KØBENHAVNS UNIVERSITET

Prn917 & Xkh771

No of characters: 28,500 (11.88)

Introduction to Social Data Science

How does accessibility to public transit impact property prices?

Contribution

Prn917: 1., 3.3, 3.4, 4., 4.2, 5.1, 6.

Xkh771: 1., 2., 3.1, 3.2, 4.1, 4.3, 5.2, 5.3, 6.

Aug 28, 2024

1 Introduction

This paper sets out to answer the question: How does proximity to public transit via bus and metro affect house prices? The relationship between accessibility to public transit and house prices is of significant interest due to it's implication for urban planning, real estate investment, and public policy. A greater understanding can help both urban planners make more efficient designs for public transit, and it can help investors make more informed decision. Research has been heavily focused on the effect of railroads on property prices (E.g Rodriguez and Targa (2004) and Bowes and Ihlanfeldt (2001)), but research into the impact from an improved network of busses has shown to be scarce (Yang et al., 2019). This study aims to fill this gap in existing literature, with an econometric analysis of the effect of accessibility of busses and metros on property prices in the Greater Copenhagen Area. We do this from both a local accessibility point of view, that describes how easily a resident can get to a given transit stop, and from a proxy of regional accessibility, that describes how well-connected a given property is to the rest of the transit network. We deploy a standard hedonic model, a quantile regression model, and a spatial model to shed light on the topic from multiple angles.

In our research, we find that homeowners list their property with a premium when closer to the nearest transit stop, and when having multiple stops nearby. This results persists across all our regression models. The effect is strongest using the local accessibility measure. However, we prefer the regional accessibility measure, as it is more realistic in magnitude. When we look across different quantiles of listed prices, the effect is stronger for lower and middle valued properties which are listed at a higher premium when having good accessibility to public transit. This could be caused by the fact that the lower quantiles are the furthest from the city centre, and they are dependent on public transit to commute to work and other activities, whereas the potential residents of the most expensive houses would also have the disposable income not to be dependent on public transit. This is in accordance to research done by DTU Transport Transport (2014), that show that the lowest income groups travel the most by busses followed by the middle income group. The rapport also shows that the highest income group will travel longer distances when going with public transport, thus meaning that they are dependent on the train stations rather than bus stops. To ensure that, our results were not affected by underlying factors such as distance to the city centre, socio-economic characteristics of a specific area, we employed a spatial regression model. Here, we found that there still exists a premium for regional accessibility but of smaller magnitude.

The rest of the paper is structured as follows: In section 2 we will brush over the relevant literature on the topic; in section 3 we will explain what and how we have gathered the data to make the analysis, and the ethical considerations in doing so; in section 4 we go over the Gauss-Markow conditions and perform the analysis on a hedonic model, a quantile model, and a spatial model; in section 5 we discuss our findings and avenues for future research.

2 Literature Review

In the literature highlighting the impact of transit accessibility on property prices, there is a large number of papers researching the benefit of being located near a fixed railroad station. There has been established a relatively consistent, yet inconclusive, relationship between distance to a railroad or bus rapid transit (BRT) station and economic advantage (E.g. Bowes and Ihlanfeldt (2001), Rodriguez and Targa (2004), So et al. (1997)), who show that residents are willing to pay a premium on properties ranging between 2.4%-3.5% based on proximity to nearest railroad station, whereas they are willing to pay around 6.8%-9.3% for proximity to a BRT station. Others has only been able to find a marginal or even insignificant effect (E.g Landis et al. (1995), Gatzlaff and Smith (1993)).

Relatively limited attention has been paid to the effect of having a bus available close by, most studies have not focused specifically on this issue, and of those who have, conflicting results has been drawn. A majority of the studies have shown that better accessibility to bus transit is insufficient to impact property prices (E.g So et al. (1997), Wen et al. (2014)), while some has even found a negative relationship (Cao and Hough (2008)). Cao and Hough (2008) discusses whether this negative effect represents a spurious relationship. They argue that proximity to bus stops and amount of bus stops nearby might work as a proxy for spaciousness of apartments: the closer an apartment is to a central bus stations the smaller it usually is, thus decreasing the price.

All of these studies have been limited to looking at the local accessibility impact of transport. That is the ease of transporting yourself to a place of transport, this could be given in number of metres to nearest station (Yang et al., 2019) chose to look at both the local and the regional accessibility of transport. They also stand out from the rest of the papers by performing a spatial regression, taking into account the externalities of other properties in the same area having higher prices, which has been shown to significantly improve the models that aim to estimate house prices (Osland, 2010). Yang et al. (2019) find a positive effect of accessibility (both local and regional) on property prices in the area. Generally, in research attempting to quantify the relation between characteristics of properties and the selling price differs in not only sign and size of coefficients but even significance. To mitigate this Zietz et al. (2008) implements a quantile regression, as some characteristics might not be priced the same across the distribution of house prices

3 Data collection

3.1 Ethical considerations

As data scientists, we hold significant responsibility to ensure that our practices align with ethical standards. We consulted with all Terms and Conditions of the websites to make sure we aligned with the demands. We made sure that all data presented is publicly available on a non-aggregated level, eliminating the risk of exposing sensitive information. We utilized headers with full contact details, to allow recipients to understand who were requesting from their website. Further, we applied time.sleep(1) to minimize the load we expose the servers to.

3.2 Property data

We use data for listed properties from Boligsiden, which is one of Denmark's largest sites for both secondhand and newly built properties. Much of the research that has been done so far has been using data from sold properties rather than listed properties. This is a limitation in our analysis, as we focus on seller's expectations rather transaction prices.

