Statistical Analysis of the Relationship between Public Transport Accessibility and Flat Prices in Riga

Pavlyuk, Dmitry

Transport and Telecommunication Institute

14 August 2009

Statistical Analysis of the Relationship between Public Transport Accessibility and Flat Prices in Riga

Pavlyuk Dmitry
Candidate of Economic Science, Transport and Telecommunication Institute
Lomonosova 1, Riga, LV-1019, Latvia
Phone: (+371)29958338. Email: Dmitry.Pavlyuk@tsi.lv

ABSTRACT

The relationship between public transport accessibility and residential land value is a point of interest of many recent researches. A hedonic price regression model, widely used in this research area, has one very important shortcoming – it calculates an "average" influence of factors on land value in the analysing area. Usually spatial effects present in data, and the influence of public transport accessibility can be distributed over the area non-uniformly.

In this study we apply a comparatively new modification of the regression model – geographically weighted regression – to examine the relationship between public transport accessibility and residential land value (in a form of rent and sell prices) in Riga. The proposed method allows taking into account spatial effects present in the relationship.

We use information about geographical locations of urban public transport stops and routes to calculate a level of transport accessibility.

Together with the transport accessibility level and a common set of property-specific parameters (floor area, number of rooms, etc.) we consider additional hedonic properties of a flat location such as distances to supermarkets, higher schools and natural attractors like large parks, the river, and the seaside.

Keywords: geographically weighted regression, hedonic price model, public transport accessibility

Introduction

Nowadays public transport is one of the main components of a city infrastructure. Improvements of public transport routes presents in almost all city development plans ([1]), and the main reason for it is a classical assumption about a positive relationship between the quality of the public transport infrastructure and land value. Good quality of public transport increases people's mobility, improves living conditions in city outskirts, raises business activities, and, as a result, makes the land value higher.

Classically the land value can be considered from two main view points – business activities and residential living. In this research we use the second approach and consider the residential land value in two forms – flat sell prices and flat rent prices. Both sell and rent prices indicate the residential land value, but with different aspects. Sell prices shows the long-term land value and influenced mostly by stable factors like living utilities and public transport around the house. Rent prices are significantly oriented by a number of temporary habitants like tourists and students, and that's why influenced by location factors – distances to tourist attractions, sea/river sides, higher education institutes. The influence of tourists on rent prices are especially significant for Riga, where the number of tourists arriving every quarter is similar to the total city population, and every third tourist chooses a non-hotel accommodation.

The usual econometric approach widely used ([2]) in this area is a hedonic price regression model. The hedonic regression can be formulated as:

$$Price = \mathrm{F}(PropertyFactors, LocationFactors, TransportFactors) + \varepsilon,$$

where the dependent variable *Price* is explained by three sets of factors – property, location, and transport characteristics [3].

A usual regression model shows how factors influence a dependent variable *in average*. This is good for understanding of dependencies in cross-sectional non-spatial data, but this is a huge disadvantage in a spatial case like housing data. As an example let us consider the dependence between a flat floor area and a flat price. If we estimate the usual regression in this case, we will receive an average price per square metre in a city. But this number will not be useful in case of significant deviations of this indicator in the city parts; the average value will be bad approximation for every particular flat. The obvious way to solve this problem is including district-specific dummy variables into the regression (or cross-dummy variables as products of district-specific variables and the floor area, for example). This approach allows estimating differences in values of the dependent variable and influence of explanatory variables by city districts, but also has shortcomings. The main shortcoming is an artificial nature of city districts (administrative or study-specific).

There no any reason why two flats located across the street but in different districts should be more different than two flats located in different parts of the same district.

In this research we overcome these disadvantages using the Geographically Weighted Regression (GWR) [4], a method based on the spatial nature of housing data.

The main practical goal of the research is to answer the question – how does the public transport accessibility influence the residential land value. The first subtask we solved here is construction of an indicator of transport accessibility. There are different ideas used in researches in this area – from a simple distance to the nearest public transport stop to complex calculations of average travel times to different locations like schools, a city business centre, other attractors [5]. We constructed our own public transport accessibility index, which based on the distance decay function and used as much information as possible, but remained relatively simple in calculations. This new indicator is used in the GWR to reflect the target relationship.

