

American Journal of Economics and Business Management

Vol. 9 Issue 11 | **pp.** 466-478 | **ISSN:** 2576-5973
Available online @ https://www.globalresearchnetwork.us/index.php/ajebm

Article

Explainable Machine Learning Valuation in Thin Housing Markets: Evidence from Newcastle upon Tyne and Sunderland, England

Dr Victor Joes Obasse

1. Lakeshore Management College
* Correspondence: victorobasse@gmail.com, https://orcid.org/0009-0002-3477-0946

Abstract: This study tests how accurately open British transaction data can value homes in two thin, lower priced housing markets, and whether machine learning improves on standard methods without bias against cheaper homes. Using 51,899 standard residential sales in Newcastle upon Tyne and Sunderland from HM Land Registry Price Paid Data for 2018 to 2024, it compares a comparable sales baseline, a fixed effects hedonic regression, a random forest, gradient boosting and an ensemble. Every feature is built only from earlier sales, and the models are tested on all 6,857 sales of 2024 with a quarterly rolling origin. The ensemble achieved a median absolute percentage error of 12.6 per cent, against 20.6 per cent for the hedonic regression, and valued 41.4 per cent of homes within 10 per cent of the price. Accuracy rested mainly on the price history of the postcode unit, the street and the dwelling's own previous sale, and held up where local comparables were scarce. Grouped by sale price, all models overvalued the cheapest homes, but grouped by predicted value the ensemble's ratios stayed between 0.98 and 1.01. Bank Rate added no accuracy. Without floor area or condition data, the models suit screening rather than lending decisions on their own.

Keywords: automated valuation models; machine learning; hedonic pricing; vertical equity; thin housing markets; HM Land Registry; Newcastle upon Tyne; Sunderland; ValuInsight AI

Citation: Obasse, V. J. Explainable Machine Learning Valuation in Thin Housing Markets: Evidence from Newcastle upon Tyne and Sunderland, England. American Journal of Economics and Business Management 2026, 9(11), 466-478

Received: 13th Aug 2026
Revised: 28th Sep 2026
Accepted: 15th Oct 2026
Published: 09th Nov 2026

Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

Residential property is the largest asset most British households own, and a valuation stands behind almost every purchase, remortgage and secured loan. Lenders use automated valuation models (AVMs) for many lower risk decisions, such as remortgages at modest loan to value, yet the evidence on how well these models perform is mostly private. Vendors publish headline accuracy figures, but rarely the data, the test design or the errors broken down by the kinds of property where models struggle [1]. That gap matters most outside London and the South East. Cities such as Newcastle upon Tyne and Sunderland have lower price levels, a high share of terraced housing and many postcode sectors with only a handful of sales in a given year. In such thin markets the sales comparison approach has few comparables to work with. A fixed error in pounds also becomes a large error in percentage terms when the property is cheap [2]. The result can be valuations that are least reliable for the households with the least housing wealth, a pattern assessors call vertical inequity.

This study asks how far an open, reproducible valuation pipeline built only on public British data can go in these two markets [3]. It addresses five questions:

1. How accurate are valuation models for Newcastle upon Tyne and Sunderland when tested strictly out of time, on sales the models could not have seen?
2. Do tree based machine learning methods outperform a fixed effects hedonic regression and a simple sales comparison benchmark?
3. Where do errors concentrate, by district, property type, price band and the depth of the local market?
4. Does adding the Bank of England Bank Rate improve short horizon accuracy?
5. Which inputs drive the predictions, and are they economically sensible?

The paper makes three contributions [4]. First, it reports a complete validation on 51,899 standard residential sales recorded by HM Land Registry between 2018 and 2024, with every feature built only from sales that preceded the one being valued. Second, it reports accuracy alongside the ratio study measures used by assessors, so that vertical equity is measured rather than asserted. Third, it makes the full analysis code available, so that each figure in the paper can be reproduced from the public data [5]. The method evaluated here underpins the ValuInsight AI prototype developed by the author; the paper assesses that method on evidence and does not rely on any claim made for the product.

Hedonic pricing treats a dwelling as a bundle of characteristics, each with an implicit price revealed by market transactions. Reviews of several hundred applications show that size, age, type and location explain most of the variation in residential prices, while the functional form and the treatment of location drive much of the difference between studies. Hill shows how the same framework supports price indices that control for changes in the mix of homes sold.

British studies have used the approach to price specific amenities [6]. Cheshire and Sheppard valued land and neighbourhood attributes in Reading and Darlington. Gibbons and Machin priced access to good primary schools and to new rail stations in England. Fuerst et al. linked Energy Performance Certificate ratings to Land Registry transactions and found a price premium for more efficient homes. These papers share a lesson relevant here: location effects are large and local, often varying between neighbouring postcode sectors [7].

1.1 Repeat sales and comparable sales

A second tradition uses dwellings that sell more than once, so that fixed characteristics cancel out. Repeat sales methods are powerful when a property has a recent sale, but most dwellings in any short window have none. Practitioners rely instead on comparable sales, which work well in deep, uniform markets and poorly in thin ones. Spatial methods that borrow strength from nearby sales can reduce this problem [8].

