Testing Land Valuation Methodologies

Lot 1: Market-based statistical valuation
Lot 2: Advanced algorithmic and machine-learning applications

Prepared for the Welsh Government **March 2026**

alma economics

Table of Contents

List of Tables

List of Figures

Glossary

CoStar: A commercial real estate database providing property information.

Deprivation Index (WIMD): The Welsh Index of Multiple Deprivation, a composite measure assessing relative deprivation across Welsh areas.

Gradient Boosting Machine (GBM): An ensemble machine learning technique that sequentially combines multiple decision trees to model complex, non-linear relationships.

HM Land Registry Price Paid Data: Official data recording the prices of residential property transactions in England and Wales.

Linear Hedonic Pricing Model: A statistical model that estimates property prices based on structural (e.g., size, age), socioeconomic (e.g., neighbourhood income), and environmental (e.g., population density) characteristics.

Lower Super Output Area (LSOA): A geographic unit used in the UK for statistical reporting, typically containing around 1,500 residents.

Lasso (Least Absolute Shrinkage and Selection Operator): A statistical method similar to linear regression that performs automatic feature selection by penalising the absolute size of regression coefficients.

Mean Squared Error (MSE): A metric that quantifies the average squared difference between observed and predicted values, used to assess model accuracy.

MHCLG Energy Performance of Buildings Dataset: A dataset providing information on the energy efficiency of buildings in England and Wales.

Population Density: The number of people living per unit area, often used as an indicator of urbanisation or crowding.

Random Forest: A non-linear algorithm that builds a large number of decision trees during training, capturing complex interactions and non-linear relationships without the need to pre-specify a functional form.

R-squared (R^2): A statistical measure indicating the proportion of variation in the dependent variable explained by the model.

Spatial Hedonic Pricing Model: Similar to the Linear Hedonic Pricing Model, but also incorporates property prices in nearby locations as independent variables to capture geographic spillover effects.

Valuation Office Agency (VOA): UK government agency responsible for property valuations for tax and other official purposes.

Executive Summary

Background

The Welsh Government is seeking to understand the potential of using alternative land valuation methods to inform future policy development in Wales. The objective is to produce a real-world proof of concept for land valuation, with the primary aim of identifying practical challenges, data requirements, limitations, and the applicability of different valuation methods.

To enable comparison and a thorough assessment of the challenges and outputs associated with alternative land valuation methods, the Welsh Government commissioned research from multiple organisations across five lots, with up to three organisations delivering distinct approaches within each lot.

- Lot 1: Market-based statistical valuation
- Lot 2: Advanced algorithmic and machine-learning applications
- Lot 3: Formula-based valuation by land area
- Lot 4: Conventional valuation approaches
- Lot 5: Innovative or experimental approaches

Alma Economics was commissioned to deliver work under Lots 1 and 2.

Methodology

Across both Lots, we applied methods intended for mass valuation, which are quantitative, rely on market prices, and focus on the value of the entire property (land and buildings) rather than land alone. Specifically, our approach follows three key steps to derive land values:

- **Step 1: Property valuations.** In the first step, we use statistical analysis to model property transaction prices as a function of property characteristics and socioeconomic factors. In Lot 1, we apply traditional statistical methods, such as hedonic regressions, whereas in Lot 2, we employ machine learning techniques, including random forests and gradient boosting machine.
- **Step 2: Land-to-building ratios.** In Step 2, property prices and land transaction data are used to derive land-to-building ratios.
- **Step 3: Land valuations.** Lastly, Step 3 combines the predicted property prices for a standardised property in Step 1 with ratios derived in Step 2 to obtain land valuations.

This approach is primarily motivated by the scarcity of land-only transactions. Assessing properties as a whole is more practical and likely to generate more accurate valuations, as there are significantly more property transactions available than land-only sales.

Findings

The table below presents average valuations for the selected Lower Super Output Areas (LSOAs) as specified in the brief, reported for terraced residential and non-residential properties. The valuations are derived from the estimated models, which serve two purposes: (i) to generate valuations for a standardised property—assuming identical property size, type, and attributes across LSOAs—and (ii) to capture differences in location characteristics, proxied by measures of deprivation, household income, rurality, and population density. Consequently, the variation in the estimated valuations shown below reflects differences in location characteristics, with property characteristics held constant.

The valuations are very similar across Lots 1 and 2. There is considerable variability across LSOAs, with valuations ranging from £1.7 to £14.7 per square foot for residential properties and from £0.5 to £11.3 per square foot for commercial properties. Among these areas, Rhondda Cynon Taf (001F) has the lowest average valuations, consistent across both residential and non-residential properties. In contrast, Monmouthshire (006F) exhibits the highest residential valuations, while Cardiff has the highest non-residential land valuations.

Table 1. Land valuations by LSOA (£/sqft)

	Residential (terraced properties)		**Non-residential (commercial properties)**	
	Lot 1	Lot 2	Lot 1	Lot 2
Gwynedd 009D	1.8	1.9	0.6	0.7
Flintshire 015A	5.7	4.8	2.6	2.3
Powys 011C	5.2	5.0	2.8	2.5
Ceredigion 002D	8.5	8.4	9.7	9.9
Pembrokeshire 002F	4.2	4.9	1.8	1.9
Bridgend 019D	9.2	7.5	7.2	7.2
Rhondda Cynon Taf 001F	1.7	1.8	0.5	0.7
Monmouthshire 006F	14.7	14.7	4.6	4.6
Cardiff 032H	14.1	12.1	10.5	11.3

Source: Alma Economics analysis.

The available data do not allow for valuations to be produced at a within-LSOA level. In particular, there is no information on socioeconomic factors (e.g., deprivation, population density) below the LSOA scale, nor are there detailed indicators of location attractiveness that would support valuation estimates within LSOAs.

Conclusion and further considerations

The valuations generated across the two Lots appear reasonable when compared with other valuation benchmarks and when sense-checked against socioeconomic factors, such as population density and deprivation. However, due to data limitations, they are not sufficiently robust to support valuations for the purpose of introducing policy reforms.

The primary limitation relates to land transaction data. Our sample includes only 141 land-only transactions in Wales over the past five years. This limited evidence base constrains the ability to robustly estimate land-to-building ratios and, consequently, land values. A second major limitation concerns non-residential data, which are insufficient to support valuations using mass valuation methods. While a relatively large sample of historical non-residential transactions is available (around 2,000 transactions between 2019 and 2025), there are many LSOAs—68% of them—for which no transactions are observed.

Overall, we have high confidence in the residential property valuations, medium confidence in the non-residential property valuations, and low confidence in the land valuations.

Notwithstanding these limitations, while the approach outlined in this document is not recommended for directly generating land valuations, it could be applied if more suitable data were available—particularly for land prices—or when used alongside other valuation methods. In particular, constraints arising from the limited availability of land transaction data could be mitigated by drawing on supplementary sources, such as property listing websites, to help infer land values. Integrating more robust estimates of land prices with the proposed methodology could lead to significantly improved valuation outcomes.

An alternative way to address limitations in land price data would be to combine the statistical models presented in this research with other valuation approaches, such as residual valuation methods. In practice, this would involve estimating Step 2 using a different methodology, rather than relying solely on market transaction data.

Future research could explore how the limitations identified above might be addressed through the incorporation of additional data sources and complementary approaches, particularly by combining statistical methods with other valuation techniques.

1. Introduction

1.1. Expertise and credentials

Alma Economics is a team of economists, social researchers, data scientists, developers, and creatives working together to address important public policy questions. Our work spans multiple policy areas, including tax, housing and homelessness, education, health and social care, the environment, public finance, and international development, among others.