Within the bounds of the model we investigated the influence of other location factors like distance to supermarkets, sea/river, parks, higher education institutes, and property characteristics like area, standard flat projects and floor numbers.

Theoretical foundation

Geographically weighted regression

The traditional regression model can be written as:

$$y = X \cdot \beta + \varepsilon$$

where y is a vector of dependent variable values, X is a matrix of explanatory variables values, β is a vector of unknown coefficients and ε is a random term. Coefficients β are estimated for the global data set and shows average influences of explanatory variables.

The geographically weighted regression has a difference in coefficients [4]:

$$y = X \cdot \beta(u_i, v_i) + \varepsilon$$

where u_i and v_i are the coordinates of a house (a latitude and a longitude). Obviously, we need for additional information about the house location to estimate this regression, and the regression coefficients will be estimated for each flat separately. In this case we receive different estimates for each data point, which allows investigating the spatial behaviour of relationships.

For estimating the coefficients of the model for a particular flat i, we assume that flats located near to the flat i, have higher influence on the coefficients than flats located far from the flat i. In terms of regression, it means that flats located near to the point i, have larger weights. Usually the bi-square function is used for calculating a weight of the point j for the point i model:

$$w_{ij} = \begin{cases} \exp\left(-distance_{ij} \big/ bandwidth \right), \text{for } distance_{ij} \leq bandwidth, \\ 0, \text{for } distance_{ij} > bandwidth \end{cases}$$

where $distance_{ij}$ is a distance from the point i to the point j. The *bandwidth* reflects the speed of weight decreasing. There are two main approaches for the bandwidth selection – fixed and adaptive bandwidths. The fixed bandwidth is selected once for all data points, the adaptive can be changed from point to point depending on data density. The illustration of the adaptive bandwidth is presented on the Figure 1.

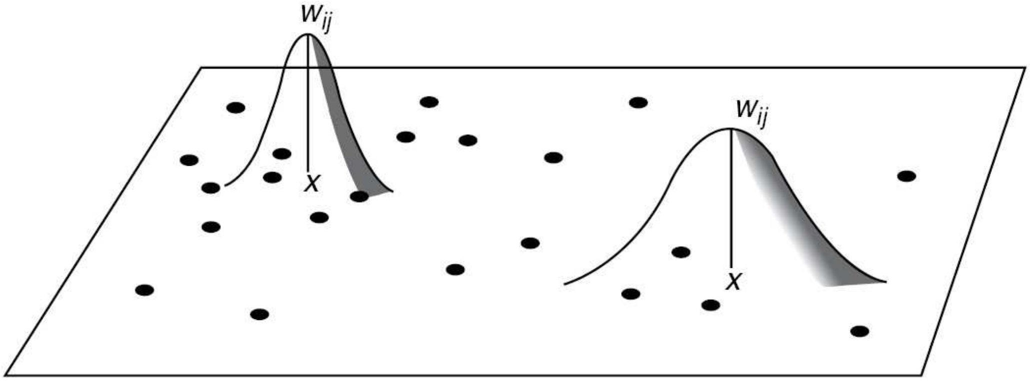

Figure 1. Geographically weighted regression with adaptive spatial bandwidths. Source: Fotheringham et al.[4]

The best bandwidth can be chosen with the Akaike Information Criterion (AIC).

In our case flats data have an inconstant density, so we used the adaptive bandwidth approach.

The coefficients are calculated for each data point using the weighted least squares method:

$$\hat{\beta}_i = \left(X^T W_i X\right)^{-1} X^T W_i Y$$

where W_i is a weight matrix with w_{ij} on the main diagonal and zeros on other positions.

The result of the estimation procedure is a set of estimates of coefficients for each data point.