We gathered data for 4,347 properties in the region "København Omegn" and the municipalities København and Frederiksberg. We choose to filter out properties of the type terraced house, full year plot, cooperative, holiday house, and farm as these are not necessarily priced by the market, but might be subject to the regulatory environment. Our remaining dataset consists of condos, villas, and villa apartments. This leaves us with of 3,760 properties. As shown in 1, the data points are spread throughout the entirety of the geographic analysis area, with the city of Copenhagen being the most densely populated.

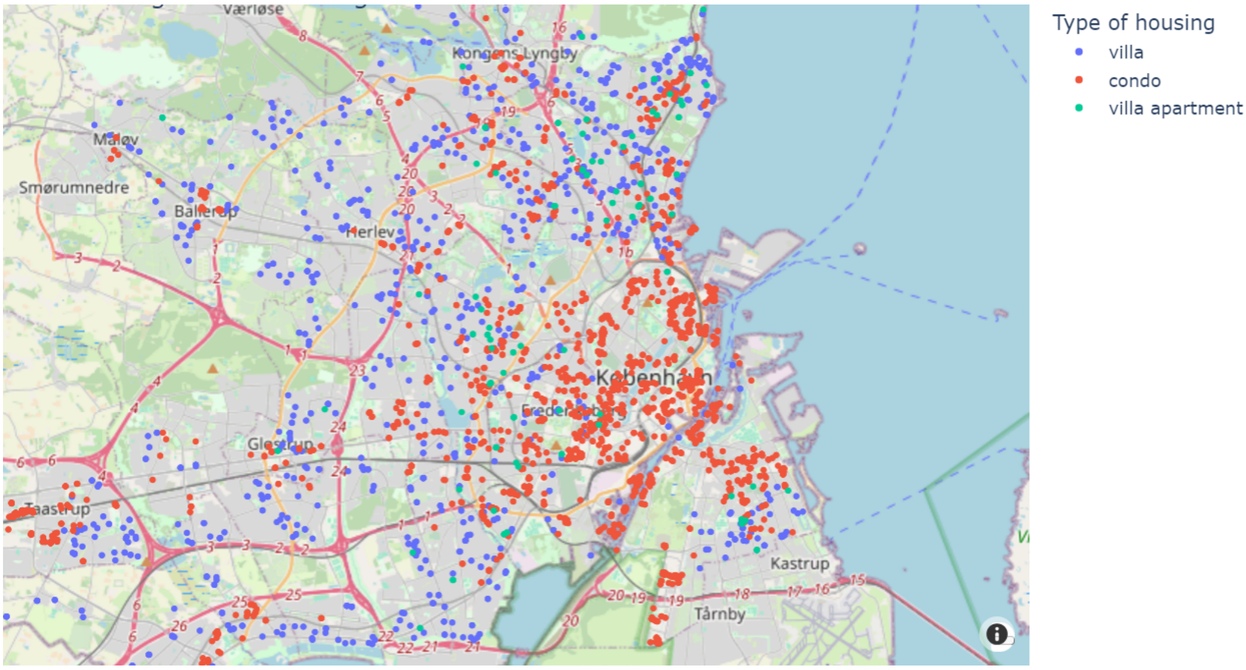


Figure 1: Properties in dataset

We have also chosen to exclude properties that deviates negatively more than two standard deviations from the average price per square meter of the given postal code. This is done to prevent right of reversion (Hjemfaldspligt) to have any significant impact. Aside from these manipulations, we have made no further attempts at excluding other observations from the dataset, as it should be considered a random sample of observations.

We used RegEx to extract the postal code and city of the observation, that are used in the spatial regression. Coordinates are left in the table, as we use them for integration across datasets. We also changed energy label of the house to be of numeric value from 1-5 instead of D-A. The variables left after cleaning can be seen from table 1 in the Appendix.

We have decided to log-transform the dependent variable, listing price of the property, as it has long tails leading to a right-skewed dataset. When log-transforming the target variable, it follows a normal distribution closely. It will also make the interpretation of the parameters easier, as they are now representing the percentage change in the target variable for a one-unit change in the control variables. The distribution of listing prices can be seen from figure 2.

3.3 Transit data

We utilize data on public bus connections from Din Offentlige Transport (DOT) which is a collaboration between DSB, Metro, and Movia and manages public train, metro, and bus transportation in Copenhagen.

To compile a comprehensive list of bus stops, we first extracted data on all bus routes

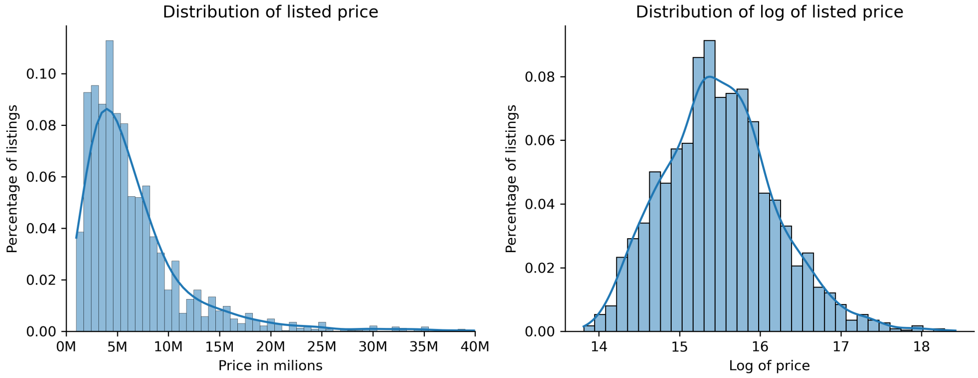


Figure 2: Distribution of listed property prices. Note that the figure on the left has been truncated to include all properties with a listing price below 40 millions.

available on DOT's website by scraping their website, resulting in 424 bus routes. Subsequently, we used the identifiers of these bus routes (e.g., 1A and 5C) to retrieve all associated bus stops, totaling 11,550. The retrieval of associated bus stops is done by entering each bus line's subpage on the DOT. Regarding the comprehensive list of bus stops, it is important to note that the same stop may appear multiple times if it is shared by different bus routes.