With regard to land, property, and tax research, our experience includes research in:

- Planning and development
- Property and land tax reforms, council tax, and non-domestic rates
- Social housing strategy, housing affordability, and housing quality standards
- House price modelling and forecasting

Our clients include most UK government departments, the Welsh Government, the Scottish Government, councils across the country, arms-length bodies, and several leading charities. Internationally, we work with major supranational organisations, including the World Bank, UNICEF, the WHO, the OECD, and the European Commission, as well as a number of national governments in Europe, Africa, the Middle East and Asia.

A sample of our research can be found here, and our team can be viewed here.

1.2. Background

Following extensive research into how local taxation could be reformed and improved, and the publication of the Welsh Government's Summary of Findings reviewing the role of local taxes within local government fiscal arrangements, the Welsh Government began exploring the idea of a local land value tax as an alternative taxation mechanism. Within this context, the Welsh Government is seeking to understand the potential of using alternative land valuation methods to inform future policy development in Wales. The objective is to produce a real-world proof of concept for land valuation, providing practical evidence to complement the conceptual research undertaken to date.

To enable comparison and a thorough assessment of the challenges and outputs associated with alternative methods of valuing land, the Welsh Government commissioned research from multiple organisations across five lots, with up to three organisations delivering distinct approaches within each lot.

- **Lot 1: Market-based statistical valuation.** This lot aims to test the use of market transaction data alongside statistical techniques to estimate the value of land in Wales at a granular level.
- **Lot 2: Advanced algorithmic and machine-learning applications.** This lot aims to test the application of advanced automation, supervised and/or unsupervised machine-learning, and/or AI techniques to facilitate the mass appraisal of land in Wales at a granular level.
- **Lot 3: Formula-based valuation by land area.** This lot aims to identify and apply transparent, low-input formulae to approximate a land value by parcel size that can

be benchmarked against more complex and costly valuation methods.

- **Lot 4: Conventional valuation approaches.** This lot involves well-established approaches used by the valuation profession to land in Wales, with an emphasis on extending expertise to value-built land.
- **Lot 5: Innovative or experimental approaches.** This lot category invites methods that are not otherwise covered in the other categories.

The findings from this research will not be used to deliver policy reforms at this stage, but will help identify the practical challenges, data requirements, limitations, and applicability of different valuation methods.

1.3 Related literature

A recent study for Wales developed land valuations primarily using econometric techniques. For residential land values, the study used a two-step approach to address the challenge posed by limited land transaction data: first, estimating residential property values via a hedonic regression model, then calibrating underlying land value by combining these estimates with a sample of land value data from Zoopla. Estimating commercial land prices proved more challenging due to limited data on property characteristics, including basic details, such as size. As a result, the study linked commercial land values to residential estimates.

The ONS has previously used residual valuation and simple formulae to estimate land values in England by land use (residential, industrial, agricultural, commercial). The Republic of Ireland presents an interesting case: while it does not tax land directly, it levies a property tax and uses an online tool to support property valuation and the allocation of the tax burden. Eurostat-OECD's guide, though dated, offers useful insights into valuation methods across OECD countries, most of which lack sufficient market-based land price data. These countries typically rely on residual methods or official assessments. In Korea, for instance, about 1,300 appraisers (2011 data) value sampled plots, with prices extrapolated to adjacent plots.

Our approach aligns more closely with the Welsh study outlined above than with the international studies.

1.4 This report

Alma Economics was commissioned to deliver work under Lots 1 and 2. This report presents that work and is structured as follows:

- Section 2 describes the methodology and data used.
- Section 3 summarises the findings.
- Section 4 compares the models and valuations across the two Lots.
- Section 5 sets out the conclusions and key lessons learned.
- Section 6 provides further considerations.
- Appendices A and B provide summary statistics and model outputs.

2. Methodology

2.1 Data (Lots 1 and 2)

Both Lots 1 and 2 employ quantitative approaches that rely on extensive data. To support the quantitative approaches, we utilised several data sources, including:

- **Residential property data** from two sources: HM Land Registry Price Paid Data and MHCLG's Energy Performance of Buildings dataset.
- **Non-residential property and land price** data from CoStar.
- **Socioeconomic data**, such as deprivation and income, from StatsWales and the ONS.

A summary of these data and sources is provided below.

Residential property data

Property prices were sourced from the HM Land Registry Price Paid Data (LR-PPD), which records actual sale prices and provides comprehensive coverage of all residential transactions in England and Wales. The dataset extends back to 1995 and is updated monthly, offering both a long time series and up-to-date information for model estimation. However, the LR-PPD provides limited information on property attributes, particularly property size, which are crucial for the statistical models in both Lots 1 and 2. To address this, we also used the MHCLG Energy Performance of Buildings Data (EPC), which includes floor area and additional property characteristics, such as number of rooms, construction age, energy ratings, and property type (flat, detached, semi-detached, or terraced). The two datasets were merged to create a comprehensive dataset with transaction prices and property attributes. The combined dataset includes a sample of around 220,000 observations covering 1916 Welsh LSOAs[1], over the period from 2019 to 2025. On average, the dataset provides 115 transactions per LSOA.

Non-residential property data

Non-residential data were drawn from CoStar, one of the UK's largest non-residential property databases. CoStar covers the majority of commercial transactions taking place in England and Wales and provides data on transacted prices and a range of property attributes, such as size, number of floors, age of the building, and location. It also classifies properties into categories, including retail, office, industrial, and other[2].

The dataset includes 2,000 transactions from 2019 to 2025, and covers transactions in 606 out of 1,917 Welsh LSOAs, meaning that 68% of LSOAs have no recorded transactions. Also, among LSOAs with recorded transactions, 80% have fewer than five records over the period. As discussed in the Conclusion, this represents a major limitation of the commercial property valuations.

[1] That is, the dataset covers all LSOAs except one.

[2] The database is compiled using a combination of a large research team and public property records. Initially, a census-level approach is used, with selling agents and buyers contacted to collect transaction data, which is subsequently verified against Land Registry records once they become available. An annual licence is required to access CoStar—when we accessed CoStar, the licence cost around £10,000.

Land transaction data

Land transaction data was used as the basis for calculating land-to-building ratios. This data was drawn from CoStar and includes information about transaction price, location, and size of land. The dataset does not contain details on the land's potential use—such as agricultural, residential, or commercial—or on building permissions.

Moreover, the number of transactions for Wales within the relevant timespan (2019-2025) was limited to 141. To increase the sample size, we expanded the dataset to include land transactions from comparable English regions (Yorkshire and the Humber and the West Midlands) and from previous years (2010 onwards)[3]. As a result, we obtained a larger, albeit imperfect, pool of transactions (approximately 1,500) to estimate land-to-building ratios. Notwithstanding this, as discussed in the Conclusion, the land valuation data available cannot support the development of robust valuations.

Socioeconomic data

In addition to property attributes, the models include proxies for neighbourhood and socioeconomic characteristics, which are expected to affect property prices. We collected data on measures such as LSOA average income, level of deprivation, population density, and categories of urban/rural.

Table 2. Socioeconomic data

Variable	Description	Source	Year	Geography
Welsh Index of Multiple Deprivation	Measure of deprivation, ranking areas from highest to lowest deprivation.	StatsWales	2019	LSOA
Population density	Permanent population by square kilometre.	Census	2021	LSOA
Income	Mean household disposable (net) income, equivalised to account for household size.	ONS	2020	MSOA
Rural/urban classification	Classification of geographies as either rural or urban.	Census	2021	LSOA

Note: Year refers to the year for which the data was recorded; Geography indicates the level of disaggregation of the data. In addition, the National Statistics Postcode Lookup was used to facilitate the matching between our property datasets and socioeconomic variables.

2.2 Lot 1: Rationale behind the methodology

The methods we employed in Lot 1 are designed for mass valuations, are quantitative in nature using market prices, and are based on the value of the property (land and buildings) rather than land alone. The rationale for this approach is threefold:

[3] West Midlands was chosen for its geographical proximity to Wales, and Yorkshire and the Humber for its similarity in terms of construction inputs costs. Based on the Annual Business Survey (2023), we found it to be the region most similar to Wales in terms of four metrics in the construction industry: per capita turnover, per capita approximate gross value added at basic prices, per capita purchases of goods, materials, and services, and per capita employment costs.