Public transport accessibility

Measurement of public transport accessibility is deeply investigated in some modern studies ([6]). For our investigation we need in indicator of public transport accessibility for each flat. Obviously, the indicator should be based on the house location and distances to nearest public transport stops. If we consider a utility of a distance to the stop, we will see the classic sigma-shaped function. Some first tenth metres doesn't make a difference, after that the utility is quickly decreasing, and stabilising at a lower level again after that. We chose the log-logistic distance decay function to reflect this behaviour ([7]).

Another problem was related with taking a number of stops into account. We suggest using a sum of utilities of individual stops to aggregate all stops around the house.

So the suggested indicator can be presented by the next formula:

$$PublicTransportAccessibility_i = \sum_{route\ stop \in AREA} \frac{1}{1+\exp\left(a + b\ln\left(distance_{i,route\ stop}\right)\right)}$$

where $a = -20, b = 3.7$

The values of parameters a and b are chosen on the base of our fillings. The resulting distance decay function is presented on the Fig. 2.

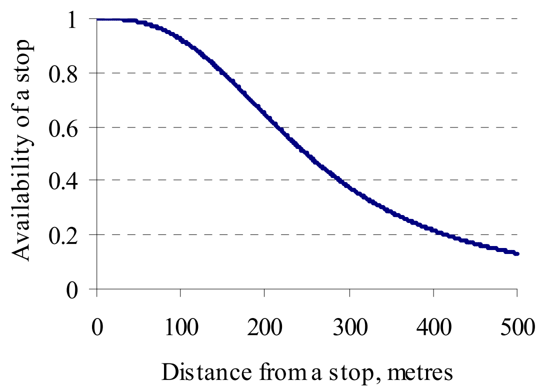

Figure 2. Distance decay function for public transport stops availability

The utility of the distance from the stop is decreasing proportionally in the interval from 100 to 400 metres; distances less than 100 and more than 400 metres don't make differences.

The sum in the formula aggregates accessibility of all route stops in the *AREA*. We use an area instead of calculations of accessibility of all transport routes, because we assume that if a stop is located significantly far from the flat it will not be used. If this research we used a circle with 1 kilometre radius to limit the *AREA*.

Data

We collected data about flats for sell and rent in Riga. Both samples include about 1000 flat notices dated May 2009. Also information about the public transport infrastructure, supermarkets, higher education institutes and other attractors' location is collected.

Data was received from the data sources below:

1. Local real estate bulletin board website (SS.lv) was used for information about flats for sell and rent. The information includes flats' characteristics – a floor area, a number of rooms, flat project plans, a floor number and a price requested. We realise that using requested prices can lead to a bias (to the higher end), but according to Hometrack [8] requested and real deal prices are highly correlated (the coefficient of correlation = 0.986) in UK and we assume the similar situation in Riga. The main goal of our research is relationships between the flat price and other characteristics, so usage of highly correlated substitute should not lead to a bias of coefficients' estimates.
   Also information about flat addresses is provided on the website. This information was translated into geographical coordinates (latitude and longitude) using the Google Earth service.
2. The information about public transport routes, stops, and schedules was received from the official site of the only public transport municipal enterprise in Riga – Rigas Satiksme.
3. The Google Local Business Centre was used for receiving information about locations of markets, supermarkets, and railway stations.
4. A number of websites were used to supplement the Google Local Business Centre information.
5. Natural attractors (the sea side, Daugava river, green parks, including Mezaparks) were added to the Google map manually.

Distances from each flat to the nearest supermarket or market, the nearest higher education institute are computed. Absolute distance values (without usage of a distance decay function) are used.

Also the indicator of public transport accessibility was calculated on the base of the formula presented above. Buses, trams, trolleybuses, and railways are considered as public transport in this research.

Results

The first practical subtask of the research was calculation of public transport accessibility values. The indicator was calculated for every flat and varies from 1 to 47 points (according to the formula, each point equals to a stop just near to the flat).

Accessibility values interpolated by the Kriging procedure ([9]) are presented on the Figure 3.