To find all metro stops, we scraped www.M.dk, which is the official website of Metro, who is in charge of the Copenhagen metro. Next, we cleaned it, so it aligned with the notation of bus stops previously scraped, such that we were able to integrate it where possible.

For integration of the public transport stop data with property listings, it was necessary to obtain geographic coordinates to calculate the distance between bus stops and properties. To achieve this, we employed Google Maps' geocoding API*, which allowed us to generate 6,977 unique sets of coordinates representing all the transit stops (metro and buses) along the identified routes. Google Maps' geocoding API comes in very handy as it is able to recognize the names of transit stops and return reliable coordinates.
Our method for retrieving transit data delivers a reliable approach as we collect it from the source who is responsible for the maintenance while combining with Google Maps' API.

3.4 Nearest neighbour

To analyze the availability of public transit, we needed to determine the most appropriate measures for our regression analysis. We will work with two metrics based on what has previously been used in the literature: local and regional accessibility, where the first refers to the ease of obtaining access to a transit stop and the latter explaining where a resident can arrive within a specified time. We measure the local accessibility as the distance to the nearest bus stop, and the regional accessibility as the number of stops within a 500-meter radius of the property. For the ease the calculations, we used the straight-line (or "as-the-crow-flies") distance, which is an approximation and likely underestimates the actual travel distance.

Figure 3 illustrates the concepts behind the regional accessibility measure. The red point represents a property listing, while the black points denote bus stops. The white circle (which appears as elliptical) highlights the 500-meter radius around the property, within

*https://developers.google.com/maps/documentation/geocoding/overview which the relevant bus stops are located. The visualization demonstrates how we evaluate the accessibility of public transit options relative to a property.

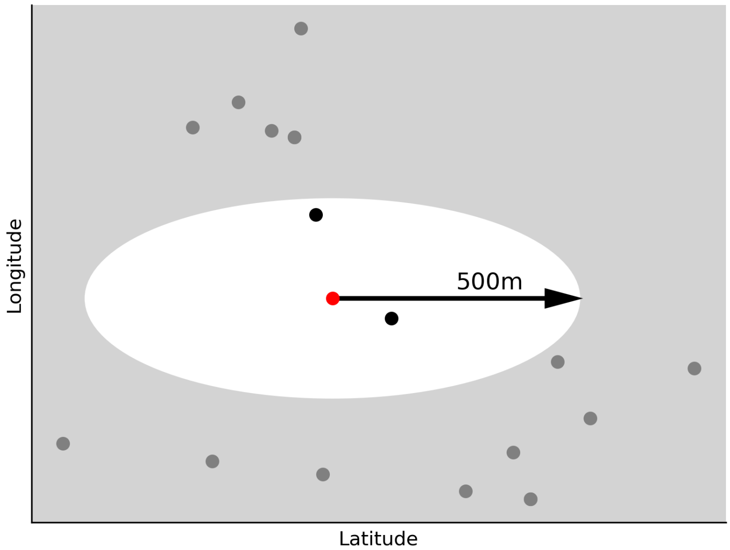


Figure 3: Example of radius measurement

There are various methods available for calculating these metrics. A straightforward approach would be to brute force where one would have to calculate the distance for each property listing to all the bus stops. While conceptually simple, this brute-force method is computationally expensive, requiring millions of distance calculations. Although feasible for our dataset, we propose a more efficient method known as the Ball Tree approach, which significantly reduces the computational load.

The Ball Tree is a space-partitioning data structure designed to organize points in a multi-dimensional space by partitioning them into a hierarchy of nested balls. This structure is particularly advantageous for nearest neighbour searches. We utilized Scikit-learn's implementation of the nearest neighbour search algorithm using a Ball Tree. In this implementation, the Haversine formula is employed to measure the distance between two points on the surface of a sphere. This allowed us to efficiently and precisely compute the distance from any property listing to the nearest bus stop, as well as identify all bus stops within a 500-meter radius.

In figure 4, we show how the local and regional accessibility are distributed for all listings. We see that most of the properties have below 500 metres to the nearest transit stop and only three properties have more than 1 kilometre. Likewise most of the properties have between 2 and 7 stops within 500 metres while 3.5% have none.

