Mapping The Price-Size Gradient: Implications for Automated Valuation Models in London's Housing Market

Wade Rogers[1], Faculty Mentor: Dr. Jon Reades[2]

[1]Laidlaw Scholars, Massachusetts Institute of Technology; [2]Center For Advanced Spatial Analysis, University College London

INTRODUCTION

- **Automated Valuation Models (AMVs)** mathematically estimate property values at scale, making them a promising avenue for fairer tax reform; in fact, the Local Government Finance (Wales) Act 2024 has already deployed AVMs to revaluate 1.5 million homes for Welsh Council Tax band adjustments.
- Internal floor area is a significant factor in home valuation, yet existing aggregate-level analysis of floor area's relationship to price disregards spatial variations in this relationship. AVMs assuming a single global risk systematically misvaluing properties, undermining the fairness they're meant to bring.
- This study sampled **25,000 London homes sold for full market value in 2024** to determine the degree of spatial heterogeneity in floor area-price paid per square meter (ppsqm) relationships by iteratively adjusting a geographically weighted regression model with properties' structural and locational attributes that AVMs can realistically use.

METHODS

- The dataset, provided by University College London's Center for Advanced Spatial Analysis (CASA), linked Energy Performance Certificate (EPC) data with Price Paid Data (PPD) along the address field. These datasets provide information about structural attributes (namely floor area) and price, respectively. Both floor and area and ppsqm were log-transformed to address skew and non-linearity, giving each coefficient an elasticity interpretation. An n=25,000 sample (of 87,267 valid records) was used.
- **Predictors were added in three stages:** floor area alone, plus structural predictors (flat status, age, attachment style, etc.), plus locational predictors (distance to London's center, transit access).
- Comparing global Ordinary Least Squares (OLS) regression against Geographically Weighted Regression (GWR), which estimates a distinct coefficient at every location according to the equation:

$$\ln(\text{ppsqm})_i = \beta_0(u_i,v_i) + \beta_1(u_i,v_i)\cdot\ln(\text{floor\_area})_i + \Sigma\beta_k(u_i,v_i)\cdot X_{ki} + \varepsilon_i$$

- OLS residuals showed significant spatial clustering (Moran's I: 0.576–0.409 across models, p<0.001), motivating this spatially-varying approach. Bandwidth was selected via AICc-minimizing search for each model.

Figure 1: OLS Residual Map

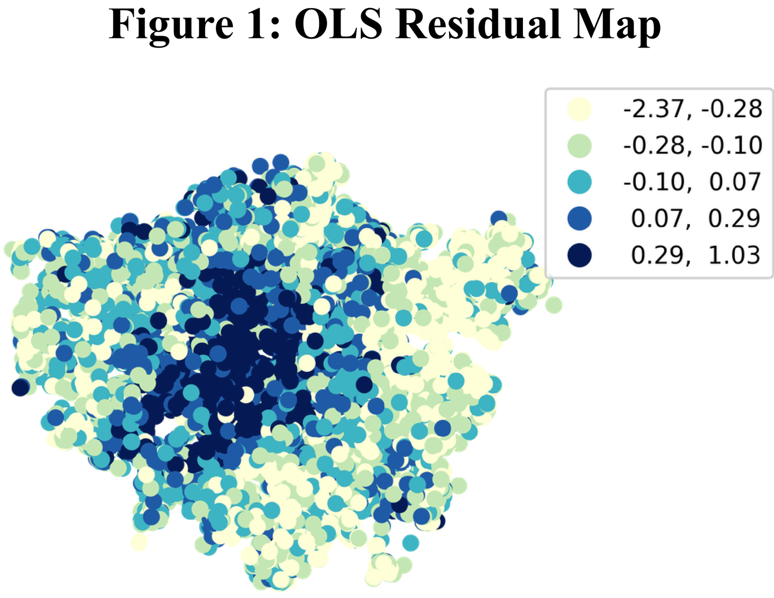

Figure 1: Banded OLS residual (log ppsqm) map for the base model; systematic clustering can be observed.