1.2 Machine learning and automated valuation

Tree ensembles such as random forests and gradient boosting capture non linear effects and interactions without the analyst specifying them. Antipov and Pokryshevskaya found random forests outperformed several alternatives for mass appraisal of apartments. Mayer et al. compared six methods on about 123,000 Swiss sales, judging both accuracy and the stability of valuations over time; gradient boosting with expanding training windows was among their preferred methods, alongside robust and mixed effects regression [9]. Kok, Koponen and Martínez Barbosa set out how such models move valuation from manual appraisal towards automated systems, and Glumac and Des Rosiers summarise the advantages and limits of AVMs for practice. Two cautions run through this work. First, results depend heavily on the test design. Random splits let a model learn from sales that happened after the one it values, which inflates accuracy relative to real use. Second, a single headline error figure says little. Steurer, Hill and Pfeifer argue for a battery of metrics, including the share of valuations within 10 and 20 per cent of the price, and for reporting error by segment [10].

1.3 Explainability and fairness

Opaque models raise concerns about accountability in high stakes decisions, and some argue that interpretable models should be preferred where they perform comparably. Post hoc tools such as SHAP values and permutation importance offer partial insight into what a model relies on. In valuation, fairness has a precise measurable form.

Ratio studies compare predicted values with sale prices and test whether the ratio drifts with price; the coefficient of dispersion and the price related differential are standard measures. McMillen and Singh show that assessments are often regressive, overvaluing cheap homes and undervaluing expensive ones [11].

1.4 Macroeconomic drivers

Interest rates, credit conditions and incomes drive house prices over the cycle, and taxes on transactions alter the timing and level of sales. The 2020 to 2021 stamp duty holiday and the rise in Bank Rate from 0.10 per cent in December 2021 to 5.25 per cent in August 2023 make the study period a demanding test. Whether macroeconomic variables help a valuation model over a horizon of weeks, rather than years, is an empirical question this paper tests directly [12].

1.5 The gap

Published UK AVM evidence rarely combines open data, a strictly time ordered test, segment level error and ratio study measures for a secondary city market. This paper supplies that combination for Newcastle upon Tyne and Sunderland.

1.6 Data Source and study area

The study uses HM Land Registry Price Paid Data, which records the price, date, full postcode, property type, new build status and tenure of every residential sale for value lodged for registration in England and Wales. The yearly files for 2018 to 2024 were downloaded and filtered to the two local authority districts of Newcastle upon Tyne and Sunderland, both in Tyne and Wear. The data are published under the Open Government Licence v3.0.

The two districts make a useful pair. They are neighbours and share a regional economy, yet they differ in price level, housing stock and market depth. Newcastle has a large stock of city centre and student flats; Sunderland is dominated by semi detached and terraced houses [13].

1.7 Cleaning

The raw extract held 65,701 records. The following were removed: 13,686 category B transactions (additional price paid entries such as repossessions and transfers under a power of sale, buy to let purchases identified by a mortgage, and transfers to non private buyers; 3 records without a postcode; and 113 records priced below £20,000 or above the 99.9th percentile of £1.44 million. No records had property type O. The final dataset holds 51,899 standard sales. Sales from 2018 serve only as history for the look back features described in Section 4, so the modelling sample runs from January 2019 to December 2024 and holds 44,841 sales [14].

1.8 Descriptive statistics

Table 1 summarises prices by district and type over 2018 to 2024. Newcastle prices are higher across every type, and the gap is widest for flats, where the Newcastle median is £130,000 against £72,000 in Sunderland.

Table 1. Standard residential sales, 2018 to 2024

District	Type	Sales	Median price (£)	Mean price (£)
Newcastle upon Tyne	Detached	4,964	312,972	343,642
Newcastle upon Tyne	Semi detached	8,093	183,300	212,125
Newcastle upon Tyne	Terraced	8,191	166,400	204,787
Newcastle upon Tyne	Flat	6,039	130,000	146,170
Newcastle upon Tyne	All	27,287	182,950	219,251

District	Type	Sales	Median price (£)	Mean price (£)
Sunderland	Detached	5,148	255,000	276,344
Sunderland	Semi detached	9,145	145,000	152,975
Sunderland	Terraced	8,327	106,000	116,602
Sunderland	Flat	1,992	72,000	83,020
Sunderland	All	24,612	140,000	160,812

Source: author's calculations from HM Land Registry Price Paid Data.

New builds make up 18.5 per cent of Newcastle sales and 14.0 per cent of Sunderland sales; leasehold makes up 29.9 and 17.2 per cent respectively. The data cover 53 postcode sectors in Newcastle and 42 in Sunderland. Market depth differs sharply: the median Newcastle sector records 46 standard sales a year against 85 in Sunderland, and 39.6 per cent of Newcastle sectors average fewer than 25 sales a year, compared with 11.9 per cent in Sunderland [15]. Newcastle is therefore the thinner market at sector level, despite its larger total volume.