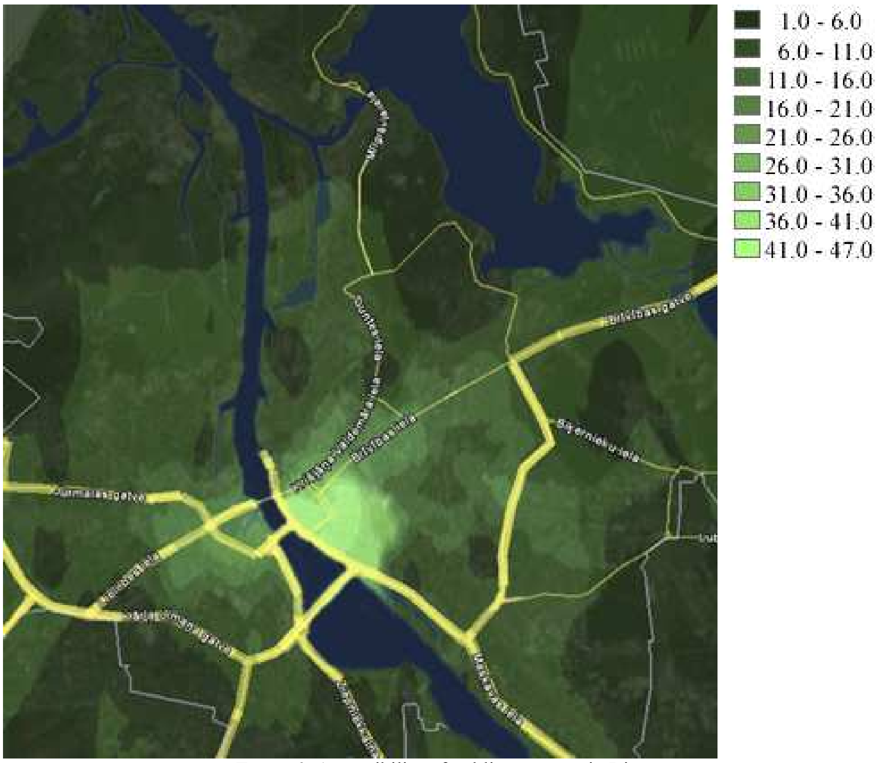

Figure 3. Accessibility of public transport in Riga

On the Fig.3 brighter areas mean higher, darker – lower public transport accessibility. As expected the city centre has the highest values, and outskirts have lower accessibility level.

Public transport accessibility doesn't mean its utility; habitants of the city centre have a possibility to travel to any part of the city from a stop near to the house, but they don't use all these possibilities. Habitants of outskirts usually have one-two options to travel on public transport (usually to the city centre), but they use this option very often.

Let consider the usual global hedonic regression models for rent and sell prices (Table 1).

Table 1. Global hedonic regression models

Dependent Variable	*Rent Price*		*Sell Price*	
	Coefficient	P-value	Coefficient	P-value
FloorArea	2.8	0.000	794.6	0.000
DistanceToWater	-9.2	0.019	-4042.5	0.011
DistanceToGreen	6.1	0.376	5704.8	0.050
DistanceToSupermarket	-23.8	0.012	-5853.6	0.170
DistanceToHigherEducation	-16.1	0.000	-2265.6	0.185
PublicTransportAvailability	1.2	0.000	810.2	0.000
FirstFloor	-3.4	0.598	-3686.5	0.179
Project602	-4.1	0.649	-303.1	0.930
ProjectNew	91.1	0.000	26334.6	0.000
ProjectRenov	73.8	0.000	22310.5	0.000
Project119	-20.9	0.089	-5427.2	0.227
Constant	14.3	0.213	-15726.3	0.001

We can see significant relationships in both regressions which can be analysed and explained.

The indicator of public transport accessibility has highly significant positive coefficients in both regressions, so we can make a conclusion about positive influence of the factor. But let compare the influence of a new route in a city outskirt and in the city centre. We think that a new route or a new stop is much more important for the outskirt and completely unimportant for the city centre. So the average value, estimated by the global regression is correct in average, but doesn't reflect spatial features.

Another similar example is the absence of an effect of the first floor. Usually flats on the first floor are cheaper, but it also depends on the house position. We guess that the relationship, insignificant in average, can be significant for some city parts.

