areas, and flatter prices for the outer London areas. This fits with our understanding of the market dynamics at the time: During the pandemic, if people had to be stuck at home, they preferred to be any place other than in the middle of a city which had all of its most appealing features temporarily closed. This trend was further compounded by the broad adoption of working from home.

The important point is that these microgeographic distinctions in price movement would be lost in a composite "London" index.

From March of 2020 to March of 2021, our index picks up approximately a 10% drop in the WC postal area and a 20% drop in the EC postal area.

- If one were to apply the index to bring a £1000 market rent for a property in EC2 from April 2021 up to present market value, the overall London index would suggest a present value of £1219, while the EC index suggests a present value of £1364, a nearly 12% difference in value over just two critical years.
- Again, we can see the improved accuracy in indexed estimates that arises from applying the most specific and relevant price indices possible. Conversely, we can see the accuracy risks of applying an overly broad index to properties whose local markets are importantly different.