Annual median prices rose from £161,400 in 2018 to £199,950 in 2024 in Newcastle, and from £128,000 to £150,000 in Sunderland. Most of that growth came between 2020 and 2022, during and after the stamp duty holiday; medians were flat from 2023 as Bank Rate peaked. Transaction volume fell to 930 sales in the second quarter of 2020, during the first national lockdown, and peaked at 2,516 in the third quarter of 2021.

1.9 What the data cannot see

Price Paid Data hold no floor area, room count, condition, plot size or energy rating. The models in this paper must therefore infer quality from location, type and sale history. This is a deliberate test of what the open transaction record alone can support, and it sets a floor for accuracy rather than a ceiling [16].

2. Materials and Methods

2.1 Features built only from the past

Every feature for a sale on day t is computed only from sales registered before day t. This rule prevents leakage, in which a model learns from the very price, or from later prices, it is meant to predict. The features fall into five groups.

a. **Dwelling attributes.** Property type, new build status and tenure, as recorded by HM Land Registry.
b. **Market level and momentum.** The mean log price of the same property type in the same district over the previous 6 months, which serves as the anchor for each valuation, and the gap between the 6 month and 24 month means, which measures momentum.
c. **Local comparables.** The mean log price and the number of sales over the previous 24 months in the same postcode sector, and in the same sector and property type. The same measures are computed over the previous 10 years for the same street and the same full postcode unit. Each is expressed relative to the district mean, so that it measures a location premium rather than a price level.
d. **Repeat sales.** For dwellings sold before within the data, the previous price is carried forward by the change in the district mean since that sale, together with the years elapsed.
e. **Macroeconomic.** The Bank of England Bank Rate in force on the sale date (Bank of England, 2025), used only in the ablation test of Section 5.4.

2.2 Models

Five approaches are compared.

a. **Comparable sales baseline.** The 24 month mean log price of the same type in the same postcode sector, adjusted by district momentum. Where fewer than five such sales exist, the sector mean for all types is used, shifted by the district gap between that type and all types; failing that, the district mean for the type is used. This mimics a simple desk valuation from recent local sales.
b. **Hedonic regression.** Log price regressed on property type, new build status, tenure, the interaction of type and new build, postcode sector fixed effects and calendar month fixed effects. Sectors with fewer than 10 training sales are merged into their postcode district. Sales in a month beyond the training data take the last estimated month effect. A negligible ridge penalty (alpha of 1) is applied for numerical stability.
c. **Random forest** (Breiman, 2001) with 300 trees.
d. **Gradient boosting**, using the histogram based implementation in scikit-learn (Pedregosa et al., 2011), which follows the design of LightGBM (Ke et al., 2017), with an absolute error loss and early stopping.
e. **Ensemble.** A weighted average of the log predictions of models 2 to 4, with weights chosen on the validation year.

All prices are modelled on a log scale. This is standard in housing studies for two reasons. Price effects then read as percentages, so being detached adds a share of value rather than a fixed sum in pounds. Logs also stop a few very expensive homes from dominating the fit.

2.3 How the tree models value a home (Equation 1)

ln(Predicted price) = Market anchor + Adjustment for this home
Written with symbols: $\ln \hat{P}_{i,t} = A_{i,t} + f(x_{i,t})$

In words, the valuation has two parts. The first part, the market anchor A, is the average log price of the same property type in the same district over the six months before the sale. It sets the general price level and moves with the market. The second part, f(x), is the premium or discount for this particular home that the random forest or gradient boosting model has learned from past sales. It depends on the features x listed in Section 4.1, such as the price history of the street and postcode, and the home's own previous sale. The anchor answers the question of what a typical home of this type sells for here at the moment. The adjustment answers the question of how this home differs from that typical home.

As an illustration only: if semi detached houses in Sunderland have recently sold for about £150,000 on average, and the model judges a particular semi to be worth about 8 per cent more because of its street and its last sale, the valuation is about £150,000 × 1.08, or £162,000.

The anchor is used because tree models cannot predict values above or below the range of prices they have already seen. Leaving the price level to the anchor means the trees only need to learn relative differences, which stay stable from one quarter to the next. In this study, where prices were broadly flat in 2024 and the models were retrained every quarter, the anchor made little difference to accuracy; it matters most when prices move quickly.

2.4 The hedonic regression (Equation 2)

ln(Price) = Constant + Type effect + New build effect + Leasehold effect + Type and new build interaction + Postcode sector effect + Month effect + Error
Written with symbols: $\ln P_{i,t} = \alpha + \beta_k \text{Type}_{ik} + \gamma \text{New}_i + \delta \text{Lease}_i + \theta_k (\text{Type}_{ik} \times \text{New}_i) + \mu_s + \tau_t + \varepsilon_{i,t}$ Each term has a plain meaning:

a. **Constant (α):** the baseline log price of an existing, freehold terraced house, the reference home against which every other effect is measured.

b. **Type effect (β):** the extra or lower value of a detached house, a semi detached house or a flat compared with a terraced house.
c. **New build effect (γ):** the premium paid for a newly built home.
d. **Leasehold effect (δ):** the difference in value between leasehold and freehold.
e. **Interaction (θ):** allows the new build premium to differ by property type, since a new detached house and a new flat need not carry the same premium.
f. **Postcode sector effect (μ):** the location premium or discount of each postcode sector, such as NE3 1 or SR2 8, holding the other attributes fixed.
g. **Month effect (τ):** the general level of prices in each calendar month, which captures market movements over time.
h. **Error (ε):** everything the model cannot see, such as floor area, condition and the circumstances of the sale.

