

BCI AVM

WHITE PAPER



BLOCKCHAIN & CLIMATE
institute 

The logo graphic for the Blockchain & Climate Institute, featuring a stylized 'B' and 'C' made of teal squares with a grid pattern.

INTRODUCTION

BCI AVM is an automated valuation model built for the UK real estate market, and specifically for the purposes of the Blockchain Climate Institute. Using advanced machine learning technology, it focuses on providing industry best-practice valuations which assist BCI asset risk models, CREM models, and others who in turn provide a basis for residential property asset risks due to climate change, anywhere in UK. This paper explains the core concepts and functionality of BCI AVM.

COVERAGE

BCI AVM UK valuations are limited to single family homes, condominiums, apartments, and townhomes. Mobile homes, land, and commercial properties are not valued and are excluded from the database used during AVM development at this time, so that they do not bias value estimates of the included property types. BCI AVM provides valuations for 16M+ properties across all major UK cities and regions.

Individual valuations are based on local deviations between properties within the same postcode or the nearest comparable real estate market. In all cases, data used for development of the valuation model has been captured within 24 months of the AVM valuation date.

METHODOLOGY

AVM values are derived using technologically advanced machine learning methods, including:

- Supervised algorithms
- Tree-based & deep learning algorithms
- Feature engineering derived from small clusters of similar properties
- Ensemble (value blending) approaches

Both point values as well as value ranges are provided for each property that is valued. Value ranges are developed through a statistical bootstrapping approach that allows for asymmetrical ranges around each point value to be provided. The bootstrapping approach is flexible enough to accommodate reasonable BCI research quality standards for price recommendations. Thus, the upper and lower range price estimates indicate a range within which the most likely AVM error will be found. We derive this range by calculating the known errors found when testing against up to 15-100 randomly selected comparable properties.

CONFIDENCE SCORE AND FORECAST STANDARD DEVIATION

Every property receives a unique confidence score. This confidence score represents the precision of the AVM estimate and measures the deviation between the range of values and the point value itself. To provide the most transparent measure of variability in the estimates, we provide the larger of differences between the low and high value and point estimate. The low and high values represent the 16th and 84th percentile errors found during our bootstrap random re-sampling approach which are then applied to the point estimate, providing a 68% confidence interval around the point value, given the set of randomly generated comparable properties valued nearby. In this way, asymmetrical distributions of potential values are conservatively represented by the confidence score.

Confidence scores are calculated as 100 minus the forecast standard deviation, where the forecast standard deviation is the greater of the difference between the low range and the high range value and the point estimate, divided by the point estimate. This is an industry-standard way of computing confidence scores.

SCORING & ERROR METRICS

To gauge the error of BCI AVM we have compared our AVM values to over 100,000 properties listed on BCI over a recent period. These properties were selected to be representative of postcode, and local neighborhood characteristics such as property types, and sizes, and their proportional distribution over all of UK properties listed in EPC/Price data. For each new model version, these approx. 100,000+ properties are set aside at the outset of model training and development, in order to ensure that none of our production models ever sees these specific properties prior to AVM scoring.

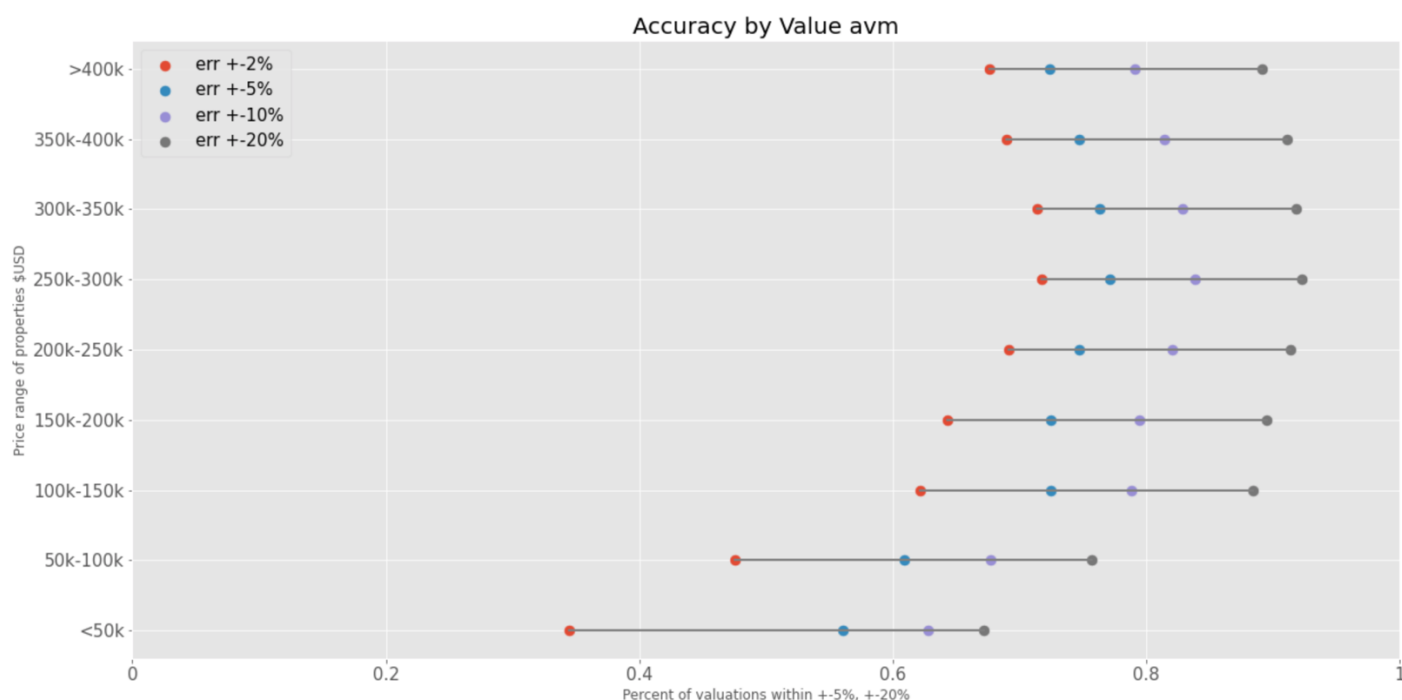
For the most recent scoring session, the median difference between our AVM estimates and the true price was approx. \$1639 USD, meaning that one half of all valuations are within \$1639 USD of the true price of the property. Additionally, 70%+ of valuations are within 5% of the true price, and 80%+ within 10%.

