Chapter VII: Bus Transport Accessibility and Rental Prices in Greater Manchester

7.1 Problem statement

This section investigates the relationship between bus accessibility and rental prices in Greater Manchester using hedonic price modelling. The focus is on understanding whether high accessibility, characterised by the presence of unique bus lines and routes within 200m and 400m proximity of rental listings, is positively associated with rental prices. The second objective is to investigate whether the association between bus accessibility and rental prices is stronger in suburban areas compared to central areas of Greater Manchester.

7.2 Data source and preprocessing

The primary dataset comprises 5,703 unique rental listings collected from PropertyData (2024) across all districts of Greater Manchester. This dataset provides information on rental prices (asking price) and property attributes, such as the number of bedrooms, property type, and geographical coordinates of the rental listings. To assess the relationship between bus transportation accessibility and rental prices, several key variables were derived using the bus timetable data and geospatial matching with GeoPandas, which include buffers of 200 meters and 400 meters around each property to calculate the number of unique bus lines and routes within these distances.

As per the methodology followed in similar hedonic modeling by Huang et al. (2017), identifying spatial autocorrelation is crucial to assess the validity of certain models and incorporate models that allow for spatial dependencies. To investigate this, Moran's I which accounts for the gloabal spatial autocorrelation and (Local Sparial Autocorrelation) LISA introduced in the section 6.2.2 were calculated. The calculated value for Moran's statistic on log-transformed rents is -0.00041, with a negative value indicating a weak tendency towards dispersion. Although the almost zero Moran's I value does not strongly confirm dispersion, a very small p-value signifies statistical significance, meaning the data does not follow a purely random spatial pattern either. The LISA statistic revealed the presence of high-high (HH), low-low (LL), high-low (HL), and low-high (LH) clusters, indicating local variations in rental prices that are not captured by the global Moran's I. The absence of significant global spatial autocorrelation, coupled with the presence of local clusters identified by LISA, underscored the importance of considering local features, such as distances to amenities in the analysis. As a consequence, using the Google Places API, distances to the nearest key amenities such as hospitals, parks, train stations, and schools were calculated employing the haversine distances formula 0.1.1. Rental listings were matched with MSOAs through their associated bus stops, which were linked to LSOAs and then to the Index of Multiple Deprivation (IMD) for each MSOA.

7.3 Models

Similar to Huang et al. (2017), this study employs spatial analysis and a traditional Hedonic Price Model (HPM) approach, a type of multivariable regression analysis that uses a market indicator—the asking rent price—as the dependent variable, while various quantifiable characteristics of the property and its neighborhood serve as independent variables. The hedonic price function used for rental prices P_i is represented as follows:

$$\ln(P_i) = \beta_0 + \sum_{k=1}^{m} \alpha_k \text{Property}_{ik} + \sum_{j=1}^{n} \beta_j \text{Neighborhood}_{ij} + \sum_{l=1}^{p} \gamma_l \text{Transport}_{il} + \epsilon_i \quad \text{(VII.1)}$$

In equation VII.1, P_i denotes the natural logarithm of the asking price of rental property i, β_0 is the intercept, α_k represents the coefficients for property attributes, which include *Bedrooms* and *Property Type*. β_j represents neighborhood attributes like the *Index of Multiple Deprivation* and nearest distances to amenities like *Distance to Hospital*, *Distance to Park*, and *Distance to Supermarket* among others, and γ_l represents transportation accessibility attributes, including *Unique Lines (200m)*, *Unique Routes (200m)*, *Unique Lines 400m*, *Unique Routes 400m*, and *Distance to Nearest Train Station*. These attributes are shown in the feature column of Table 7.1. Finally, ϵ_i is the irreducible error term.

To determine the functional form that most explicitly explains the relationship between rental prices and explanatory variables, both semi-log and double-log models were employed in the hedonic regression models. Variance inflation factors (VIFs) were calculated for each predictor to address potential issues with multicollinearity. Additionally, the Breusch-Pagan test was conducted to check for heteroskedasticity and the Shapiro-Wilk test for normality of residuals was performed to assess the distribution of residuals once the hedonic model was fitted. In accordance with the local clusters determined in LISA, Geographically Weighted Regression (GWR) was employed to account for spatial heterogeneity in hedonic pricing models, as highlighted by Huang et al. (2017). The GWR model allows the coefficients to vary with geographical location, capturing the spatial variability in the data. The GWR model is expressed as:

$$\ln(P_i) = \beta_0(\phi, \lambda) + \sum_{k=1}^{m} \alpha_k(\phi, \lambda) \text{Property}_{ik} + \sum_{j=1}^{n} \beta_j(\phi, \lambda) \text{Neighborhood}_{ij} + \sum_{l=1}^{p} \gamma_l(\phi, \lambda) \text{Transport}_{il} + \epsilon_i \quad \text{(IX.3)}$$