- Mass valuations allow efficient assessment of a large number of locations (e.g., LSOAs), which is essential for enabling future policy design.
- The quantitative approach draws on extensive real-world market transactions and valuations, reducing the need for strong assumptions or judgment.
- Valuing properties rather than land alone is also practical, given the limited availability of land-only transactions.

A key limitation of this approach is that it still requires a large volume of land transaction data. In addition, mass valuation methods primarily reflect national and area-level factors and do not capture idiosyncratic features. As a result, they cannot account for parcel-specific characteristics or variations in valuations within LSOAs.

2.3 Lot 1: Research design

Our approach uses three key steps to generate land values.

Step 1: Property valuations. In the first step, we use statistical analysis to model property transaction prices as a function of property characteristics and socioeconomic factors. Estimations are performed separately for residential and non-residential properties using granular data by Lower Super Output Area (LSOA) from 2019 to 2025. The methods used in Lot 1 draw on traditional statistical approaches. These models are used to generate valuation estimates for a “standardised” property across LSOAs. In practice, the approach removes the influence of property attributes such as type, size, and age, which vary across LSOAs, normalising prices and enabling valuations to be generated for an equivalent property across LSOAs.

Step 2: Land-to-building ratios. In Step 2, property prices and land transaction data are used to derive land-to-building ratios. These ratios are calculated across different property and LSOA types to account for variations in the relationship between land and property values across areas and property categories.

Step 3: Land valuations. Lastly, Step 3 combines the predicted property prices for a standardised property in Step 1 with ratios derived in Step 2 to obtain land valuations.

2.4 Lot 1: Statistical analysis

Two types of models were estimated for both residential and non-residential property prices: a Linear Hedonic Pricing model and a Spatial Hedonic Pricing model. A non-technical description of these models is provided in Box 1. In all models, the dependent variable—the variable we modelled—was the natural logarithm of the price per square foot. The independent variables captured a set of property characteristics and socioeconomic factors. For each of the models for Lot 1, we estimated several alternative specifications by varying the type of independent variables and other model characteristics (e.g., the spatial lag specification), and then selected the preferred model based on in-sample (e.g., R-squared, AIC, BIC) and out-of-sample (e.g., R-squared, mean squared error) model fit criteria. For Lot 1, our preferred model is the Spatial Hedonic Pricing Model—however, due to practical limitations, the Linear Hedonic Pricing Model was used to generate valuations for commercial properties. A summary of the models for residential and commercial property prices, including coefficients and model fit diagnostics, is provided in Appendix B.

After selecting the preferred model, we used it to predict property prices for a standardised property to ensure that comparisons across LSOAs reflect differences in land value rather than differences in the composition of properties. The standardised property was defined for each property type in both datasets (flat, detached, semi-detached, terraced, commercial, industrial). For each property type, we constructed the standardised property using the median values of all structural characteristics, such as size and number of rooms[4].

Box 1: Hedonic pricing models

We used two types of models for **Lot 1**: Linear Hedonic Pricing Model and Spatial Hedonic Pricing Model.

- **Linear Hedonic Pricing Model:** This model relies on the assumption that a property's price is determined by the specific bundle of attributes it possesses. These attributes, often classified as structural (e.g., size, age), socioeconomic (e.g., neighbourhood income), and environmental (e.g., population density), can be used to estimate the extent to which each of them affects price.
- **Spatial Hedonic Pricing Model:** This model is similar to the Linear Hedonic Pricing Model, but in addition to property and area characteristics, it uses property prices in nearby locations as independent variables to capture geographical spillover effects. This model offers two advantages over linear models. First, it can increase valuation precision (Bourassa et al., 2020). Second, it prevents artificial "cliff edges" between neighbouring areas, leading to more realistic valuations.

2.5 Lot 1: Land-to-building ratios

We calculated the land-to-building ratio using the property valuations predicted by the model above and the actual land prices. We carried out the analysis by property type and by LSOA. Specifically, we undertook the following steps.

1. **Property prices:** Estimated the price per square foot for all properties as per Step 1 and the approach described in the previous section.
2. **Median property prices by type:** Calculated the median price by property type (flat, detached, semi-detached, terraced, commercial, and industrial) and by LSOA type—LSOAs were classified into categories based on their average net income and their level of urbanity or rurality.
3. **Median land prices by type:** Calculated the median land price per square foot by LSOA type.
4. **Land-to-building ratio:** Divided the median property prices by the median land prices to calculate the land-to-building ratio by property type and LSOA type.

The table below presents the estimated average ratios by property type. For residential properties, the ratios range from 4.2% to 5.7%, compared with 5.8% for commercial properties and 12.7% for industrial properties.

[4] We generated valuations assuming that we are in 2024Q4—our estimates are valuations for 2024Q4.

Table 3. Land-to-building ratio by property type (Lot 1)

Property type	Ratio
Detached	4.2
Semi-Detached	4.9
Flat	5.0
Terraced	5.7
Industrial	12.7
Commercial	5.8

Source: Alma Economics analysis based on data from LR-PPD, EPC, and CoStar.

2.6 Lot 2: Rationale behind the methodology

Similar to Lot 1, the methods we employed in Lot 2 are designed for mass valuations, are quantitative in nature using market prices, and are based on the value of the property (land and buildings) rather than land alone. The rationale for this approach is threefold:

- Mass valuations allow efficient assessment of a large number of locations (e.g., LSOAs), which is essential for enabling future policy design.
- The quantitative approach draws on extensive real-world market transactions and valuations, reducing the need for strong assumptions or judgment.
- Valuing properties rather than land alone is also practical, given the limited availability of land-only transactions.

A key limitation of this approach is that it still requires a large volume of land transaction data. In addition, mass valuation methods primarily reflect national and area-level factors and do not capture idiosyncratic features. As a result, they cannot account for parcel-specific characteristics or variations in valuations within LSOAs.

2.7 Lot 2: Research design

Our research design is similar to that used in Lot 1 and is built around three key steps to generate land values. The only difference is that, in Step 1, Lot 2 uses machine learning models instead of traditional hedonic pricing models.

Step 1: Property valuations. In the first step, we use machine learning approaches to model property transaction prices as a function of property characteristics and socioeconomic factors. Estimations are performed separately for residential and non-residential properties using granular data LSOA from 2019 to 2025. These models are used to generate valuation estimates for a “standardised” property across LSOAs. In practice, the approach removes the influence of property attributes such as type, size, and age, which vary across LSOAs, normalising prices and enabling valuations to be generated for an equivalent property across LSOAs.

Step 2: Land-to-building ratios. In Step 2, property prices and land transaction data are used to derive land-to-building ratios. These ratios are calculated across different property and LSOA types.

Step 3: Land valuations. Lastly, Step 3 combines the predicted property prices for a standardised property in Step 1 with ratios derived in Step 2 to obtain land valuations.

2.8 Lot 2: Statistical analysis

Three types of models were estimated for both residential and non-residential property prices: Least Absolute Shrinkage and Selection Operator, Random Forest, and Gradient Boosting Machine. A non-technical description of these models is provided in Box 2. In all models, the dependent variable—the variable we modelled—was the natural logarithm of the price per square foot. The independent variables or predictors captured a set of property characteristics and socioeconomic factors. After selecting the preferred model, we used it to predict property prices for a standardised property to ensure that comparisons across LSOAs reflect differences in land value rather than differences in the composition of properties. The standardised property was defined for each property type in both datasets (flat, detached, semi-detached, terraced, commercial, industrial). For each property type, we constructed the standardised property using the median values of all structural characteristics, such as size and number of rooms[5].