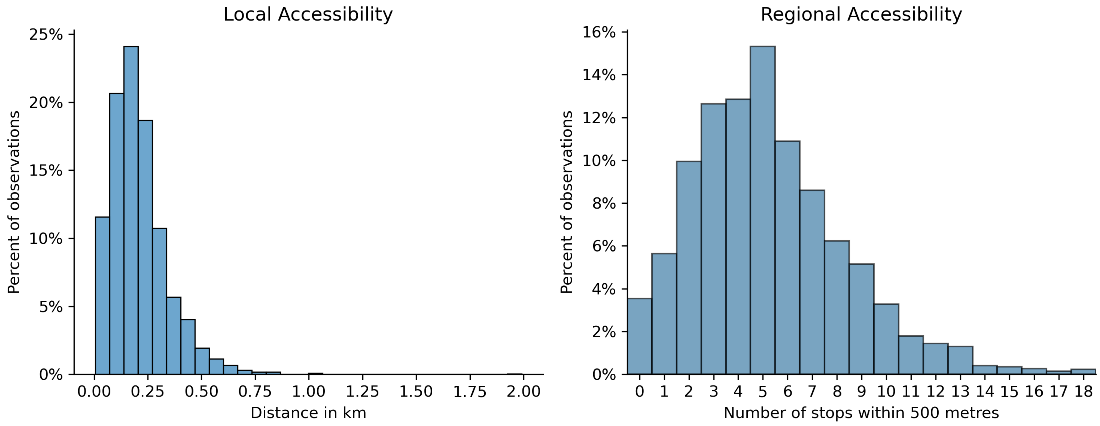


Figure 4: Distribution of local and regional accessibility measures

4 Analysis

The starting point of our analysis is the hedonic model. Next, we examine how the coefficients differ between high-valued and lower valued property listings. Finally, we refine our analysis by accounting for the spatial correlation, recognizing that nearby houses may share similar values due to their proximity to one another.

4.1 Hedonic Model

To be able to make any valuable inference, we need to comply with the five Gauss-Markow conditions, where we would argue that by design the conditions of linearity in parameters and random sampling has been met. The remainder of the conditions are discussed below.

No perfect multicollinearity: We have chosen not to use number of bedrooms nor bathrooms as this has been shown to have a high degree of collinearity with square footage (housingArea) So et al. (1997). To check for any issues with multicollinearity, we created a matrix over the control variables, as can be seen in figure 5. As some of the covariances between the control variables seem rather high, we choose to do a Variance Inflation Factor test (VIF test). The VIF test is a measure of multicollinearity in a model, and it tests how much the variance of a regression coefficient is inflated by multicollinearity. It is calculated as $VIF_i = \frac{1}{1-R_i^2}$ where R_i^2 is the R^2 obtained from regressing the i-th variable on the remaining independent variables. As a general guideline, any VIF coefficient below 5 demands no further investigation, thus our VIF test, which can be found in the appendix in table 2, reassures us that there is no prominent multicollinearity in our model.

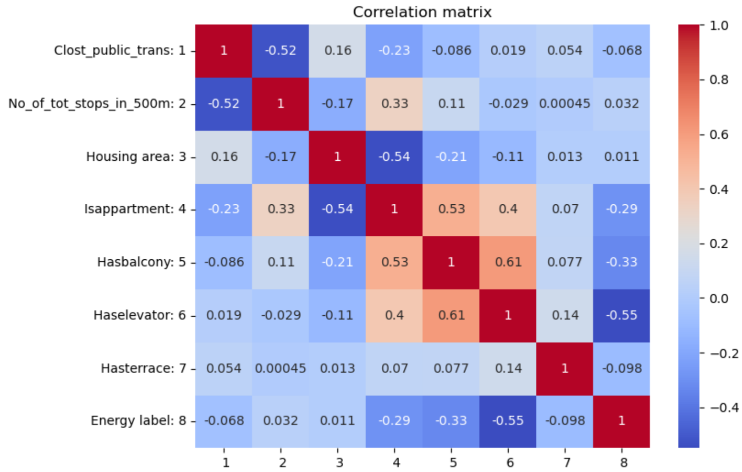


Figure 5: Covariance matrix

Zero conditional mean: When taking the log of listing price of properties, we reduce the variance of the errors. In figure 6, we have shown the distribution of the errors from the hedonic models, which shows that they conform to a normal distribution around zero.

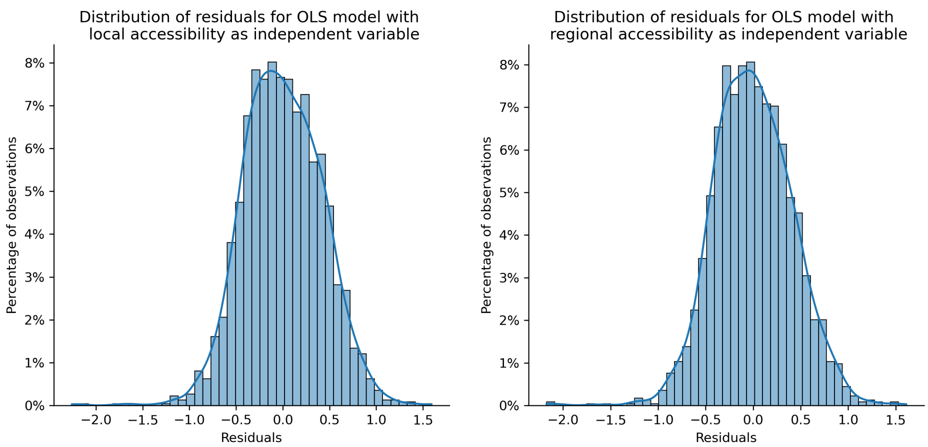


Figure 6: Distribution of errors from hedonic model

Homoskedasticity: To check for heteroskedasticity, we perform a Breusch-pagan test using the statsmodel.stats.api package. In general all of our models suffer from heteroskedasticity on 5% significance level. In order to do inference, we use robust standard-errors when needed.