Figure 2: Moran's Scatterplot

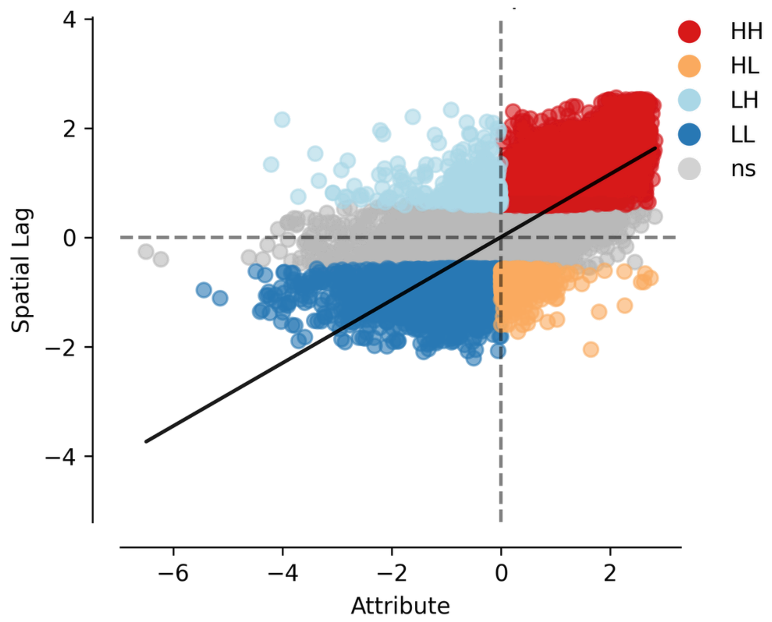

Figure 2: Positive slope indicates spatial autocorrelation (Moran's I = 0.576, p<0.001), motivating a spatially-varying model.

RESULTS

- The relationship between floor area and price per square metre was modelled in three stages, adding structural and locational predictors sequentially.

