

HOUSING PRICES IN CENTRAL BEDFORDSHIRE

Ewan Halcro

University of Glasgow, School of Mathematics and Statistics



Central Bedfordshire is a region in the south of England home to almost 300 000 people. This analysis used a residential property price dataset from 2021, which comprises of over 30 000 properties within the region. The dataset contains a wide range of variables about the properties themselves, including postcode, year of purchase, and property type. It also contains data on Energy Performance Certificates (EPCs). This analysis aims to investigate the effect, if any, on house price from three features: **property size**, **energy efficiency**, and the **geographical area**.



Methodology

The analysis used a negative binomial generalised linear model with property price as the response variable and 7 explanatory variables. Exploratory analysis was carried out to investigate the structure of the data and the relationships between relevant variables. Log transformations were applied to some variables and polynomial terms were also utilised to address remaining patterns in the data.

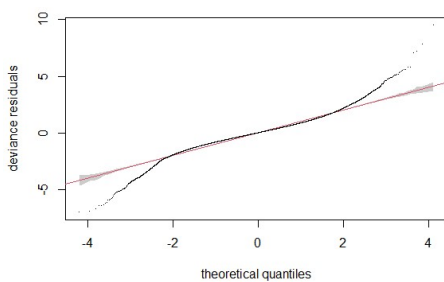


Figure 1: Q-Q plot

The following variables were chosen for the model: **total floor area**; **number of rooms**; both **current and potential energy efficiency**; **town/city**; **year of sale**; and **type of property**.

Due to the house price variable being an integer, it can be treated as count data which led to the choice of the Negative-Binomial model as it is a flexible model in this case. This model yielded an acceptable fit than a linear model as shown in the Q-Q plot in Figure 1. Whilst the fit does not appear to be perfect with the deviating tails, this could potentially be explained by overdispersion in the data.

There also appeared to still be some pattern visible when the deviance residuals were plotted. However, the AIC compared well to previous models tested and around 80% of the deviance was explained by the model.

Results

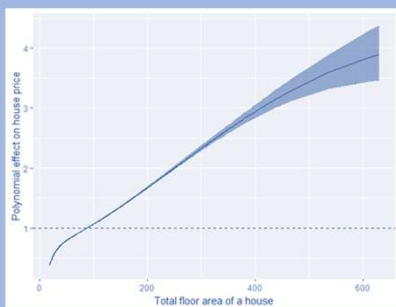


Figure 2: property size

Figure 2 shows the effect of the floor area of a property on its price. It is clear that there is a positive non-linear relationship. As total floor area increases, so does the property price. The wider ribbon at the higher values indicate there is greater uncertainty, perhaps because there are fewer observations at these values.

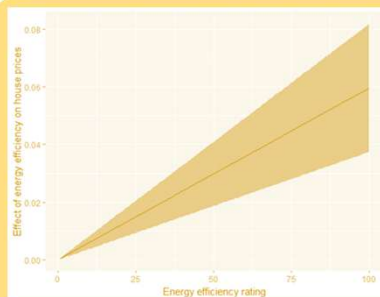


Figure 3: energy efficiency

Similarly, Figure 3 shows the effect of energy efficiency on a property's price. The price appears to increase with energy efficiency. There is again a wider ribbon at the higher values, perhaps for similar reasons as in Figure 2.

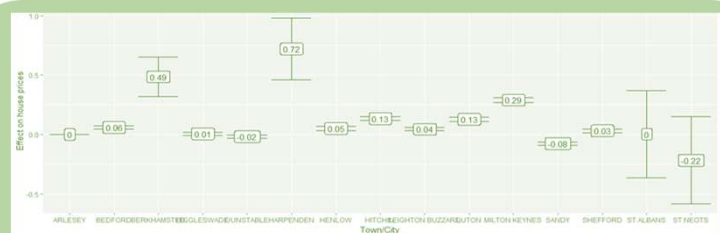


Figure 4: geographical area

Figure 4 shows the plots for house prices in different areas within Central Bedfordshire. It appears that Harpenden is the most expensive area and St Neots is the most affordable. However, there is lots of variability in these results.