2.9 Lot 2: Land-to-building ratios

We calculated the land-to-building ratio using the property valuations predicted by the model above and the actual land prices. We carried out the analysis by property type and by LSOA. Specifically, we undertook the following steps:

1. **Property prices:** Estimated the price per square foot for all properties as per Step 1 and the approach described in the previous section.
2. **Median property prices by type:** Calculated the median price by property type (flat, detached, semi-detached, terraced, commercial, and industrial) and by LSOA type—LSOAs were classified into categories based on their average net income and their level of urbanity or rurality.
3. **Median land prices by type:** Calculated the median land price per square foot by LSOA type.
4. **Land-to-building ratio:** Divided the median property prices by the median land prices to calculate the land-to-building ratio by property type and LSOA type.

The table below presents the estimated average ratios by property type. For residential properties, the ratios range from 3.9% to 5.3%, compared with 7.7% for commercial properties and 11.1% for industrial properties.

Table 4. Land-to-building ratio by property type (Lot 2)

Property type	Ratio
Detached	3.9
Semi-Detached	4.7
Flat	4.8
Terraced	5.3
Industrial	11.1
Commercial	7.7

Source: Alma Economics analysis based on data from LR-PPD, EPC, and CoStar.

5 We generated valuations assuming that we are in 2024Q4—our estimates are valuations for 2024Q4.

Box 2: Machine learning models

We used three types of models for **Lot 2:** Least Absolute Shrinkage and Selection Operator, Random Forest, and Gradient Boosting Machine.

- **Least Absolute Shrinkage and Selection Operator (Lasso):** Lasso operates by adding a penalty term to the standard linear regression loss function. This penalty forces the model to shrink the coefficients of less important variables toward zero, thereby acting as an automatic feature selection tool. This approach yields a more parsimonious model compared to standard regression models. Furthermore, this method helps to mitigate the effects of multicollinearity and prevents overfitting. Also, compared to other machine learning methods, Lasso is less of a "black box" than other methods and is valued for its ability to deliver strong predictive performance without sacrificing transparency.
- **Random Forest:** Random Forest is a non-linear algorithm that operates by constructing a multitude of decision trees during training. Each individual tree is built using a random subset of the training data and a random subset of the available features. The final prediction is determined by averaging the predictions of all the individual trees (a process called bootstrap aggregation). This averaging process is its core strength, as it significantly reduces variance and prevents the model from overfitting the data, which is a common issue with single decision trees. Random Forest can inherently capture complex interactions and non-linear relationships without requiring pre-specification of a functional form. However, it is less transparent than traditional models, as the model selection process is data-driven and does not directly reveal the impact of individual variables on the outcome of interest.
- **Gradient Boosting Machine:** Gradient Boosting Machine is an ensemble machine learning technique that combines the outputs of multiple sequential decision trees. In this process, each successive tree is trained to minimise the residual errors of the combined predictions from all previous trees. This minimisation is achieved through the iterative application of the gradient descent optimisation algorithm. A key characteristic of Gradient Boosting Machine is its ability to achieve high predictive accuracy by systematically correcting the weaknesses of prior models. While it often outperforms Random Forest in handling complex non-linearities, it is more sensitive to overfitting and slower to train. As is the case with Random Forest, the Gradient Boosting Machine is less transparent than traditional models in the sense that we cannot recover the impact of individual variables on the outcome of interest.

3. Findings

3.1 Lot 1

The table below summarises land valuations based on the models used in Lot 1 across different property types. Residential land values are higher than commercial and industrial ones, reflecting the patterns observed in property prices. Average residential land valuations are estimated at around £10.5 per sqft, while commercial valuations are significantly lower at £6.8 per sqft.

Table 5. Land valuations by property type (Lot 1)

	£/sqft
Detached	10.2
Semi-Detached	10.5
Terraced	10.7
Flat	10.1
Commercial	6.8
Industrial	5.9

Source: Alma Economics analysis.

The table below presents average valuations for the selected LSOAs specified in the brief, reported for terraced residential and non-residential properties. Valuations for other property types are reported in Appendix C. The valuations are derived from the estimated models, which serve two purposes: (i) to generate valuations for a standardised property—assuming identical property size, type, and attributes across LSOAs—and (ii) to capture differences in location characteristics, proxied by measures of deprivation, household income, rurality, and population density. Consequently, the variation in the estimated valuations shown below reflects differences in location characteristics, with property characteristics held constant[6].

There is considerable variability across LSOAs, with valuations ranging from £1.7 to £14.7 per square foot for residential properties and from £0.50 to £10.50 per square foot for commercial properties. Among these areas, Rhondda Cynon Taf (001F) has the lowest average valuations, consistent across both residential and non-residential properties. In contrast, Monmouthshire (006F) exhibits the highest residential valuations, while Cardiff has the highest non-residential land valuations.

[6] The valuations are based on the full set of available transaction data rather than data specific to individual LSOAs. The residential model draws on over 220,000 transactions across the sample period, while the commercial model uses approximately 2,000 transactions. These transactions are used to estimate the models and quantify the effects of property and area characteristics on prices. By relying on large samples, this approach reduces the impact of sparse or missing transaction data at the LSOA level and limits sensitivity to irregular or outlier observations. Irregular transaction prices—that is, very low or very high prices per square foot—were identified using summary statistics at the national level and were excluded from the analysis.

The available data do not allow for valuations to be produced at a within-LSOA level. In particular, there is no information on deprivation or population density below the LSOA scale, nor are there detailed indicators of location attractiveness that would support valuation estimates within LSOAs.

Table 6. Land valuations by LSOA (Lot 1, £/sqft)

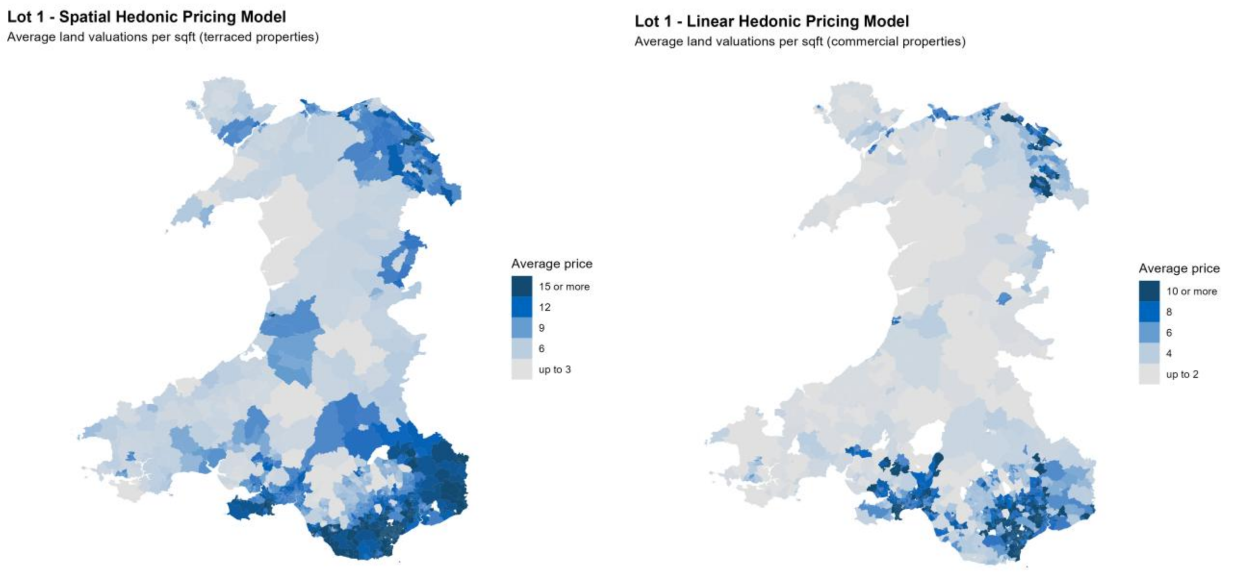
	Residential (terraced properties)	Non-residential (commercial properties)
Gwynedd 009D	1.8	0.6
Flintshire 015A	5.7	2.6
Powys 011C	5.2	2.8
Ceredigion 002D	8.5	9.7
Pembrokeshire 002F	4.2	1.8
Bridgend 019D	9.2	7.2
Rhondda Cynon Taf 001F	1.7	0.5
Monmouthshire 006F	14.7	4.6
Cardiff 032H	14.1	10.5

Source: Alma Economics analysis.