Analysis

A widely used model in the literature for analyzing property prices has been the hedonic model. This model is particularly effective for decomposing the price of a property into the value of its characteristics. By doing so, the hedonic model allows for testing hypotheses about how specific factors influence property prices which makes it an appropriate choice for this analysis.

We propose the following regression model:

$$\log(p_i) = \beta_1 \text{transit\_acc}_i + \sum_{j=2} \beta_j x_{j,i} + \epsilon_i \tag{1}$$

where $\log(p_i)$ is the log price of property listing i, transit_acc_i is one of the two accessibility measures (see above), $x_{j,i}$ represents a set of control variables listed in Table 1, and ϵ_i is the error term.

The regression is run with heteroskedasticity-robust standard errors and with an intercept term, yielding an R^2 of 0.65, which is smaller than that reported in similar studies such as Yang et al. (2019) and Rodriguez and Targa (2004). All coefficients presented in Table 3 are significant at the 5% level.

Our results from the hedonic model indicates that as the distance to the nearest public transportation increases by 1 kilometre, the price of the property decreases with 32%. Likewise, as the number of transit stops increases by 1, the listed property price tends to be higher by 3%. This is in line with our initial hypothesis, that properties are listed with a premium closely connected to the accessibility of public transit. It is also in line with the findings of Yang et al. (2019), although comparatively our coefficients for the local accessibility measure show a significantly higher premium, than they find. However, the coefficient for the regional accessibility measure is more inline with Yang et al. (2019).

4.2 Quantile Regression

Next, we employ a quantile regression as a robust alternative to OLS regression. Unlike OLS, which assumes a uniform effect across the price spectrum, quantile regression reveals the heterogeneity of effects, showing that proximity to public transportation might be more significant for lower-priced homes, where residents are more dependent on transit.
Additionally, quantile regression is less skewed by extreme values that disproportionately influence the mean estimates which makes it more robust to outliers. The impact of outliers is reduced as it minimizes the absolute residuals at different quantiles. This robustness is particularly crucial in this case as property prices vary widely as seen in figure 2

The model is defined as:

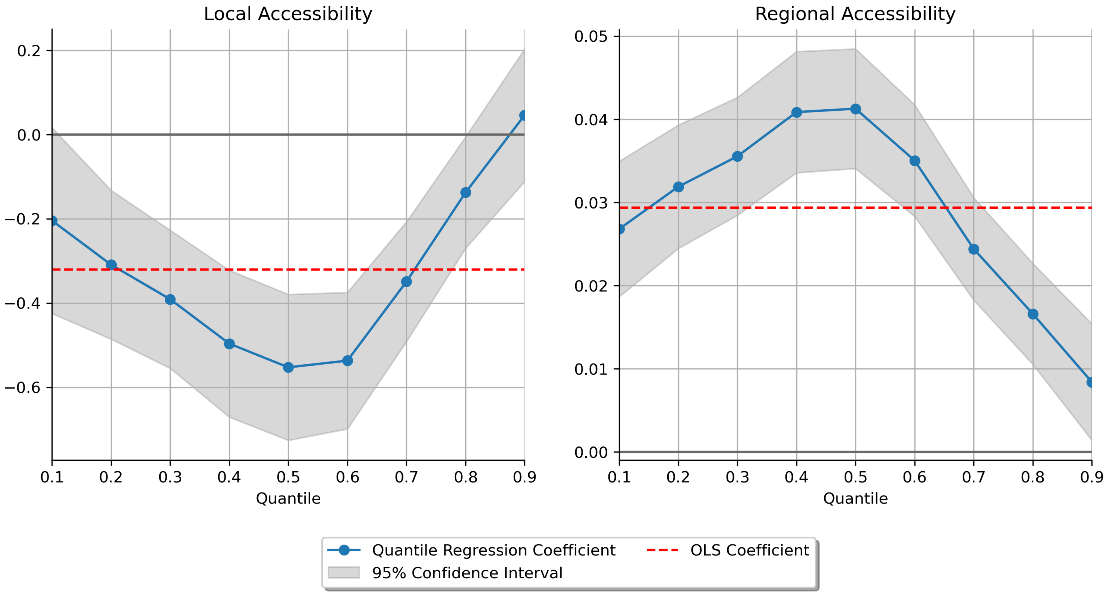
$$Q_\tau(\log p_i) = \beta_1(\tau)\text{transit\_acc}_i + \sum_{j=2}^{k} \beta_j(\tau) x_{j,i}$$

where Q_τ represents the τ-th quantile of the log of the listed price and $\beta_j(\tau)$ measure the effect of the explanatory variable x_j. The coefficients are estimated by solving:

$$\min_{\beta_1(\tau),\ldots,\beta_k(\tau)} \sum \rho_\tau \left(y_i - \sum_{j=1}^{k} x_{j,i}\beta_j(\tau) \right)$$

where $\rho_\tau(\cdot)$ is a stepping function: $\rho_\tau(u) = \tau \max(u, 0) + (1 - \tau)\max(-u, 0)$ where u is the difference between predicted value and actual value.

We employ statsmodels' implementation of quantile regression for 10 quantiles, where we again do the regression for both the local and regional accessibility measure. All estimated coefficients for each quantile are depicted in table 4 and table 5. We note that all estimated parameters for the control variables are qualitatively identical to the OLS regression. However, the level and significance differ across quantiles. We will not delve into the control variables but instead focus on transportation accessibility measures.



..