Also we note other coefficients which don't much our initial expectations (a negative relationship with distances to water, supermarkets, and higher education institutes).

Dummy variables for flat's projects have expected values (*Project602* and *Project119* are standard flat projects popular in Riga, *ProjectNew* separates new buildings. *ProjectRenov* separates old buildings after renovation).

So we note potential problems in our global regression, possibly related with the spatial effects. One of formalised tests for spatial correlation is the Moran I coefficient ([9]). Calculated values for residuals of global hedonic sell and rent price models are presented in the Table 2.

Table 2. Moran's I values

	200 metres area		1 kilometre area	
Residuals of the global regression model for	Moran's I	p-value*	Moran's I	p-value*
Rent Price	0.066	0.000	0.036	0.000
Sell Price	0.113	0.000	0.098	0.000

* one-tail test

We calculated Moran I values for nearest neighbourhood (200 metres around a flat) and a mini-district (1 kilometre around the flat). In both cases for residuals of both model we received highly significant values, which indicate the presence of spatial relationships in models' error terms. The positive sign of Moran I's attests similar residual values of flat prices in one mini-district.

On the base of previous reasoning and Moran I's values we decided to use the geographically weighted regression. In this paper we present the most interesting moments of the resulting regression only.

The first practical question for the regression was a relationship between prices and public transport accessibility. According to the GWR approach coefficients are calculated for every data point. The most interesting things for us are significance and direction of the relationship. Both characteristics can be presented by coefficient's t (Student) value. The interpolations of t-values (for the rent price model) are presented on the Figure 4.

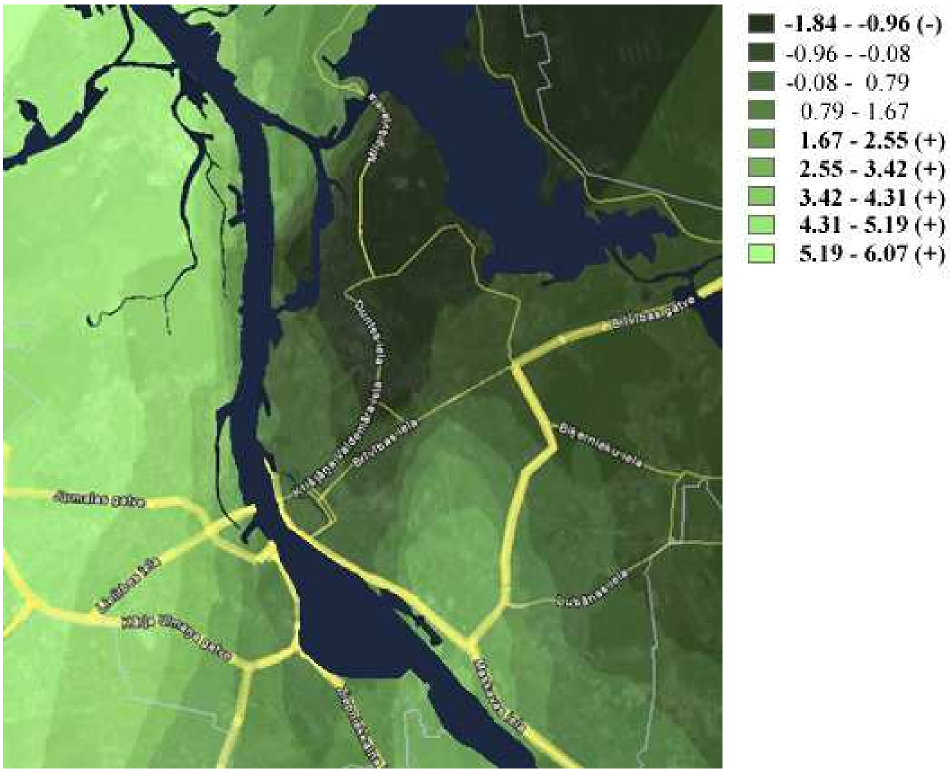

Figure 4. The relationship between public transport accessibility and flat rent prices (t-values)

Darker areas reflect negative relationship, brighter areas – positive relationship (significant areas are marked in the legend with "–" and "+" signs accordingly).