Because the model is in logs, each coefficient converts to a percentage price effect with the formula (e raised to the coefficient) minus 1. For example, the semi detached coefficient of 0.215 in Table 4 means $e^{0.215} - 1 = 0.240$, so a semi detached house sells for about 24 per cent more than an otherwise similar terraced house in the same sector and month.

2.5 Tuning and testing

The test design follows how an AVM is used in practice. Hyperparameters and ensemble weights were chosen by training on the 31,076 sales of 2019 to 2022 and validating on the 6,908 sales of 2023. The grids were small and are reported in full in the replication files: for the random forest, minimum leaf sizes of 3 and 10 and feature fractions of 0.33 and 0.6; for gradient boosting, learning rates of 0.03 and 0.08, 31 or 63 leaves and minimum leaf sizes of 20 and 60. The selected settings were a leaf size of 3 with a feature fraction of 0.33 for the forest, and a learning rate of 0.03 with 31 leaves and a minimum leaf size of 60 for boosting. The validation set gave the ensemble weights of 0 for the hedonic model, 0.35 for the random forest and 0.65 for gradient boosting. The final test is a rolling origin over the four quarters of 2024. Before each quarter, every model is refitted on all sales from January 2019 to the day before the quarter begins, then asked to value that quarter's 1,457 to 1,941 sales. In total, 6,857 sales from 2024 are valued by models that never saw them, nor any later sale.

2.6 Measures of accuracy and equity

Accuracy is reported with the measures recommended by Steurer, Hill and Pfeifer:

a. Mean absolute error (MAE) and root mean squared error (RMSE), both in pounds;
b. Mean and median absolute percentage error (MAPE and mdape);
c. R squared on the price scale;
d. PE10 and PE20, the shares of valuations within 10 and 20 per cent of the sale price.

Equity is assessed with the ratio study measures of IAAO. The ratio is the predicted value divided by the sale price. The coefficient of dispersion (COD) is the mean absolute deviation of the ratio from its median, as a percentage of the median. The price related differential (PRD) is the mean ratio divided by the ratio of summed predictions to summed prices; values above 1.03 indicate regressivity. IAAO guidance suggests a COD of 15 or below for residential property in older or more varied areas.

Differences in MAE between models are tested with a paired bootstrap of 2,000 resamples of the 2024 sales. Explainability is examined in two ways. The first uses the coefficients of a full sample hedonic regression with heteroskedasticity robust (HC1) standard errors. The second uses permutation importance, measured as the rise in absolute error when each feature is shuffled, for the gradient boosting model of the fourth quarter. SHAP values were not computed because the SHAP library could not be installed in the computing environment; permutation importance answers the same global question with less detail on individual valuations.

3. Results

3.1 Overall accuracy on 2024 sales

The tree based models cut the typical valuation error by about a third relative to both benchmarks. Table 2 reports the rolling origin test on all 6,857 standard sales of 2024. The ensemble's median absolute percentage error is 12.6 per cent, against 19.8 per cent for the comparable sales baseline and 20.6 per cent for the hedonic regression. Its mean absolute error is £34,006, compared with £52,707 and £54,645 [17].

Table 2. Out of time accuracy, Newcastle upon Tyne and Sunderland, all 2024 sales (n = 6,857)

Model	MAE (£)	RMSE (£)	MAPE (%)	MdAPE (%)	R^2	PE10 (%)	PE20 (%)	Median ratio	COD	PRD
Comparable sales baseline	52,707	86,592	28.3	19.8	0.572	27.1	50.4	0.987	28.6	1.130
Hedonic regression	54,645	89,062	29.1	20.6	0.547	25.7	49.0	0.978	29.7	1.138
Random forest	34,883	59,212	18.3	13.0	0.800	40.3	68.6	0.996	18.4	1.060
Gradient boosting	34,144	59,444	17.9	12.4	0.798	41.5	70.2	0.996	18.0	1.061
Ensemble	34,006	58,842	17.9	12.6	0.802	41.4	70.3	0.994	18.0	1.061

Notes: rolling origin by quarter; each model refitted on all sales from January 2019 to the day before the quarter. The ensemble weights are 0.35 for the random forest and 0.65 for gradient boosting; the validation year gave the hedonic model a weight of zero [18]. The gain over the benchmarks is large and precise. In the paired bootstrap, the ensemble's MAE is £20,639 lower than the hedonic model's (95 per cent interval £19,431 to £21,853) and £18,701 lower than the comparable baseline's (£17,627 to £19,808). Gains among the tree models are small. Gradient boosting beats the random forest by £740 (£361 to £1,117), and the ensemble's margin over gradient boosting alone, £138, has an interval that touches zero (−£2 to £272). The ensemble therefore adds stability rather than accuracy [19]. The results are stable. On the 2023 validation year the ensemble scored an MdAPE of 13.6 per cent and an MAE of £35,957, close to the 2024 figures. Across the four test quarters its MdAPE ranged from 12.4 to 13.1 per cent. All models are close to unbiased at the median, with median ratios between 0.978 and 0.996 [20].