Figure 7: Estimated coefficients for local and regional accessibility for different quantiles

In Figure 7, we plot the estimated coefficients of our accessibility measures across quantiles. Firstly, it can be seen that the qualitative effects are somewhat close to the OLS estimates for both measures. Secondly, we note that the impact of accessibility increases up until the 4th-5th quantiles before diminishing. The coefficients for distance to the closest transit stop range from −55% to 0%, while they for the number of transit stops within 500 metres ranges from 0% to 4%. The magnitude of the local accessibility measure is unrealistically high, why we prefer to look at the regional measures.

Overall, the quantile regression supports the conclusion from the OLS that proximity to transit stops increase the price of a property listing. For low-valued homes, having an extra transit stop within 500 metres increases the listed price with roughly 3.5%, while the effect is small (to non-existing) for high-valued homes. This conclusion aligns with what we would have expected.

4.3 Spatial regression

Since property prices are influenced not only by the characteristics of the property itself but also by the values of nearby properties and local amenities such as schools and parks, we will now account for spatial dependency. Spatial dependence is "the co-variation of variables within a geospace" (Yang et al., 2019).

Other factors could be geographical like proximity to city center and environmental quality can impact property prices (Se Can and Megbolugbe, 1997). These geographical factors might be correlated with our local accessibility measure, which would explain the unrealistically high coefficients. To explicitly capture the spatial dependence, we created a variable (postalcode_avg_price) that captures the average price per square metre of properties within the same postal code. In figure 8, log of average price per square metres of property has been mapped in different postal codes. We include it in the set of control variables $x_{i,j}$ in equation 1 and then run the regression as previously with robust standard errors, but now only for the regional accessibility measure.

All estimated coefficients are depicted in table 3 in the appendix. As expected the new coefficient is significant and positive meaning that when the price of properties in the same postal code is higher, this increase the listed price of the properties. The rest of the control variables are qualitatively the same with some smaller quantitatively differences. Lastly, we note that the spatial regression explains more of the variance compared to the simple OLS regression.

In regards to the accessibility, the estimated coefficient for number of transit stops within 500 metres is still positive meaning that homeowners list their property with a premium when their home is more connected by public transport. This is in conformity with our original hypothesis. The coefficient is quantitatively smaller than our coefficient from the OLS regressions, which could be a sign that there might be a more dense network of public transport in areas that are closer to the city centre, and previous estimations of this parameter has captured some of this effect.

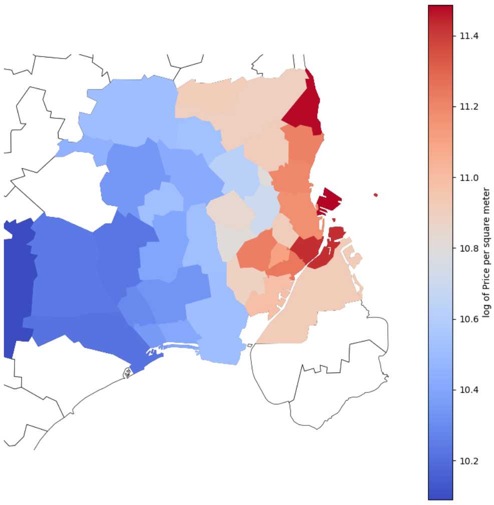


Figure 8: Log of listed price pr. square metre across postal codes

5 Discussion

5.1 Ranking of transit stops

One potential limitation of our analysis is we create our local and regional accessibility measures based solely on the presence of any transit stop, without differentiating between the types of transit or the quality of the stops. This means that our analysis does not distinguish between bus stops and metro stations, nor does it consider whether a stop is well-serviced by multiple transit lines or if it is a less frequented stop. For example, a high-quality transit stop might be one where multiple bus lines converge or where both buses and metro services are available, offering more extensive and desirable routes. By not accounting for these differences, our analysis does not fully capture the expectations of property sellers regarding transit accessibility. Likewise, our transit data contains many more bus stops than metro stops thus diluting the importance of metros in our analysis.

A refinement of our research would be to rank the bus and metro. The ranking could be based on usage data of the busses or metros. Alternatively, a simpler approach that does not require any additional data would be to rank the stops based on how many different transportation lines that has that particular stop in its routes. However, implementing this approach would necessitate redefining our local accessibility measure to account for the varying significance of different stops.

5.2 Control variables - additional data

In our models, we have used some control variables, that are descriptive of a property's characteristic. In our readings of the literature covering the subject of determining factors of house prices, we have stumbled upon many more, that would enrich our model. This could be either variables characterising the socio-economic traits of the neighborhood: surrounding environment, employment rate of the area, or educational characteristics as used in Wen et al. (2014) or it could be variables better describing the regional or local accessibility to public transport; time from property to key locations in the city as in Yang et al. (2019). There are also plenty of other variables describing the characteristics of the house, that we could include: number of bathrooms, tiling of the house, or whether they have a pool, as Zietz et al. (2008) have used successfully. These are all characteristics that individuals price into their decision about whether to buy a given property or not. Data on socio-economic variables for specific areas is available behind a paywall on DST.dk, the National Danish Bureau of Statistics. Taking this into account, we have been limited in what could be accomplished regarding the gathering of characteristics, hence we choose to focus on proxy variables, such as the postal code average price, or we have omitted them, as the benefits was outweighed by the effort needed to collect them.

5.3 Alternative research setups: Natural Experiment

A limitation of our analysis is that busses are subject to changes whereas metro and rail stations are fixed in place. This also hints at why bus traffic has been overlooked in the literature regarding public transit and property prices. Also, busses are believed to be a symptom of a high traffic area, rather than a tool to create urbanization and guide urban expansion (Yang et al., 2019).