We can note that the most outskirts have an expected positive relationship, which indicates that every new transport route and stop will increase flat prices. The areas with a significant positive relationship cover a sizeable share of the map, which leads to a significantly positive average value (in the global regression).

As we expected the city centre doesn't have a significant relationship between flat prices and transport accessibility. The city centre is rich with transport routes and new routes will not affect flat prices.

There are areas with a significant (or near to be significant) negative relationship. The areas cover Mezaparks and adjacent regions, which is a place with good ecological environment and beautiful nature located not far from the city centre, and, therefore, a place of settlement of people with higher income level. This factor leads to higher level of private car usage by habitants and higher utility of peace of quiet, hardly compatible with intensive public transport movements. Surprisingly, the negative relation is near to be significant in the Sarkandaugava district, which additionally can be explained by roads with intensive transport movements used by public transport. The nearness of thoroughfare, not included into the models directly, can overpower positive effects of the public transport.

We can compare global regression and geographically weighted regression models using the Akaike information criterion, AIC (Table 3)

Table 3. Model comparison on the bas of Akaike Information Criterion

	Global regression	Geographically Weighted regression
Rent Price	9994	9893
Sell Price	13576	13465

Smaller values of AIC indicate better models; a difference in more than about 3 point is a significant one. As we can see the GWR regressions are significantly better in both cases.

Relationships between a flat price and other explanatory variables also have spatial effects. For example, the influence of the first floor on the price was insignificant in the global regression model, but using the GWR we discovered local dependencies. In the sell price model the first floor is highly significant in the city centre (and in the Pardaugava district situated on the opposite river side) and has a negative effect. For all other city regions the influence of the first floor on the price is insignificant.

Conclusions

In this research we constructed the indicator of public transport accessibility and estimated its relationship with the residential land value (in forms of flat rent and sell prices).

The public transport accessibility indicator was constructed on the base of information about spatial positions of houses, and stop and routes of public transport. The calculated values have an expected distribution – the city centre has higher values of public transport accessibility and city outskirts have lower values. The suggested indicator can be used for further research of the transport infrastructure.

Analytical reasoning and formalised tests lead to a conclusion about unsatisfactory results of the usual global hedonic regression model. We used a relatively new method – the geographically weighted regression – to reflect spatial effects in the data. Estimating the GWR parameters we note significant differences in influence of public transport accessibility on flat prices. We ascertained city areas with positive and negative relationships, and also the city centre with absence of the significant relationship.

We investigated influence of other flat characteristic like floor area, floor number, distances to supermarkets, higher education institutes and natural attractors, and also discovered spatial features.

The results of the research can be used in composing of public transport infrastructure chapter of the city development plan.

References

1. *Spatial Plan of Riga for 2006-2018.* Riga City Council, 2005, 214 p.
2. *Report for the Royal Institution of Chartered Surveyors and the Office of the Deputy Prime Minister,* Atis Real Weatheralls, University College London, and Symonds Group Land Value and Public Transport, 2002, 36 p.
3. Fik, T. J. and Ling, D. C., and Mulligan, G. F. Modeling spatial variation in housing prices: a variable interaction approach, *Real Estate Economics*, 4, 2003, pp. 623-646.
4. Fotheringham A. S., Brunsdon C., and Charlton M. E. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley, 2002. 282 p.
5. Du H. and Mulley C. Transport accessibility and land value: a case study of Tyne and Wear, *RICS Research paper series*, Vol. 7 No 3, 2007, 52 p.
6. Levinson D. M. and Krizel K.J. *Access to Destinations*. New York: Elsevier Science, 2005. 422 p.
7. Geurts K.T., Ritsema van Eck J.T. Accessibility Measures: review and applications, *RIVM Report 408505 006*, Utrecht University, 2001, p. 147

8. *Hometrack report. Key housing market statistics for England and Wales*, UK House Price Statistics, 2005, 8 p.
9. Haining R. *Spatial Data analysis. Theory and Practice.* Cambridge: Cambridge University Press, 2004. 431 p.