Figure 3: Base Model Coefficient Map

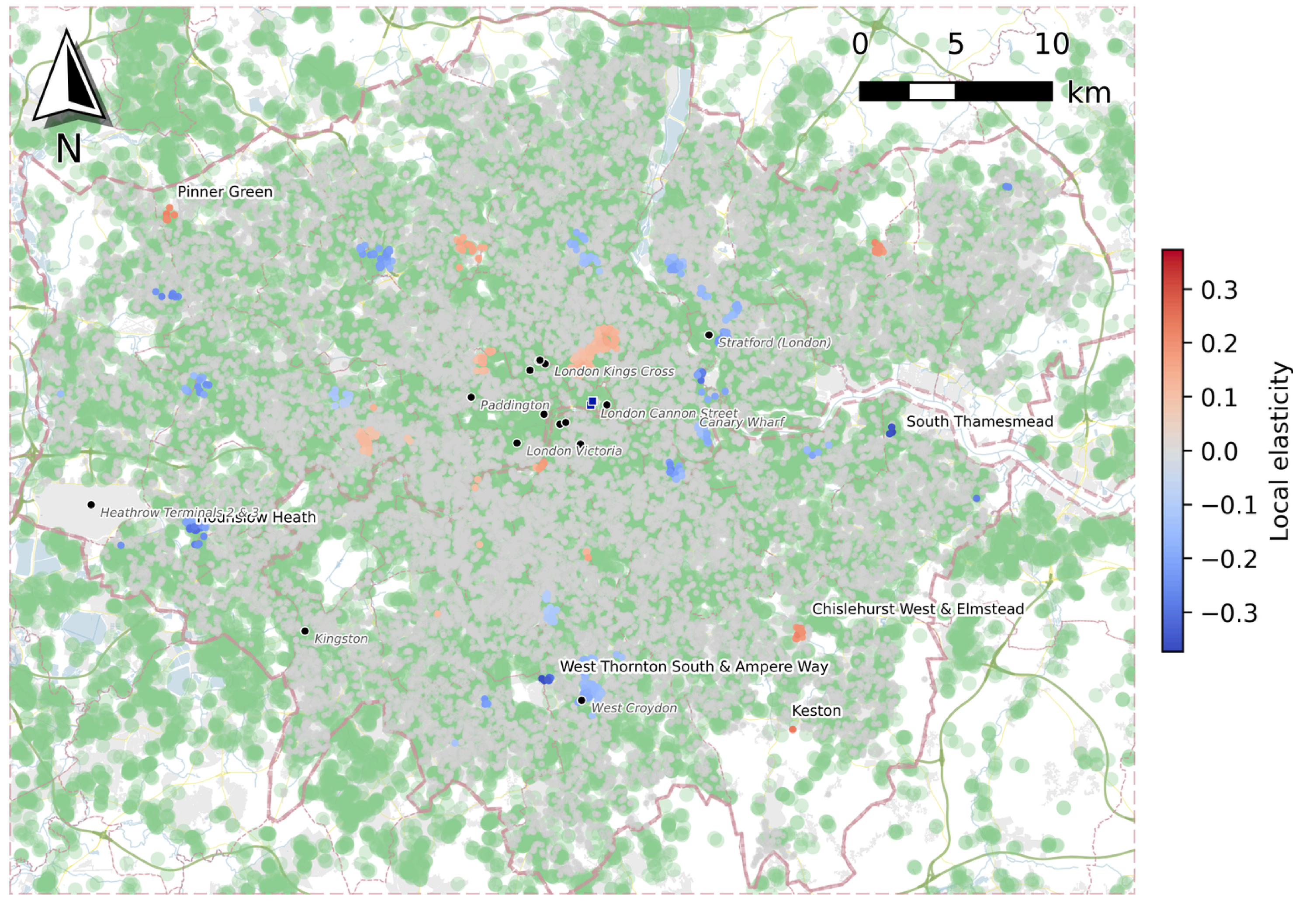

Figure 3: Local GWR floor area coefficient, base model. Grey points not statistically significant at 95% confidence.