In addition to the values presented in the table, valuations have been produced for the entirety of Wales. Residential valuations were calculated using the Spatial Hedonic model, while commercial valuations employed the Linear Hedonic model, even though the Spatial model showed stronger statistical performance. The Linear model was used due to insufficient data for all LSOAs, as the Spatial model requires complete coverage to generate valuations.

A summary of these valuations is presented in the heat map below. The highest valuations are concentrated in the southeast and northeast of Wales, particularly in (but not limited to) urban areas such as Cardiff, Bridgend, and Wrexham. In contrast, land valuations are significantly lower in rural areas. As discussed above, data availability does not allow for valuations to be produced at a within-LSOA level.

Figure 1. Land valuation heat map (Lot 1)



Source: Alma Economics analysis.

3.2 Lot 2

The table below summarises land valuations based on the models used in Lot 2 across different property types. Similar to Lot 1, residential land values are higher than commercial and industrial ones, reflecting the patterns observed in property prices. Average residential land valuations are estimated at around £10.5 per sqft, while commercial valuations are significantly lower at £6.8 per sqft.

Table 7. Land valuations by property type (Lot 2)

	£/sqft
Detached	10.1
Semi-Detached	10.5
Terraced	10.3
Flat	10.5
Commercial	6.8
Industrial	5.7

Source: Alma Economics analysis.

The table below presents average valuations for the selected LSOAs specified in the brief, reported for terraced residential and non-residential properties. Valuations for other property types are reported in Appendix C. The valuations are derived from the estimated models, which serve two purposes: (i) to generate valuations for a standardised property—assuming identical property size, type, and attributes across LSOAs—and (ii) to capture differences in location characteristics, proxied by measures of deprivation, household income, rurality, and population density. Consequently, the variation in the estimated valuations shown below reflects differences in location characteristics, with property characteristics held constant[7].

There is considerable variability across LSOAs, with valuations ranging from £1.8 to £14.7 per square foot for residential properties and from £0.7 to £11.3 per square foot for commercial properties. Among these areas, Rhondda Cynon Taf (001F) has the lowest average valuations, consistent across both residential and non-residential properties. In contrast, Monmouthshire (006F) exhibits the highest residential valuations, while Cardiff has the highest non-residential land valuations.

The available data do not allow for valuations to be produced at a within-LSOA level. In particular, there is no information on deprivation or population density below the LSOA scale, nor are there detailed indicators of location attractiveness that would support valuation estimates within LSOAs.

Table 8. Land valuations by LSOA (Lot 2, £/sqft)

	Residential (terraced properties)	Non-residential (commercial properties)
Gwynedd 009D	1.9	0.7
Flintshire 015A	4.8	2.3
Powys 011C	5.0	2.5
Ceredigion 002D	8.4	9.9
Pembrokeshire 002F	4.9	1.9
Bridgend 019D	7.5	7.2
Rhondda Cynon Taf 001F	1.8	0.7
Monmouthshire 006F	14.7	4.6
Cardiff 032H	12.1	11.3

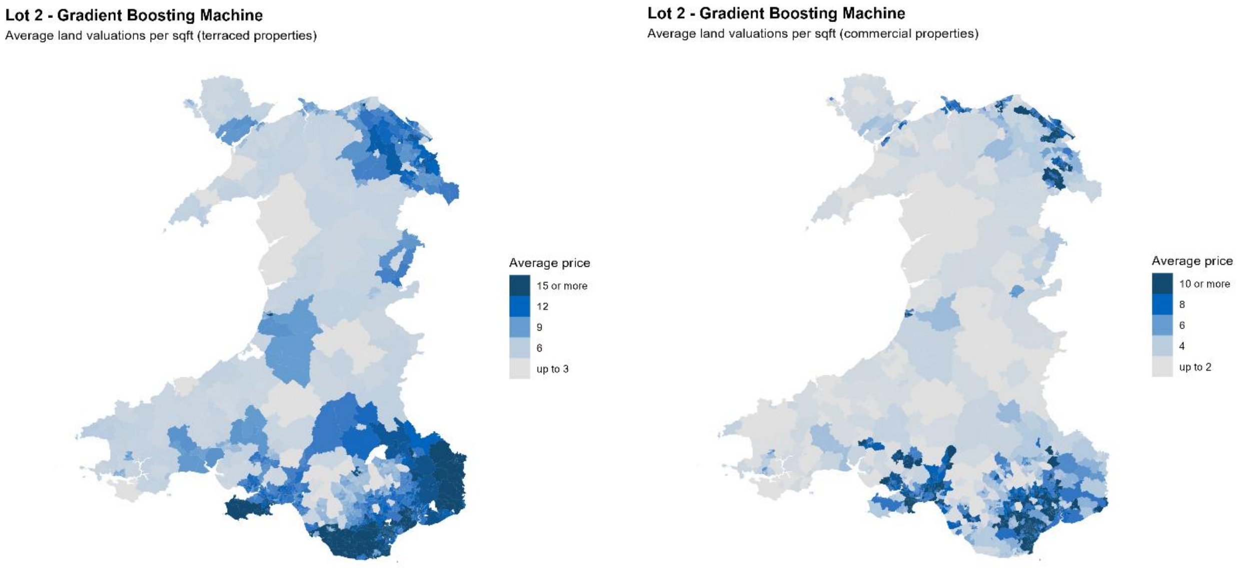
Source: Alma Economics analysis.

Beyond the values shown in the table, valuations have been generated for the whole of

[7] The valuations are based on the full set of available transaction data rather than data specific to individual LSOAs. The residential model draws on over 220,000 transactions across the sample period, while the commercial model uses approximately 2,000 transactions. These transactions are used to estimate the models and quantify the effects of property and area characteristics on prices. By relying on large samples, this approach reduces the impact of sparse or missing transaction data at the LSOA level and limits sensitivity to irregular or outlier observations. Irregular transaction prices—that is, very low or very high prices per square foot—were identified using summary statistics at the national level and were excluded from the analysis.

Wales. A summary of these valuations is presented in the heat map below. The highest valuations are concentrated in the southeast and northeast of Wales, particularly in (but not limited to) urban areas such as Cardiff, Bridgend, and Wrexham. In contrast, land valuations are significantly lower in rural areas. As discussed above, data availability does not allow for valuations to be produced at a within-LSOA level.

Figure 2. Land valuation heat map (Lot 2)



Source: Alma Economics analysis.

4. Comparison Between Lots 1 and 2

This section synthesises the findings from Lots 1 and 2, comparing the alternative models and the valuations they produce across the two Lots. It also includes a summary of a benchmarking exercise that compares the resulting valuations with other available valuations and data.

4.1 Statistical performance

We evaluated the performance of the different models across Lots using out-of-sample statistics. We trained the models using 80% of the data and then tested them on the remaining 20%. The out-of-sample performance of these models is shown in the table below.

Table 9. Model fit statistics

	Residential					Commercial				
	Lot 1		Lot 2			Lot 1		Lot 2		
Model fit	Linear	Spatial	Lasso	Random forest	Gradient boosting	Linear	Spatial	Lasso	Random forest	Gradient boosting
R-squared	0.511	0.628	0.517	0.511	**0.648**	0.248	0.304	0.186	0.281	**0.430**
MSE	0.088	0.067	0.087	0.088	**0.064**	0.689	0.638	0.680	0.601	**0.476**

Source: Alma Economics analysis.