In IX.3, the coefficients $\beta_0(\phi, \lambda)$, $\alpha_k(\phi, \lambda)$, $\beta_j(\phi, \lambda)$, and $\gamma_l(\phi, \lambda)$ vary with geographical coordinates ϕ and λ (latitude and longitude). Of note, the selection of the optimal bandwidth in GWR is crucial, as it determines the extent of spatial variation captured by the model. A range of bandwidths from 2500 to 4000, increasing in increments of 100, was tested. The optimal bandwidth was determined by minimising the corrected Akaike Information Criterion (AICc), balancing model complexity and fit. To address non-linear relationships between independent variables and property prices, machine learning algorithms such as Random Forest 0.1.3 and Extreme Gradient Boosting 0.1.4 were employed. These models were optimised using Bayesian Optimisation 0.1.6, a method that efficiently explores the hyperparameter space to enhance model performance as demonstrated in prior studies, including Liu et al. (2024).

7.4 Results

The semi-log model was chosen over the double-log model due to its higher R^2 and lower Mean Squared Error (MSE), indicating better model performance. Variance Inflation Factors (VIFs), as depicted in Table 7.1, are all less than 6, indicating that multicollinearity does not pose a significant problem.

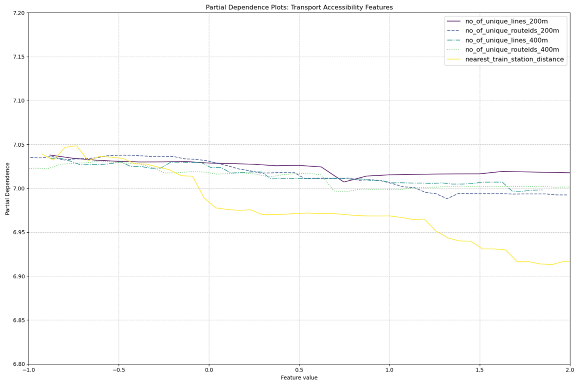
The Breusch-Pagan test for heteroskedasticity (HC) revealed significant HC, with a Lagrange multiplier statistic of 348 and a p-value of 2.6451e-62, strongly rejecting the null hypothesis of homoscedasticity. Consequently, HC-consistent standard errors (HC3) were employed to adjust the coefficient estimates. The Shapiro-Wilk test for normality of residuals yielded a test statistic of 0.86 and a p-value of 2.0085e-30, strongly rejecting the null hypothesis of normality. Given the non-normality of residuals, the OLS estimates in Table 7.1, though not robust, provide insight into the associations between the hedonic variables and rent prices.

7.4.1 Property characteristics

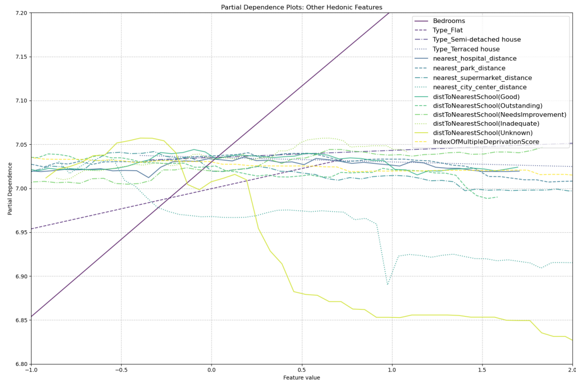
The number of bedrooms shows a significant positive impact across all models, being the top feature in terms of importance for both XGBoost and Random Forest, with a coefficient of 0.25 in HC3. Flats exhibit a significant positive impact in OLS, Q10, and Q50 models, suggesting that the effect of being a flat on rental prices is more pronounced at the lower and median rental price levels but less so at the higher end, or when spatial variability is considered. Semi-detached properties are significant in Q10 and negative in Q50 and Q90, indicating a negative impact on rental prices at median and higher levels despite positive significance at the lower end. Terraced properties, while not significant in OLS and HC3, are significantly positive in Q10 but significantly negative in Q50 and Q90, suggesting a positive impact on lower rental prices and a negative impact on median and higher rental prices.