These figures are useful but modest. The models see no floor area or condition, and it shows. Fewer than half of the ensemble's valuations (41.4 per cent) fall within 10 per cent of the price, and its COD of 18.0 exceeds the IAAO ceiling of 15. The models are useful for screening and triage, not for lending decisions on their own.

3.2 Where errors concentrate

Table 3 breaks the error down by segment for the two benchmarks and the ensemble.

Table 3. Accuracy by segment, 2024 sales

Segment	Sales	Comparable MdAPE (%)	Hedonic MdAPE (%)	Ensemble MdAPE (%)	Ensemble PE10 (%)
Newcastle upon Tyne	3,614	19.1	20.0	11.9	43.2
Sunderland	3,243	20.7	21.2	13.3	39.4
Detached	1,313	14.8	17.0	11.3	45.8
Semi detached	2,234	18.7	18.0	11.8	44.0

Segment	Sales	Comparable MdAPE (%)	Hedonic MdAPE (%)	Ensemble MdAPE (%)	Ensemble PE10 (%)
Terraced	2,244	23.8	24.6	13.4	38.7
Flat	1,066	21.8	25.2	14.1	36.0
New build	948	14.4	14.1	9.8	51.7
Existing dwelling	5,909	20.9	22.2	13.1	39.8
Previous sale on record	1,521	19.0	20.1	9.2	53.3
No previous sale	5,336	20.0	20.8	13.6	38.0
Thin market: under 10 prior sales of type in sector	165	19.4	18.4	12.8	42.4
Medium: 10 to 29 prior sales	662	20.6	23.0	13.5	38.7
Deep: 30 or more prior sales	6,030	19.7	20.3	12.5	41.7

Four patterns stand out. First, terraced houses and flats are the hardest to value. These are the most varied stock, where size and condition, which the data do not record, differ most within a street. Second, new builds are the easiest, because developers sell similar units at similar prices. Third, a previous sale on record lowers the ensemble's MdAPE from 13.6 to 9.2 per cent, the largest single gain in the table. Fourth, thin markets do not collapse [21]. The ensemble holds an MdAPE of 12.8 per cent where fewer than 10 comparable sales exist, because street and postcode unit history fill the gap. The thin group is small (165 sales), so this last result needs confirming on more data. Sunderland is valued slightly less accurately than Newcastle despite its deeper sector markets. A likely reason is its lower price level: an error of £15,000 is 15 per cent of a £100,000 terrace but 6 per cent of a £250,000 semi [22].

3.3 Vertical equity

Whether valuations are regressive depends on how the question is asked. Grouped by sale price, every model looks strongly regressive. In the lowest fifth of 2024 sale prices (under £110,000), the ensemble's median ratio is 1.128, an overvaluation of 12.8 per cent. In the highest fifth (over £281,910) it is 0.924, an undervaluation of 7.6 per cent. The PRD of 1.061 lies outside the IAAO range of 0.98 to 1.03.

Part of this pattern is mechanical. Sorting by sale price places in the bottom group the sales that happened to sell below their expected value, so the ratio rises there even for an unbiased model. Grouped instead by predicted value, the ensemble's median ratio stays between 0.979 and 1.007 across all five groups (Figure 1) [23]. Grouped that way, the comparable baseline and hedonic model still drift, with ratios between 0.963 and 1.014. The ensemble's PRD of 1.061 is, however, much lower than the 1.130 and 1.138 of the benchmarks, so it is the least regressive of the five on either measure. For lending, the overvaluation of the cheapest homes remains a risk that any user of these models should test for in their own portfolio [24].