For residential properties, all models provide a high R-squared, greater than 0.5[8]. For Lot 1, our preferred model is the Hedonic Spatial Model, and for Lot 2, our preferred model is the Gradient Boosting Machine, based on both the R-squared and Mean Squared Error (MSE). The Gradient Boosting Machine provides the best fit across all models, including the Spatial Hedonic Pricing model. For commercial properties, R-squared is significantly lower, ranging from 0.19 to 0.43. As in the case of residential properties, the Gradient Boosting Machine is the best-performing model.

Nonetheless, the valuations produced by the models are very similar, as evidenced by the high correlations shown in the table below. This suggests that different model types are likely to generate broadly comparable valuations.

[8] R-squared indicates how well the model fits the data and takes values between 0 and 1. Although there is no established threshold for what constitutes a low or high value, an R-squared of 50% is typically considered satisfactory or high, particularly for cross-sectional data and for outcomes such as property prices, which may be influenced by unobserved factors.

Table 10. Correlation between model valuations (Residential)

	Linear HPM	Spatial HPM	Lasso	Random Forest	Gradient Boosting Machine
Linear HPM	1.00				
Spatial HPM	0.95	1.00			
Lasso	1.00	0.94	1.00		
Random Forest	0.98	0.94	0.98	1.00	
Gradient Boosting Machine	0.98	0.96	0.98	0.98	1.00

Source: Alma Economics analysis.

4.2 Valuations

The table below summarises land valuations based on the preferred models used in Lots 1 and 2 across different property types. Overall, the valuations are closely aligned between the two Lots in line with the correlation analysis presented above.

Table 11. Land valuations by property type (£/sqft)

	Lot 1	Lot 2
Detached	10.2	10.1
Semi-Detached	10.5	10.5
Terraced	10.7	10.3
Flat	10.1	10.5
Commercial	6.8	6.8
Industrial	5.9	5.7

Source: Alma Economics analysis.

The table below presents a comparison of average valuations for the selected LSOAs set out in the brief across Lots 1 and 2. Overall, the two Lots generate similar residential land valuations, with notable differences in Cardiff and Bridgend, and smaller divergences in Flintshire and Pembrokeshire. Commercial valuations show a higher degree of consistency between the two Lots. Understanding the observed differences and similarities between Lots 1 and 2 is challenging. There is no clear relationship between valuation differences and either rural/urban classification or the overall size of the valuation (for example, Lot 1 is not consistently higher than Lot 2). Furthermore, the nature of the Machine Learning models, which prevents tracing the impact of individual factors, makes it difficult to determine the key drivers behind the observed differences and similarities.

Table 12. Land valuations by LSOA (£/sqft)

	Residential (terraced properties)		Non-residential (commercial properties)	
	Lot 1	Lot 2	Lot 1	Lot 2
Gwynedd 009D	1.8	1.9	0.6	0.7
Flintshire 015A	5.7	4.8	2.6	2.3
Powys 011C	5.2	5.0	2.8	2.5
Ceredigion 002D	8.5	8.4	9.7	9.9
Pembrokeshire 002F	4.2	4.9	1.8	1.9
Bridgend 019D	9.2	7.5	7.2	7.2
Rhondda Cynon Taf 001F	1.7	1.8	0.5	0.7
Monmouthshire 006F	14.7	14.7	4.6	4.6
Cardiff 032H	14.1	12.1	10.5	11.3

Source: Alma Economics analysis.

4.3 Benchmarking

We compared the land value estimates against a series of quantitative benchmarks to sense-check our estimates; both the overall level of the valuations and their distribution were reasonable.

VOA estimations

We compared our land value estimates with a sample of valuations provided by the Valuation Office Agency (VOA) for England. The VOA developed these estimates specifically for policy appraisal purposes. Although no estimates are available for Wales, we found that our valuations were broadly in line with the VOA figures in comparable regions (West Midlands and Yorkshire and the Humber), indicating that our approach produces, on average, sensible estimates.

Construction costs

We estimated land valuations using property prices and construction costs. Land values were derived as the residual obtained by subtracting per-square-foot construction costs from per-square-foot property prices. The results show a strong correlation between valuations from this method and our approach (a correlation of 0.69 for residential properties), with broadly similar overall levels. However, notable discrepancies arise in some locations. In particular, the construction-costs approach appears to capture greater variability between LSOAs, producing, for example, substantially higher estimates in more densely populated areas.

Population density and WIMD

Lastly, we compared our estimated land valuations to population density and the deprivation index (WIMD) at the LSOA level. The rationale for comparing valuations with population density is that areas with higher population density typically experience greater demand for space, leading to higher land values. Indeed, we observe a positive correlation between land valuations and population density (correlation of 0.44 for residential and 0.43 for non-residential properties). Similarly, we find a positive, albeit smaller, correlation between the deprivation index and both residential (0.30) and non-residential (0.17) land valuations, reflecting the expected pattern that more deprived areas tend to have lower land values.

5. Conclusions

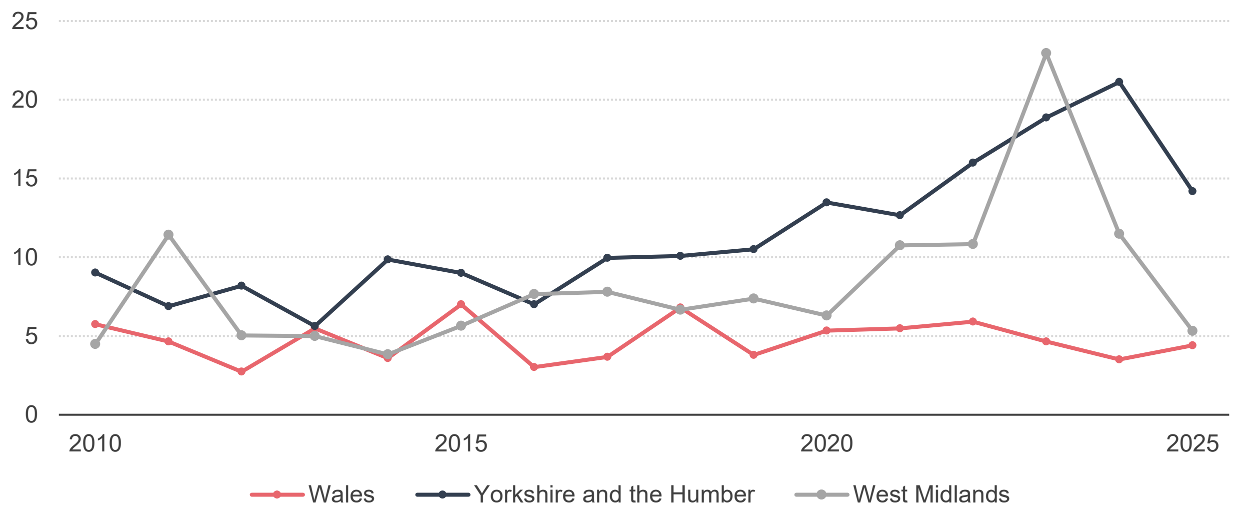
The valuations we have generated across the two Lots are sensible when compared with other valuations and when sense-checked against socioeconomic factors (population density and deprivation). **However, due to data limitations, they are not robust enough to determine valuations for the purpose of introducing policy reforms.** These data limitations relate primarily to the sample of land and non-commercial property transactions, discussed below in detail.

Limitation 1: Land transaction data are not sufficient to support valuations based on mass valuation methods

Sample size: The primary limitation concerns the small number of transactions—only 141 for Wales over the past five years. This limited sample restricts the ability to robustly estimate land-to-building ratios and, consequently, land values. Although we expanded the dataset by including transactions from a longer historical period and comparable regions in England, the total number of observations remains relatively low, at around 1,500 over 15 years.

