Analysis of property prices in Portsmouth

University of Glasgow

Ivan Carnevali 2946257C@student.gla.ac.uk

Scan to download the poster!

Analysis plan

Few variables of interest will be extracted from the dataset in order to investigate, after fitting an appropriate model, how the sale price is influenced by different properties of the houses.

Variables of interest

Explanatory variables

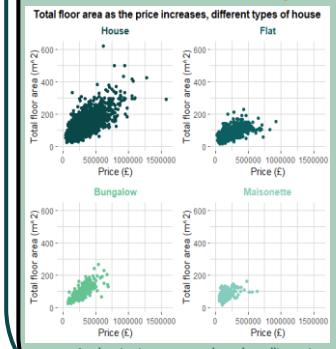
- Total floor area
- Postcode
- Energy efficiency

Covariates

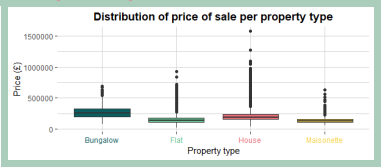
- Property type
- Year of sale

The total floor area will be used after a log transformation

Exploratory analysis



As the size increases so does the selling price. However the relation between them depends on the type of property.



The distribution varies according to property types. Thus this variable is included as a covariate in our model.

The average price of sale has increased over the years. This could be due to the housing market. That is why the variable year is included in the model.

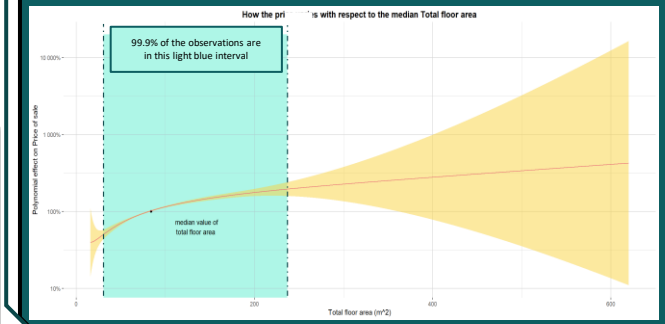
Modelling method

The final model is a GAM model, which can capture the quartic relationships between the total floor area and the price.

A Gamma distribution and a log link function are used.

Effect of the total floor area on the price

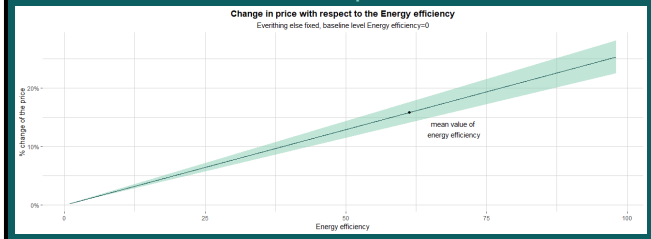
From the figure below is possible to see that there is a non-linear increase in the price as the size of the property increases.



In the graph the red line is the regression line. The yellow shaded area is the confidence interval for the model. The light blue shaded area is where the 99% of the obersvations are located. This is why the confidence interval is so large for properties with bigger size. The black dot is the median value of the total floor area, which results to be the reference for the change of price.

Do people pay more for energy efficient homes?

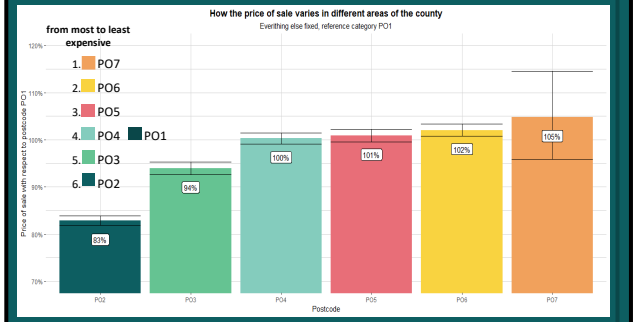
The figure below shows that the price increases linearly with the energy efficiency of the house. For a unit increase in energy efficiency there is a 0.25% increase in the sale price.



The intercept is zero because the baseline for the energy efficiency is the level 0. The black dot in the graph is the mean value for the energy efficiency (61.22). A property with this energy efficiency level costs on average 16% more than a property with energy efficiency equal to zero. Holding all other variables fixed, a property with the highest efficiency costs 25% more than one with the lowest level of efficiency.

Price differences between areas of the county

There are areas where, holding all other variables fixed, the price is higher. In particular the most expensive zone is the PO7 postcode zone, even if with wider confidence bands. The cheapest zone, on the other hand, is the PO2 postcode zone.



It is interesting to note that although the PO7 postcode area is the most expensive (5% more expensive than the reference category PO1), its confidence intervals (black bands) are so wide that not much reliance can be placed on this statement. This is due to the very few observations recorded in the PO7 postcode.

Conclusions

At the end of this analysis it's possible to say that:

- The larger the size, the higher the price
- People pay for more energy efficient houses
- Some areas are more expensive than others

It should be considered that the model's diagnosis presents a slight deviation from the normality assumption, demonstrated by the qq-plot showing a lighter than expected right tail. This is why it is necessary to use this model for forecasting with caution.

Data sources: The dataset used combines information from the Land Registry Price Paid Data and information about property size and energy efficiency from the Energy Performance Certificates (EPCs)