MEDIAN ABSOLUTE ERROR	\$1639 USD
VALUES WITHIN 5% OF TRUE PRICE	70%+
VALUES WITHIN 10% OF TRUE PRICE	80%+

DRAFT

```
OrderedDict([('MAE', 22522.85361974734),
             ('Root Mean Squared Error', 146534.3687459181),
             ('R2', 0.8444558060554183),
             ('MSE', 21472321223.7647),
             ('MedianAE', 1639.53125),
             ('MaxError', 20698366.75),
             ('ExpVariance', 0.8444577730092291)])
```

The chart below highlights the overall distribution of the absolute differences between our AVM values and observed true prices across UK postcodes.



To provide transparency to the valuation process, the following outlines how a typical property is valued with BCI AVM.

- 1. Employ hyper-tuned UK valuation model:** One stacked pipeline, hereafter referred to as “predictive model” or “avm”, which has been hyper-tuned to achieve low error valuations for the entire UK area and property is deployed to create a point estimate for the target property.
- 2. Find ‘comparable’ properties.** Next, we select a set of up to 100 comparable properties with known true prices, based on the geographically nearest and most recent (within one year) listings of homes of the same or most similar type. When we are not able to identify a minimum of 15 comparable properties, we expand the criteria used to qualify what is a “comparable” property. In very rare cases, such as in rural settings, we will expand the geographical region until a property of similar type is found. Through use of a robust statistical procedure and randomly re-sampling the comparable properties, many of the initial comparable sales are down weighted or eliminated due to dissimilarity with the subject property or other issues. Not all properties are valued with all models as the use of some models is contingent upon available data, such as images, or listing descriptions. At the end of this step, we will have made predictions on all randomly sampled comparable properties with known true prices, and thereby calculated the error for each prediction.

3. Determine range estimate(s): We then rank the errors of estimated values from low to high. We use the 16th and 84th percentile values as the low-end and high-end range of the value estimate, and we apply these error percentages to our original point value estimate.

4. Determine the Confidence Score: Next, we look at the percentage difference between the low (16th percentile) and high (84th) values and the original point value estimate. We take the larger of the two (the FSD Max) and subtract from 100 to create a confidence score. By taking the larger of the two, our confidence score accounts for skewed distributions of values and offers a conservative measure of our confidence in the AVM value.

5. Choosing a model: Prior to valuing properties with each new AVM version release, we conduct regular AVM scoring tests where we compare our AVM values to known true prices of properties that were recently listed on UK EPC/Price databases. Using this information, we then hyper-tune model training parameters using search algorithms which seek to find those parameters which provide the lowest error. The most accurate data processing pipeline + model(s) get rolled-up into a unified pipeline which we refer to as our model, used for each estimated AVM value, along with its respective low and high values, FSD and confidence score.

PROPERTIES WE DON'T VALUE

There are several property types we intentionally don't value, and don't include in our comparable selection. We eliminate these property types for two reasons. First, these property types have a negative impact to the predictive value of other homes. Second, these property types require different data and modeling and don't belong in an AVM for standard single-family residences.

1. **Mobile homes:** In the case of mobile homes, a two-bedroom two-bathroom home should value lower than a standard fixed foundation of the same configuration.
2. **Multi-million-dollar homes:** These properties often have unique features or locations that influence the value, making comparable selection impossible.
3. **Homes on farm or agricultural land:** In this case, the land and its use influence the value of the home because they are sold as a single unit of real property. Because of this, the valuation requires data and modeling from similar comparable properties.
4. **Multi-unit homes:** These properties rely on rental data to determine value because they are most typically investment properties that are under multi-tenant lease. As such, they require different data and modeling to derive an accurate value.

DRAFT