7.4.2 Neighbourhood characteristics

Distance to the nearest hospital shows a significant negative impact in HC3, median, and upper quantiles, indicating that properties with better healthcare accessibility demand higher rental prices. Distance to the nearest parks has a mixed impact, with negative effects in the upper quantile and GWR models, suggesting spatial variance, although it is not very significant. Distance to supermarkets is significant in lower and median quantiles but not in GWR, suggesting that the effect is not spatially uniform, likely due to the abundance of supermarkets in Greater Manchester. However, it is consistently negative, though not significant. Distance to the nearest schools (including various types, see Table 7.1), is significant in all quantiles but insignificant in GWR. Notably, good schools are significant in all quantiles; schools needing improvement are significant in lower and median quantiles; unknown schools are significant in median and upper quantiles; and outstanding schools are significant in the median quantile, although not strongly. The partial dependence plots in Figure 7.1b clearly demonstrate that distances from amenities are negatively correlated with rental prices, indicating that properties closer to amenities command higher prices.



(a) Transport accessibility features



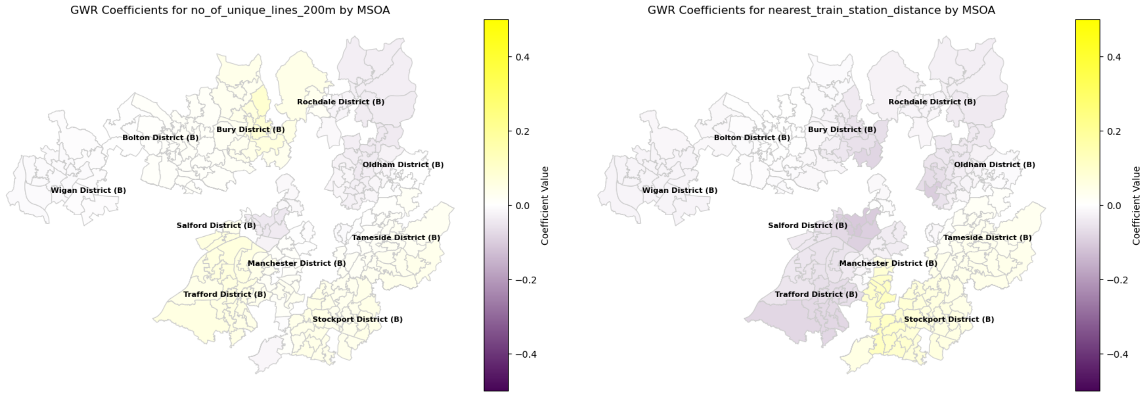
(b) Other hedonic features

Figure 7.1: Partial dependence plots for the feature sets (XGBoost)

The IMD has a consistently negative and significant impact across all quantiles, HC3, and OLS models. However, it shows a larger coefficient (0.456) and is weakly significant and positive in the GWR model, suggesting spatial variability. This indicates that while deprivation generally correlates with lower rental prices, the spatial distribution of deprived areas in Greater Manchester, especially in central parts, introduces more complexity. This aligns with findings in section 5.4.2.

7.4.3 Transport accessibility

The number of unique bus lines within 200 meters proximity of rental listings shows a negative and significant impact in HC3 and all three quantile models, while it is only marginally significant in the GWR model. This suggests that closer proximity to multiple bus lines generally increases rental prices, particularly at lower and median rent price levels. For unique bus routes within 200 meters, the impact is slightly significant in the median quantile and negative otherwise but insignificant in the GWR model, suggesting no spatial variability. The number of unique bus lines within 400 meters is slightly significant in the upper quantile and highly significant in GWR, showing spatial variability. Conversely, the number of unique routes shows a positive effect significant in HC3 and all quantiles but not in GWR, indicating no spatial variability. This could be due to uniform impacts as proximity to a diverse set of bus routes might lead to more bus routes in densely populated or central parts of Greater Manchester. Distance to the nearest train station is significantly negative across all models except the GWR model, suggesting that closer proximity to train stations is a strong predictor of higher rental prices due to the connectivity they offer. The partial dependence plots showcased in Figure 7.1a for XGBoost show the marginal effects of these transport features on rental prices, confirming that proximity to bus lines and train stations influences rental prices.