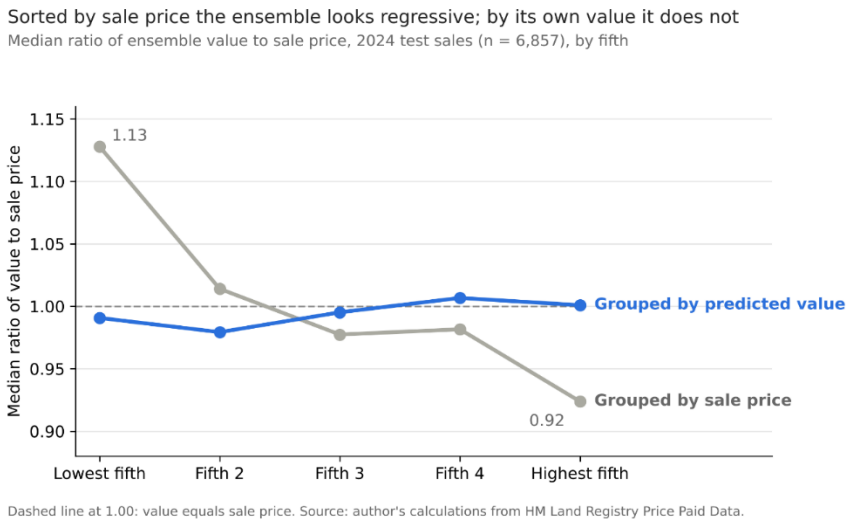


Figure 1. Median ratio of the ensemble value to the sale price, by fifth of the 2024 sales, grouped two ways.

3.4 Does Bank Rate help?

Adding Bank Rate to the gradient boosting model left accuracy unchanged: the MAE moved by £40, with a 95 per cent interval of −£135 to £60, and the MdAPE moved from 12.4 to 12.5 per cent [25]. This is not evidence that interest rates do not matter. Their effect on prices is already captured by the 6 month market anchor, which moves as the market reacts to rate changes. Over a horizon of one quarter, the recent level of local prices carries the information that the policy rate would add.

3.5 What drives the valuations

The full sample hedonic regression (44,841 sales from 2019 to 2024, with district by quarter and 90 location effects; R^2 of 0.636 on log price) gives economically sensible implicit prices. Table 4 reports them against an existing freehold terraced house [26].

Table 4. Hedonic implicit prices, 2019 to 2024

Attribute	Coefficient (log)	Robust s.e.	Price effect (%)
Detached	0.757	0.007	+113.3
Semi detached	0.215	0.005	+24.0
Flat	−0.485	0.009	−38.4
New build (terraced)	0.272	0.010	+31.2
New build × detached	−0.213	0.012	−19.2
New build × semi detached	−0.061	0.011	−5.9
New build × flat	−0.005	0.024	−0.5
Leasehold	0.004	0.007	+0.4

Notes: the dependent variable is log price; HC1 standard errors; the price effect is the exponential of the coefficient minus one.

The new build premium is about 31 per cent for terraced houses and flats, 23 per cent for semi detached houses and 6 per cent for detached houses [27]. Once location and type are held fixed, leasehold carries no measurable discount, perhaps because most leasehold sales here are flats, whose discount the type effect already captures. Location effects are wide: the 90th percentile sector effect is about three times the 10th. Permutation importance for the gradient boosting model (Figure 2) tells the same story from the machine learning side. The model relies most on the price history of the same postcode unit, then on property type, the dwelling's own previous sale, and street and sector premia [28]. Calendar time and district carry no importance once the market anchor is in place. This is what a valuer would expect: what nearby, similar homes sold for, and what this home sold for last time [29].

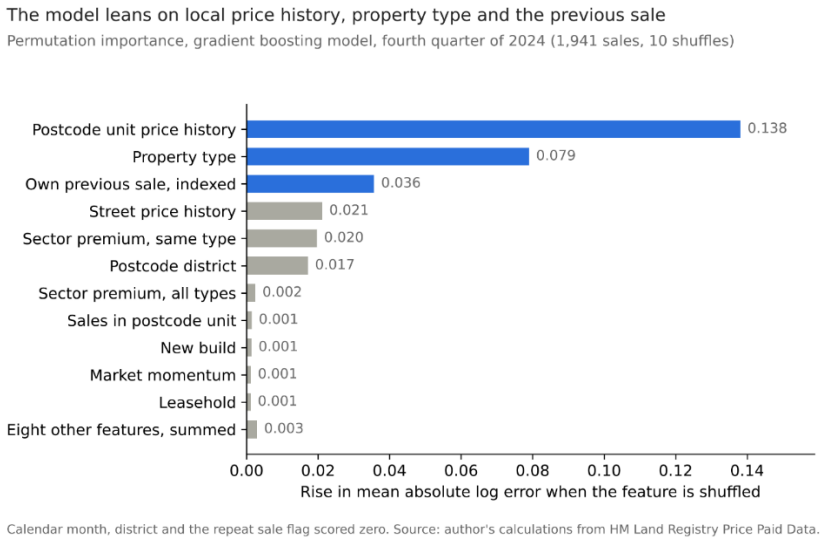


Figure 2. Permutation importance of the gradient boosting model for the fourth quarter of 2024. The eight lowest features are combined into one bar.

4. Discussion

4.1 What the evidence supports

Three findings are robust. First, tree ensembles built on open transaction data value homes in these two districts far better than a hedonic regression or a simple comparable approach on the same inputs, with a reduction in typical error of about a third. Second, the gain comes mainly from fine grained price history: the postcode unit, the street and the dwelling's own previous sale. This matches the case for local, repeat sales information in the literature [30]. Third, the ensemble is less regressive than the benchmarks, and much of its apparent regressivity is the artefact of grouping by sale price. The paper does not support several stronger claims. It does not show that the models match the accuracy a lender needs for a decision without inspection. A COD of 18.0 and a PE10 of 41.4 per cent are well short of that standard. Nor does it show that macroeconomic data improve valuation over a horizon of weeks. Finally, it does not show that the method generalises beyond Tyne and Wear [31].