In this study, we have utilized traditional econometric methods to analyze the impact of public transport on house prices. A new trend in the research literature is natural experiments which are being utilized more frequently (Fuchs-Schündeln and Hassan, 2016). Some researchers have used this approach to look into the effect of public transport on property prices, and they have found positive and significant effects from being closer to a transport hub (E.g. Song et al. (2019), Dorantes et al. (2011)). Natural experiments offer a quasi-experimental setting, which can be more useful to control for causality.

6 Conclusion

In this study, we examined the impact of proximity to public transit on listed property prices in the greater Copenhagen area using a range of econometric models.

The hedonic model provided a straightforward analysis, revealing that properties closer to transit stops or with many transit stops within 500 meters were listed at higher prices. Our quantile regression supported this finding, indicating that low- and medium-valued properties were listed with a premium for transit proximity, while high-valued properties did not show a similar effect. When we use the local accessibility measure, we get very high and unrealistic estimates of this premium. This suggest the effect of having a short distance to the nearest transit stop reflects other underlying factors. These other factors could be things such as the distance to the city centre or areas with high job availability, such that one does not have the same need for public transport. By employing a spatial regression model, which accounts for spatial effects across properties within each postal code, we account for some of these underlying factors. Our results from the regression showed that the positive effect of regional transit accessibility persisted, though at a reduced magnitude.

Overall, our analysis underscores the complexity that is the relation between transit accessibility and property prices.

References

D. R. Bowes and K. R. Ihlanfeldt. Identifying the impacts of rail transit stations on residential property values. *Journal of urban Economics*, 50(1):1–25, 2001.

X. Cao and J. A. Hough. Hedonic value of transit accessibility: An empirical analysis in a small urban area. In *Journal of the Transportation Research Forum*, volume 47, pages 170–183, 2008.

L. M. Dorantes, A. Paez, and J. M. Vassallo. Analysis of house prices to assess economic impacts of new public transport infrastructure: Madrid metro line 12. *Transportation research record*, 2245(1):131–139, 2011.

N. Fuchs-Schündeln and T. A. Hassan. Natural experiments in macroeconomics. In *Handbook of macroeconomics*, volume 2, pages 923–1012. Elsevier, 2016.

D. H. Gatzlaff and M. T. Smith. The impact of the miami metrorail on the value of residences near station locations. *Land economics*, pages 54–66, 1993.

J. Landis, S. Guhathakurta, W. Huang, M. Zhang, and B. Fukuji. Rail transit investments, real estate values, and land use change: A comparative analysis of five california rail transit systems. 1995.

L. Osland. An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, 32(3):289–320, 2010.

D. A. Rodriguez and F. Targa. Value of accessibility to bogotá's bus rapid transit system. *Transport Reviews*, 24(5):587–610, 2004.

A. Se Can and I. Megbolugbe. Spatial dependence and house price index construction. *The Journal of Real Estate Finance and Economics*, 14(1):203–222, 1997.

H. M. So, R. Y. Tse, and S. Ganesan. Estimating the influence of transport on house prices: evidence from hong kong. *Journal of property valuation and investment*, 15(1):40–47, 1997.

Z. Song, M. Cao, T. Han, and R. Hickman. Public transport accessibility and housing value uplift: Evidence from the docklands light railway in london. *Case studies on transport policy*, 7(3):607–616, 2019.

D. Transport. Transportvaneundersøgelsen, 2014.

H. Wen, Y. Zhang, and L. Zhang. Do educational facilities affect housing price? an empirical study in hangzhou, china. *Habitat International*, 42:155–163, 2014.

L. Yang, J. Zhou, O. F. Shyr, et al. Does bus accessibility affect property prices? *Cities*, 84: 56–65, 2019.

J. Zietz, E. N. Zietz, and G. S. Sirmans. Determinants of house prices: a quantile regression approach. *The Journal of Real Estate Finance and Economics*, 37:317–333, 2008.

7 Appendix

Control variables:	Description
housingArea	Numeric: states number of square metres of floor area in property
energyLabel_cat	Numeric: Contains information about energy consumption with lower numbers being more energy efficient
addressType	String: Contains information about type of house Types: Condo, Villa, Villa apartment
isApartment:	Boolean: 1 if property is an apartment, 0 otherwise
hasBalcony	Boolean: 1 if property has a balcony, 0 otherwise
hasElevator	Boolean: 1 if property has an elevator, 0 otherwise
hasTerrace	Boolean: 1 if property has a terrace, 0 otherwise
Explanatory variables:	
no_of_tot_stops_in_500m	Numeric: number of total public transport stop within 500m of property. Includes both bus and metro stations
clost_public_trans	Numeric: distance in km of nearest transit stop. Includes both bus and metro stations

Table 1: Variables used in regressions

Variable	VIF coefficient Local	VIF coefficient Regional
housingArea	3.516	3.089
energyLabel_cat	3.443	3.501
isApartment:	3.033	4.148
hasBalcony	2.806	2.802
hasElevator	2.508	2.468
hasTerrace	1.086	1.082
Transit variable	3.047	4.208

Table 2: Variance inflation factor as measure for multicolinerity for OLS regressions