(a) Number of unique lines 200m

(b) Distance to city centre

Figure 7.2: Spatial coefficients from GWR analysis

Table 7.1: Results of hedonic pricing models estimated by OLS, heteroskedasticity-adjusted OLS, quantile regression, geographically weighted regression (GWR), random forest (RF), and XGBoost

Feature	OLS Coef.	VIF	HC Coef.	Q10 Coef.	Q50 Coef.	Q90 Coef.	GWR Coef.	RF Imp.	XGB Imp.
Intercept	7.0340	-	7.0327	6.8258	7.0420	7.2460	0.0552***	-	-
Bedrooms	0.2366***	1.5008	0.2352***	0.2125***	0.2266***	0.2263***	0.2205***	0.3404	0.4486
Unique Lines (200m)	-0.0062	3.3632	-0.0077	-0.0040	0.0007	-0.0144**	0.0080**	0.0128	0.0175
Unique Route IDs (200m)	0.0041	5.0847	0.0066	-0.0014	-0.0084*	0.0088	-0.0060	0.0155	0.0168
Unique Lines (400m)	-0.0067	4.3968	-0.0053	0.0120*	-0.0056	-0.0110**	-0.0129***	0.0204	0.0209
Unique Route IDs (400m)	0.0256***	5.9358	0.0238***	0.0153**	0.0383***	0.0215***	0.0069	0.0280	0.0261
Distance to Hospital	-0.0138***	1.2845	-0.0147***	-0.0056	-0.0077**	-0.0154***	0.0194	0.0199	0.0133
Distance to Park	0.0056*	1.1958	0.0075**	-0.0002	0.0111***	0.0071***	-0.0065	0.0267	0.0092
Distance to Supermarket	-0.0103***	1.7459	-0.0083	-0.0188***	-0.0127***	-0.0005	-0.0057	0.0946	0.0133
Distance to City Center	-0.0352***	3.4438	-0.0400***	-0.0728***	-0.0520***	-0.0056	-0.0067	0.0669	0.0369
Distance to Train Station	-0.0538***	2.2578	-0.0538***	-0.0749***	-0.0520***	-0.0444***	-0.0436	0.0419	0.0237
Distance to School (Good)	0.0159***	1.3376	0.0165***	0.0119***	0.0185***	0.0196***	-0.0260	0.0294	0.0200
Distance to School (Outstanding)	-0.0056	1.2466	-0.0057	-0.0089*	-0.0059**	0.0005	-0.0037	0.0287	0.0209
Distance to School (Needs Improvement)	0.0234***	1.5253	0.0219***	0.0312***	0.0229***	-0.0084**	-0.0062	0.0472	0.0165
Distance to School (Inadequate)	-0.0356***	3.7270	-0.0363***	-0.0585***	-0.0511***	-0.0078	-0.0124	0.042	0.0242
Distance to School (Unknown)	-0.0470***	4.3503	-0.0457***	-0.0176*	-0.0389***	-0.0926***	0.0414	0.1481	0.0817
Index of Multiple Deprivation	-0.0317***	1.1821	-0.0301***	-0.0385***	-0.0347***	-0.0101***	0.4562*	0.0153	0.0155
Property Type: Flat	0.0711***	2.3934	0.0702***	0.1346***	0.0241***	-0.0045	-0.0131	0.029	0.1211
Property Type: Semi-detached	0.0171***	1.8495	0.0188***	0.0802***	-0.0056	-0.0281***	0.1019***	0.0054	0.0368
Property Type: Terraced	-0.0037	2.0542	-0.0029	0.0604***	-0.0246***	-0.0464***	0.0525***	0.0065	0.0371
R^2	**0.6764**	-	**0.6764**	**0.4155**	**0.4383**	**0.4546**	**0.7050**	**0.8678**	**0.8888**

Notes:
*** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$
The R^2 values for the models are provided in the bottom row.

7.5 Conclusion

The results from the GWR and XGBoost models, which account for spatial variability and non-linearity respectively, highlight that the presence of unique bus lines and routes within 200 meters and 400 meters of rental listings are significantly associated with rental prices. The analysis reveals that routes in the east of Greater Manchester are associated with higher rental prices, while the number of unique lines shows a negative impact in central and eastern areas as showcased in Figure 7.2 suggesting that central areas already have established connectivity, making rental prices less impacted by additional lines. Conversely, in less central areas, the number of lines has a more pronounced positive association with rental prices, indicating a need for developing bus transit corridors to enhance connectivity enhancing route diversification. The findings support the hypothesis of this study that bus transport accessibility is associated with rental prices, with significant spatial variability across Greater Manchester.