4.2 Limitations

a. **Missing physical attributes.** Price Paid Data contain no floor area, rooms, condition or energy rating. Much of the remaining error is the variation in size and condition within a street. Linking Energy Performance Certificate records, which hold floor area and energy rating, is the clearest route to better accuracy.

b. **Short sale history.** Repeat sales are identified only within 2018 to 2024, so 77.8 per cent of 2024 sales have no previous price in the data. Adding the Price Paid files back to 1995 would raise this coverage.

c. **Two districts, one region.** The results describe Newcastle upon Tyne and Sunderland. Markets with different stock, such as London flats or rural detached houses, may behave differently.
d. **Standard sales only.** Category B transactions, including repossessions and many buy to let purchases, were excluded. The models are not tested on these.
e. **Matching on address text.** Previous sales are matched on postcode, house number or name and flat identifier. Changes in address format can break a match, which would understate repeat sale coverage.
f. **Small thin market sample.** Only 165 test sales fell in the thinnest group, so the finding that accuracy holds in thin markets is provisional.
g. **Limited tuning.** The hyperparameter grids were small. A wider search could improve the tree models slightly, though the gap between random forest and boosting suggests limited room.
h. **Explainability.** Permutation importance gives a global view only. Individual valuation explanations, such as SHAP values, were not computed here.
i. **Prices, not values.** The target is the recorded price, which can depart from market value through incentives, related party sales or atypical motivation. This is a standard limitation of transaction based AVMs.

4.3 Implications for practice

For lenders and valuers, the results suggest a triage role. A model of this kind can flag valuations that depart sharply from local evidence, and prioritise inspection where the model is least certain: terraced houses and flats, homes without a recent sale, and the cheapest fifth of the market. It should not replace a valuation by a RICS registered valuer, whose standards also require inspection and professional judgement.
For developers of AVMs, two lessons follow. Test designs must be time ordered, and accuracy must be reported by segment and with ratio study measures, not as a single headline figure. The ValuInsight AI prototype will be revised on this basis: its published accuracy statements will be limited to figures that can be reproduced from this study's code and data.

5. Conclusion

Open British transaction data, used carefully, support a useful but limited valuation model for Newcastle upon Tyne and Sunderland. Tested strictly out of time on all 6,857 standard sales of 2024, an ensemble of gradient boosting and random forests valued half of all homes within 12.6 per cent of the price and cut the typical error of a hedonic benchmark by about a third. Its accuracy rests on fine grained local price history, it holds up in thin sector markets, and it is less regressive than the benchmarks. It does not reach the accuracy needed to replace an inspection, and Bank Rate adds nothing over a horizon of one quarter. The next steps are to link Energy Performance Certificate data for floor area, extend the sale history back to 1995, and repeat the test in other regions.
The data are HM Land Registry Price Paid Data, yearly files for 2018 to 2024, available free under the Open Government Licence v3.0. Contains HM Land Registry data © Crown copyright and database right 2025. This data is licensed under the Open Government Licence v3.0. Bank Rate is taken from the Bank of England.

The analysis code is provided as seven Python scripts that reproduce every table and figure: data preparation, feature construction, model tuning and testing, descriptive statistics and hedonic estimation, results, the vertical equity check, and the figures. The analysis used Python 3.13, NumPy 2.5, SciPy 1.18, pandas 3.0 and scikit-learn 1.9, with a random seed of 2025. The code and the cleaned data are available from the author on request and will be deposited in a public repository on acceptance.

Funding. This research received no specific grant from any funding agency in the public, commercial or not for profit sectors.

Use of generative AI. The author used Claude (Anthropic) to assist with generating code. The author reviewed all code, results and text and takes full responsibility for the content.

Relationship to an earlier article. This study builds on Obasse (2025), which set out the conceptual case for the ValuInsight AI approach. The present article is the first full empirical test of that approach. It uses new data, new models and a new test design, and its results supersede the performance figures reported in the earlier article.

Ethics. The study uses only published, property level data with no personal information about buyers or sellers.

Competing interests. The author developed ValuInsight AI (www.valuinsight-ai.com), a property valuation prototype based on the methods evaluated here.

REFERENCES

[1] M. J. Bailey, R. F. Muth, and H. O. Nourse, "A regression method for real estate price index construction," *J. Am. Stat. Assoc.*, vol. 58, no. 304, pp. 933–942, 1963.

[2] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 4765–4774.

[3] HM Land Registry, "About the Price Paid Data." 2025. Accessed: Jul. 11, 2026. [Online]. Available: https://www.gov.uk/about-the-price-paid-data

[4] A. Fisher, C. Rudin, and F. Dominici, "All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously," *J. Mach. Learn. Res.*, vol. 20, no. 177, pp. 1–81, 2019.