Figure 4: Structual Model Coefficient Map

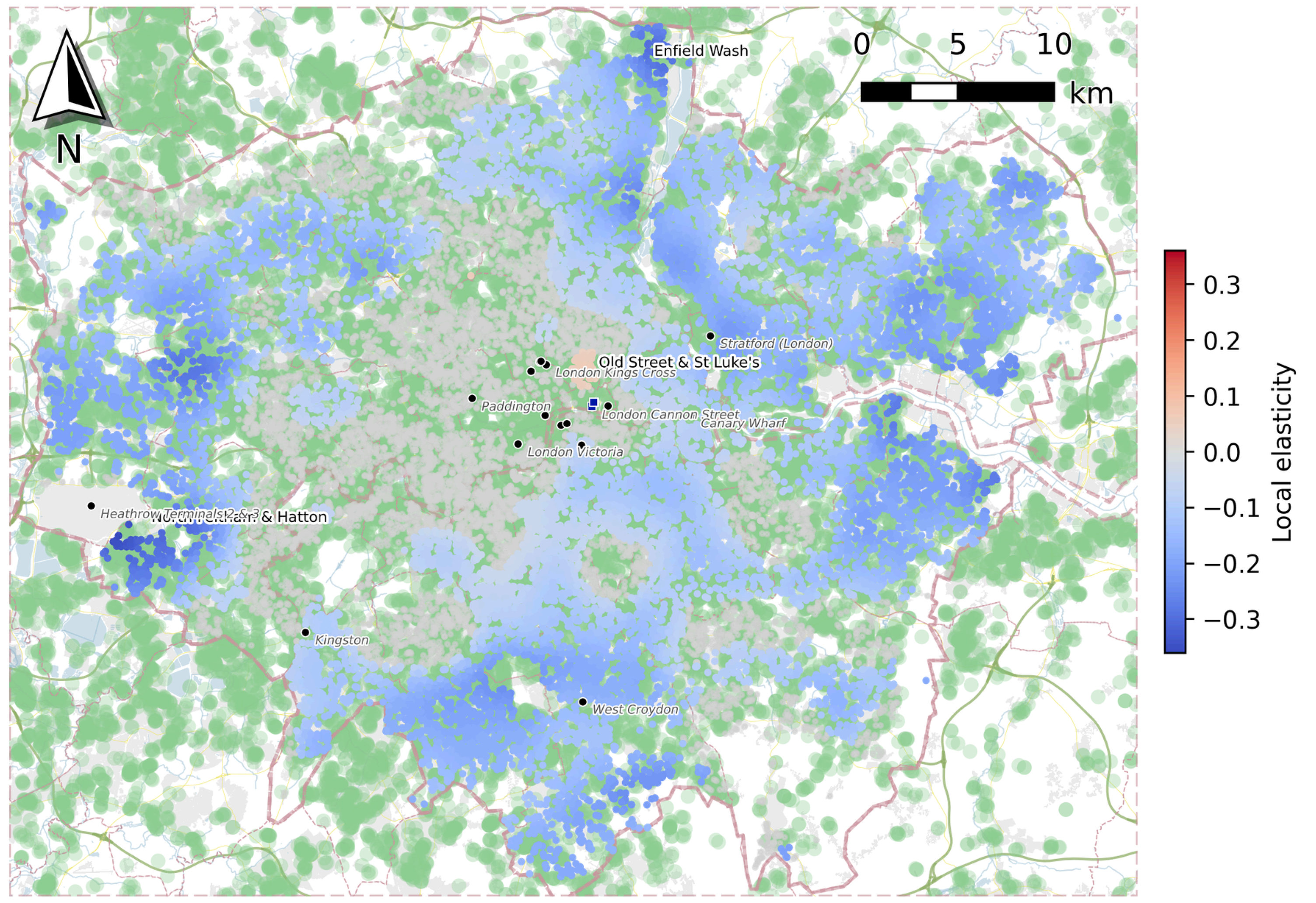

Figure 4: Local GWR floor area coefficient, structural. Grey points not statistically significant at 95% confidence.

- Adding structural predictors substantially altered the pattern; coefficients became significant across much of Outer London, while most central positive regions disappeared, consistent with flat status explaining much of the initial relationship (r = -0.584 with floor area).
- Adding locational predictors changed little further; adjusted R^2 fell from 0.634 to 0.629 and AICc rose, which actually suggested a slight decline in model fit.

Table 1: Model diagnostic across base, structural, and final models

Model	Base	Structural	Final
Predictors	1	6	9
Bandwidth	402m	1050m	1300m
OLS Adj. R^2	0.001	0.032	**0.280**
GWR R^2	0.626	**0.656**	0.649
GWR Adj. R^2	0.584	**0.634**	0.629
AICc	1060.866	**-3428.070**	-3260.873
Moran's I (OLS resid.)	0.576	0.548	**0.409**

Table 1: Comparisons of OLS and GWR diagnostics for each model. Values that support a model as the strongest are shaded blue.

Table 2: Predictor coefficients in OLS and GWR

Predictor	OLS Coefficient	GWR Range (Min-Max)
Floor Area	-0.036	-0.375 to 0.072
Nearby Bus Transport	0.206	-0.615 to 0.271
Nearby Rail Transport	0.257	-0.404 to 0.269
Current Energy Efficiency	0.077	-0.299 to 0.121
Distance from Centre	-0.506	-1.480 to 1.272
Construction Year	-0.085	-0.215 to 0.557
Attachment	0.063	-0.188 to 0.099
Flat Status	0.103	-1.460 to 0.589
New Build Status	0.188	-1.457 to 0.755

Table 2: OLS coefficients versus GWR coefficient ranges by predictor, final model. GWR range reflects all addresses; local significance varies by location.

DISCUSSION

- Each model confirmed that the **relationship between floor area and ppsqm is not uniform** throughout London. However, AMVs commonly treat it as such, potentially compromising the fairness of policy they inform.
- AICc fell by over 4000 points once structural predictors were introduced (a change of 10 is the conventional threshold for a decisively improved model), while locational predictors changed little about the model, suggesting structural factors are the primary driver of the high spatial variation observed.

WHAT'S NEXT

- **MGWR:** Employing Multiscale Geographic Weighted Regression, which allows each predictor to operate at a unique bandwidth, can potentially improve these models' accuracy.
- **Extension and Open Publication:** This work, built on Python, PySAL, and Quarto, forms part of CASA's larger 'Atlas of Small Homes' initiative, which aims to extend these methodologies to other core cities in the UK. The transparency of open code and data is salient in public-facing applications like tax reform where those affected must understand how outcomes are determined.

ACKNOWLEDGEMENTS

Thank you to the Laidlaw Foundation for the funding that made this work possible.
Thank you to Dr. Jon Reades and those from UCL's CASA who made this work fruitful, meaningful, and fun.

Full Paper and References