Volatility: The land prices are volatile (see the plot below), potentially reflecting the small sample and idiosyncrasies of the specific localities in which these transactions occurred.

Figure 3. Median land prices (£ per sqft) over time



Source: Alma Economics analysis based on data from LR-PPD, EPC, and CoStar.

Adjusting for land price inflation: We pooled land transactions across several years to increase the sample size, which may not reflect current market valuations. We tried to adjust the historical land price data for land price inflation, but the volatility characterising the data prevents us from identifying and adjusting for inflation.

Land use type: The constraints are not limited to sample size and composition but also relate to the range of available information. In particular, more detailed data on land use types would allow for a more refined approach that generates valuations by land use type.

Representativeness: Land transacted in recent years may not be representative of the land underlying existing built properties. For example, recent transactions of undeveloped land with planning permission in prime urban areas are uncommon, simply because very little unbuilt land remains in these locations. This means that, on average, the land transactions we observe could be for less desirable locations, which could bias our land valuations downwards.

Limitation 2: Non-residential data are not sufficient to support valuations based on mass valuation methods

While we have a large sample of historical non-residential transactions—2,000 transactions from 2019 to 2025—there are many LSOAs, 68% of them, for which we have no transactions. Although our approach accounts for missing data and can extrapolate valuations from LSOAs with data to those without, the lack of data for such a large proportion of LSOAs raises concerns about the robustness of the valuations. The low model fit performance of the commercial models presented in the previous section also corroborates this concern.

Limitation 3: Mass valuation methods cannot capture land-specific idiosyncratic factors

Mass valuation methods inherently reflect national and area-level factors but do not capture idiosyncratic features. In other words, the approaches presented in this report cannot account for parcel-specific factors or variation in valuations within LSOAs.

Key advantage of the approach

Notwithstanding the limitations of the land and non-residential data, the residential property data used are relatively rich, comprising a large sample of historical transactions and broad coverage of property attributes. The combination of this extensive dataset and the statistical modelling approach provides a satisfactory model fit and sensible valuations of residential property prices. A further advantage of this approach is its scalability and reproducibility: once estimated, the models can generate property valuations across all LSOAs in Wales and can be readily updated as new data become available.

Overall assessment

In summary, based on the limitations and discussion above, we have high confidence in the residential property valuations, medium confidence in the non-residential property valuations, and low confidence in land valuations, as summarised in the table below.

Table 13. Confidence level

Valuation category	Level of confidence
Residential property valuations	High
Non-residential property valuations	Medium
Residential land valuations	Low
Non-residential land valuations	Low

Lessons learned

Policy-oriented: Applications of market price-based mass valuation methods to support policy development in Wales are challenging primarily due to the lack of sufficient market data on land transactions.

Data-related: There is rich data on residential property prices and attributes that could be used to support valuations in Wales by combining information from the Land Registry and the MHCLG EPC dataset. In contrast, data on non-residential property prices and attributes are more difficult to obtain for a sufficiently large sample.

Statistical-related: Machine Learning models perform better than Linear and Spatial hedonic regression models. However, the correlation in forecasts between the different types of models is quite high. Conventional hedonic regression models, therefore, perform sufficiently well, potentially without the need to apply more computationally intensive[9] (spatial lag) or less transparent (machine learning) methods.

[9] The spatial specification adds a significant computational burden to the analysis. The main challenge in estimating a spatial lag model arises from the computational intensity of estimating the spatial weights matrix. These matrices are of size $N \times N$, where N is the sample size. Consequently, the larger the sample, the more computationally demanding it becomes to generate the spatially lagged variable(s). Given our large sample size in the residential dataset, for example, generating spatial lag variables required splitting the sample into smaller segments. Although larger samples are desirable for accuracy and robustness, they create a trade-off between precision and computational feasibility when using this method. In addition, the spatial model does not allow for out-of-sample prediction. While we can use the characteristics of the standardised property and the LSOA to predict property prices, estimating the spatial lag for LSOAs with no recorded transactions is not straightforward, because the spatial lag variables are missing.

6. Further Considerations

While the approach outlined in this document is not recommended for directly generating land valuations, it could still be used if more **suitable data were available, especially for land prices, or alongside other methods**. Specifically:

More data: Constraints related to the limited availability of land transaction data could be addressed by drawing on supplementary data sources, such as property listing websites like Zoopla and Rightmove, to infer land values. Integrating a more robust estimate of land prices with the proposed methodology could lead to notably improved valuation results.

Other methods: Another way to address the limitations of the land price data is to combine the statistical models presented in this research with alternative methods, such as residual valuation approaches. In practice, this would involve estimating Step 2 using a different methodology, rather than relying solely on market transaction data.

Future research could examine how the limitations outlined above might be addressed by incorporating additional data and complementary approaches, particularly by combining statistical methods with other valuation techniques.

From a policy perspective, future research could explore the appropriate level of geographical granularity at which valuations can be developed and applied, potentially capturing within-LSOA variations and accounting for differences in valuations related to local amenities.

Appendix A: Descriptive statistics

This section presents summary statistics for the property and land prices used in the analysis. Tables 14 and 15 report average property prices across all LSOAs, while Tables 16 and 17 present prices for the LSOAs specified in the brief. Tables 14 and 15 provide detailed statistics by property type and characteristics, whereas Tables 16 and 17 report average statistics only, reflecting the small number of transactions available.

Table 14. Average price per sqft, residential properties

Variable	Category	Average price per square foot (£/sqft)	Number of transactions
Price	All	207	221,511
Total floor area	Quartile 1 (smallest floor area)	227	57,491
	Quartile 2	193	53,542
	Quartile 3	195	55,414
	Quartile 4 (largest floor area)	210	54,377
Number of rooms	1	234	390
	2	229	5,955
	3	219	31,726
	4	199	55,761
	5	199	69,580
	6	209	31,076
	7+	218	26,336
Property type	Detached	249	59,605
	Flat	213	15,402
	Semi-Detached	208	67,400
	Terraced	171	78,417
New build	Yes	185	127
	No	207	220,697
Construction age	Before 1900	184	36,504
	1900-1929	167	41,710
	1930-1949	209	20,699
	1950-1966	203	31,774
	1967-1975	223	27,474
	1976-1982	235	12,804
	1983-1990	246	13,749
	1991-1995	244	7,782
	1996-2002	248	12,097
	2003-2006	229	9,775

	2007-2011	224	5,908
	2012-2021	247	548
Energy efficiency rating	A	244	120
	B	225	3,277
	C	219	54,462
	D	204	104,645
	E	197	43,556
	F	205	11,140
	G	204	3,624

Source: Alma Economics analysis based on data from LR-PPD and EPC.

Table 15. Average price per sqft, non-residential properties

Variable	Category	Average price per square foot (£/sqft)	Number of transactions
Price	All	136	2056
Building square feet	Quartile 1 (smallest floor area)	191	514
	Quartile 2	134	514
	Quartile 3	113	515
	Quartile 4 (largest floor area)	107	513
Building age	0 to 30 years old	184	607
	31 to 60 years old	109	666
	61 to 90 years old	128	235
	Over 90 years old	121	548
Building material	Masonry	148	1171
	Other	130	105
	Steel	120	780
Building type	Health, hospitality, speciality, sports, students	208	178
	Industrial	65	522
	Office	122	383
	Retail	167	973
Building location	Central Business District (CBD)	185	317
	Suburban	132	1294
	Urban	113	445

Source: Alma economics analysis using data from CoStar.

Table 16. Average price per sqft, residential properties

LSOA	Average price per square foot (£/sqft)	Number of transactions
Gwynedd	172	121
Flintshire	216	184
Powys	193	142
Ceredigion	188	70
Pembrokeshire	220	83
Bridgend	216	138
Rhondda Cynon Taf	182	95
Monmouthshire	312	157
Cardiff	261	220

Source: Alma Economics analysis based on data from LR-PPD and EPC.