[5] D. McMillen and R. Singh, "Assessment regressivity and property taxation," *J. Real Estate Finance Econ.*, vol. 60, no. 1, pp. 155–169, 2020, doi: 10.1007/s11146-019-09715-x.

[6] N. Kok, E.-L. Koponen, and C. A. Martínez-Barbosa, "Big data in real estate? From manual appraisal to automated valuation," *J. Portf. Manag.*, vol. 43, no. 6, pp. 202–211, 2017.

[7] J. Muellbauer and A. Murphy, "Booms and busts in the UK housing market," *Econ. J.*, vol. 107, no. 445, pp. 1701–1727, 1997.

[8] F. Fuerst, P. McAllister, A. Nanda, and P. Wyatt, "Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England," *Energy Econ.*, vol. 48, pp. 145–156, 2015.

[9] M. Mayer, S. C. Bourassa, M. Hoesli, and D. Scognamiglio, "Estimation and updating methods for hedonic valuation," *J. Eur. Real Estate Res.*, vol. 12, no. 1, pp. 134–150, 2019, doi: 10.1108/JERER-08-2018-0035.

[10] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[11] R. J. Hill, "Hedonic price indexes for residential housing: a survey, evaluation and taxonomy," *J. Econ. Surv.*, vol. 27, no. 5, pp. 879–914, 2013.

[12] S. Rosen, "Hedonic prices and implicit markets: product differentiation in pure competition," *J. Polit. Econ.*, vol. 82, no. 1, pp. 34–55, 1974.

[13] S. Malpezzi, "Hedonic pricing models: a selective and applied review," in *Housing Economics and Public Policy*, T. O'Sullivan and K. Gibb, Eds., Oxford: Blackwell, 2003, pp. 67–89.

[14] M. C. Best and H. J. Kleven, "Housing market responses to transaction taxes: evidence from notches and stimulus in the UK," *Rev. Econ. Stud.*, vol. 85, no. 1, pp. 157–193, 2018.

[15] G. Ke *et al.*, "LightGBM: a highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 3146–3154.

[16] S. Mullainathan and J. Spiess, "Machine learning: an applied econometric approach," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 87–106, 2017.

[17] E. A. Antipov and E. B. Pokryshevskaya, "Mass appraisal of residential apartments: an application of Random forest for valuation and a CART-based approach for model diagnostics," *Expert Syst. Appl.*, vol. 39, no. 2, pp. 1772–1778, 2012.

[18] M. Steurer, R. J. Hill, and N. Pfeifer, "Metrics for evaluating the performance of machine learning based automated valuation models," *J. Prop. Res.*, vol. 38, no. 2, pp. 99–129, 2021.

[19] Bank of England, "Official Bank Rate history." 2025. Accessed: Jul. 11, 2026. [Online]. Available: https://www.bankofengland.co.uk/boeapps/database/Bank-Rate.asp

[20] P. Cheshire and S. Sheppard, "On the price of land and the value of amenities," *Economica*, vol. 62, no. 246, pp. 247–267, 1995.

[21] B. Glumac and F. Des Rosiers, "Practice briefing – Automated valuation models (AVMs): their role, their advantages and their limitations," *J. Prop. Invest. Finance*, vol. 39, no. 5, pp. 481–491, 2021, doi: 10.1108/JPIF-07-2020-0086.

[22] S. C. Bourassa, E. Cantoni, and M. Hoesli, "Predicting house prices with spatial dependence: a comparison of alternative methods," *J. Real Estate Res.*, vol. 32, no. 2, pp. 139–159, 2010.

[23] HM Land Registry, "Price Paid Data." 2025. Accessed: Jul. 11, 2026. [Online]. Available: https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads

[24] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019.

[25] G. S. Sirmans, D. A. Macpherson, and E. N. Zietz, "The composition of hedonic pricing models," *J. Real Estate Lit.*, vol. 13, no. 1, pp. 3–43, 2005.

[26] K. E. Case and R. J. Shiller, "The efficiency of the market for single-family homes," *Am. Econ. Rev.*, vol. 79, no. 1, pp. 125–137, 1989.

[27] T. Besley, N. Meads, and P. Surico, "The incidence of transaction taxes: evidence from a stamp duty holiday," *J. Public Econ.*, vol. 119, pp. 61–70, 2014.

[28] S. Gibbons and S. Machin, "Valuing English primary schools," *J. Urban Econ.*, vol. 53, no. 2, pp. 197–219, 2003.

[29] S. Gibbons and S. Machin, "Valuing rail access using transport innovations," *J. Urban Econ.*, vol. 57, no. 1, pp. 148–169, 2005.

[30] V. J. Obasse, "ValuInsight AI: a catalyst for the UK's real estate economy," *TAJMMR Trans Asian J. Mark. Manag. Res.*, vol. 14, no. 10–12, 2025, doi: 10.5958/2279-0667.2025.00014.5.

[31] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM, 2016, pp. 785–794.