	OLS	**OLS 2**	**Spatial**
const	14.52*	14.697*	5.071*
	(0.072)	(0.075)	(0.239)
clost_public_trans	–	−0.32*	–
		(0.071)	
no_of_tot_stops_in_500m	0.029*	–	0.013*
	(0.003)		(0.003)
housingArea	0.008*	0.008*	0.007*
	(0.0)	(0.0)	(0.0)
isAppartment	−0.224*	−0.166*	−0.39*
	(0.035)	(0.035)	(0.031)
hasBalcony	0.114*	0.115*	0.084*
	(0.024)	(0.025)	(0.021)
hasElevator	0.207*	0.177*	0.113*
	(0.031)	(0.031)	(0.026)
hasTerrace	0.14*	0.15*	0.102*
	(0.036)	(0.037)	(0.03)
energyLabel_cat	−0.036*	−0.035*	−0.049*
	(0.01)	(0.01)	(0.008)
postalcode_avg_price	–	–	0.908*
			(0.024)
R^2	.65	.64	.77

Table 3: Coefficients for OLS and Spatial regression. ”-” represents that the variable is not included in the regression. Standard deviation of estimate is included in parenthesis

	Q 1	**Q 2**	**Q 3**	**Q 4**	**Q 5**	**Q 6**	**Q 7**	**Q 8**	**Q 9**
const	14.58*	14.44*	14.497*	14.541*	14.514*	14.574*	14.56*	14.567*	14.623*
	(0.069)	(0.056)	(0.054)	(0.057)	(0.057)	(0.054)	(0.048)	(0.045)	(0.054)
housingArea	0.006*	0.007*	0.008*	0.008*	0.01*	0.01*	0.011*	0.011*	0.011*
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
isAppartment	−0.357*	−0.222*	−0.197*	−0.152*	−0.091*	−0.086*	−0.042	−0.027	−0.034
	(0.032)	(0.029)	(0.029)	(0.033)	(0.034)	(0.033)	(0.031)	(0.03)	(0.037)
hasBalcony	0.035	0.092*	0.118*	0.118*	0.152*	0.14*	0.127*	0.086*	0.074*
	(0.033)	(0.03)	(0.03)	(0.033)	(0.034)	(0.032)	(0.029)	(0.027)	(0.033)
hasElevator	0.336*	0.32*	0.283*	0.218*	0.136*	0.119*	0.056	0.075*	0.032
	(0.044)	(0.038)	(0.037)	(0.04)	(0.041)	(0.039)	(0.035)	(0.033)	(0.041)
hasTerrace	0.112*	0.151*	0.178*	0.195*	0.181*	0.126*	0.162*	0.138*	0.194*
	(0.051)	(0.046)	(0.045)	(0.05)	(0.051)	(0.049)	(0.045)	(0.043)	(0.052)
energyLabel_cat	−0.052*	−0.039*	−0.04*	−0.039*	−0.034*	−0.029*	−0.02*	−0.009	−0.007
	(0.01)	(0.009)	(0.009)	(0.01)	(0.011)	(0.01)	(0.009)	(0.009)	(0.011)
clost_public_trans	−0.204	−0.309*	−0.391*	−0.497*	−0.553*	−0.537*	−0.349*	−0.138*	0.046
	(0.112)	(0.09)	(0.084)	(0.089)	(0.088)	(0.083)	(0.073)	(0.067)	(0.08)
R^2	.389	.398	.395	.388	.393	.405	.424	.453	.486

Table 4: Coefficient for regression on closest public transit. Standard deviation of estimate is included in parenthesis

	Q 1	**Q 2**	**Q 3**	**Q 4**	**Q 5**	**Q 6**	**Q 7**	**Q 8**	**Q 9**
const	14.411*	14.268*	14.291*	14.259*	14.283*	14.315*	14.356*	14.451*	14.562*
	(0.061)	(0.052)	(0.05)	(0.05)	(0.049)	(0.046)	(0.043)	(0.042)	(0.045)
housingArea	0.006*	0.007*	0.008*	0.008*	0.009*	0.01*	0.011*	0.011*	0.012*
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
isAppartment	−0.36*	−0.277*	−0.256*	−0.226*	−0.201*	−0.164*	−0.09*	−0.063	−0.055
	(0.033)	(0.03)	(0.03)	(0.031)	(0.032)	(0.032)	(0.032)	(0.032)	(0.038)
hasBalcony	0.061	0.105*	0.109*	0.135*	0.145*	0.14*	0.129*	0.084*	0.079*
	(0.033)	(0.03)	(0.029)	(0.03)	(0.031)	(0.03)	(0.029)	(0.028)	(0.032)
hasElevator	0.3*	0.32*	0.318*	0.277*	0.208*	0.171*	0.102*	0.099*	0.057
	(0.045)	(0.04)	(0.037)	(0.038)	(0.038)	(0.037)	(0.035)	(0.034)	(0.04)
hasTerrace	0.12*	0.164*	0.156*	0.177*	0.181*	0.153*	0.157*	0.134*	0.177*
	(0.05)	(0.045)	(0.044)	(0.046)	(0.047)	(0.046)	(0.044)	(0.044)	(0.051)
energyLabel_cat	−0.059*	−0.045*	−0.041*	−0.036*	−0.039*	−0.03*	−0.016	−0.008	−0.007
	(0.01)	(0.009)	(0.009)	(0.01)	(0.01)	(0.01)	(0.009)	(0.009)	(0.011)
no_of_tot_stops_in_500m	0.027*	0.032*	0.036*	0.041*	0.041*	0.035*	0.024*	0.017*	0.008*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
R^2	.398	.411	.411	.407	.410	.419	.433	.457	.487

Table 5: Coefficient for regression on number of busstops within 500 metres. Standard deviation of estimate is included in parenthesis