Table 17. Average price per sqft, non-residential properties

LSOA	Average price per square foot (£/sqft)	Number of transactions
Flintshire	38	7
Ceredigion	152	1
Rhondda Cynon Taf	17	1
Cardiff	240	80

Source: Alma economics analysis using data from CoStar.

Appendix B: Model estimates

The following tables present the model coefficients and goodness-of-fit statistics for the preferred models within the scope of Lot 1.

Table 18. Residential property model (spatial hedonic model)

Variables	**Dependent variable: ln(price per sqft)**
Total floor area in square feet	-0.000185*** (2.45e-05)
Property type, Detached (base category)	
Property type, Flat	-0.383*** (0.0121)
Property type, Semi-detached	-0.182*** (0.00506)
Property type, Terraced	-0.309*** (0.00612)
New build = Yes	-0.00221 (0.0372)
Construction year band, before 1900 (base category)	
Construction year band, 1900-1929	-0.0417*** (0.00363)
Construction year band, 1930-1949	-0.00936 (0.00575)
Construction year band, 1950-1966	-0.0486*** (0.00597)
Construction year band, 1967-1975	-0.0191*** (0.00616)
Construction year band, 1976-1982	0.00619 (0.00640)
Construction year band, 1983-1990	0.0435*** (0.00833)
Construction year band, 1991-1995	0.0646*** (0.00769)
Construction year band, 1996-2002	0.0666*** (0.00794)
Construction year band, 2003-2006	0.0802*** (0.00968)
Construction year band, 2007-2011	0.104*** (0.0109)
Construction year band, 2012-2021	0.127***

	(0.0234)
Freehold (base category)	
Leasehold	-0.0492*** (0.00651)
Number of habitable rooms: **Habitable rooms include any living room, sitting room**	-0.00900* (0.00486)
Current energy efficiency rating, A (base category)	
Current energy efficiency rating, B	0.0117 (0.0329)
Current energy efficiency rating, C	0.0136 (0.0317)
Current energy efficiency rating, D	0.0191 (0.0319)
Current energy efficiency rating, E	0.0147 (0.0325)
Current energy efficiency rating, F	0.0226 (0.0333)
Current energy efficiency rating, G	-0.0114 (0.0355)
Current energy efficiency score (numeric)	0.000950*** (0.000175)
LSOA WIMD, Most deprived areas (up to rank 582) (base category)	
LSOA WIMD, Quartile 2	0.0652*** (0.00596)
LSOA WIMD, Quartile 3	0.108*** (0.00685)
LSOA WIMD, Least deprived areas (from rank 1487)	0.132*** (0.00791)
LSOA urban/rural classification, Urban: **Nearer to a major town or city (base category)**	
LSOA urban/rural classification, Urban: Further from a major town or city	0.00629 (0.00581)
LSOA urban/rural classification, Larger rural: Nearer to a major town or city	-0.0439*** (0.0103)
LSOA urban/rural classification, Larger rural: Further from a major town or city	-0.00579 (0.00719)
LSOA urban/rural classification, Smaller rural: Nearer to a major town or city	-0.00823 (0.0118)
LSOA urban/rural classification, Smaller rural: Further from a major town or city	-0.0123* (0.00745)
MSOA average net household income	8.22e-06*** (9.32e-07)

LSOA population density	5.63e-06*** (1.24e-06)
Price per SF spatial lag (exp 1)	0.837*** (0.0133)
Constant	0.693*** (0.0705)
Quarter fixed effects	Yes
Observations	220,824
R-squared	0.624
MSE	0.0676
MAPE (%)	3.790
BIC	32468
AIC	31839
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

Table 19. Non-residential property model (linear hedonic model)

Variables	**Dependent variable: ln(price per sqft)**
Building sqft	-1.50e-06*** (3.36e-07)
Age	-0.00256*** (0.000565)
Number of Floors	-0.0273 (0.0301)
Parking spaces per 1,000 square feet	0.0204** (0.00878)
Property type, Industrial (base category)	
Property type, Health, hospitality, speciality, sports, students	0.800*** (0.0975)
Property type, Office	0.545*** (0.0721)
Property type, Retail	0.775*** (0.0670)
Location type, Central business district (base category)	
Location type, Suburban	-0.457*** (0.0907)
Location type, Urban	-0.538*** (0.0897)

Building materials, Masonry (base category)	
Building materials, Other	-0.149 (0.101)
Building materials, Steel	-0.0362 (0.0666)
LSOA WIMD, Most deprived areas (up to rank 379) (base category)	
LSOA WIMD, Quartile 2	-0.0374 (0.0763)
LSOA WIMD, Quartile 3	0.246*** (0.0730)
LSOA WIMD, Least deprived areas (from rank 1142)	0.163* (0.0962)
LSOA urban/rural classification, Urban: Nearer to a major town or city (base category)	
LSOA urban/rural classification, Urban: Further from a major town or city	-0.157** (0.0776)
LSOA urban/rural classification, Larger rural: Nearer to a major town or city	-0.00647 (0.201)
LSOA urban/rural classification, Larger rural: Further from a major town or city	0.150 (0.111)
LSOA urban/rural classification, Smaller rural: Nearer to a major town or city	0.145 (0.116)
LSOA urban/rural classification, Smaller rural: Further from a major town or city	-0.258** (0.109)
MSOA average net household income	2.87e-05*** (8.75e-06)
LSOA population density	1.17e-05 (1.12e-05)
Constant	3.663*** (0.275)
Quarter fixed effects	Yes
Observations	2,056
R-squared	0.247
MSE	0.698
MAPE (%)	18.18
BIC	5405
AIC	5141

Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Appendix C: LSOA valuations

The tables below present the estimated land valuations across the nine LSOAs, broken down by property type. The terraced and commercial valuations are included in the main body of the report and are reproduced here for ease of comparison.

Table 20. LSOA land valuations (£/sqft): Lot 1

LSOA	Flat	Terraced	Semi-Detached	Detached	Commercial	Industrial
Gwynedd 009D	1.6	1.8	1.7	1.4	0.6	1.5
Flintshire 015A	4.9	5.7	5.4	5.2	2.6	1.8
Powys 011C	4.1	5.2	5.1	4.8	2.8	1.6
Ceredigion 002D	7.0	8.5	8.2	7.7	9.7	8.7
Pembrokeshire 002F	3.7	4.2	4.5	4.3	1.8	1.0
Bridgend 019D	6.9	9.2	8.4	7.7	7.2	6.5
Rhondda Cynon Taf 001F	1.5	1.7	1.5	1.4	0.5	1.1
Monmouthshire 006F	12.4	14.7	14.1	14.3	4.6	4.8
Cardiff 032H	14.4	14.1	13.4	13.1	10.5	7.8

Source: Alma Economic analysis.

Table 21. LSOA land valuations (£/sqft): Lot 2

LSOA	Flat	Terraced	Semi-Detached	Detached	Commercial	Industrial
Gwynedd 009D	1.8	1.9	1.8	1.6	0.7	1.2
Flintshire 015A	4.2	4.8	4.8	4.4	2.3	1.3
Powys 011C	4.6	5.0	5.0	4.8	2.5	1.9
Ceredigion 002D	8.2	8.4	8.1	6.5	9.9	5.6
Pembrokeshire 002F	4.5	4.9	4.8	4.8	1.9	1.3
Bridgend 019D	7.4	7.5	7.2	6.8	7.2	6.7
Rhondda Cynon Taf 001F	1.7	1.8	1.7	1.5	0.7	1.6
Monmouthshire 006F	12.0	14.7	13.7	12.8	4.6	4.6
Cardiff 032H	12.2	12.1	11.5	10.5	11.3	6.0

Source: Alma Economic analysis.

+44 20 8133 3192 43 Tanner Street, SE1 3PL, London, UK

LinkedIn company/alma-economics X almaeconomics

alma economics